8TH IEEE SIGNAL PROCESSING WORKSHOP ON

STATISTICAL SIGNAL AND ARRAY PROCESSING

JUNE 24–26, 1996
CORFU, GREECE

Sponsored by
The IEEE Signal Processing Society

IEEE COMPUTER SOCIETY
50 YEARS OF SERVICE •1946-1996

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	Jan 10, 1997	Technical 08|01|96 -12|31|96

4. TITLE AND SUBTITLE: 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing.

5. FUNDING NUMBERS

6. AUTHORS(S)
G.B. Giannakis and A. Swami

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Rector and Visitors of the University of Virginia
Office of Sponsored Programs
P.O. Box 9003
Charlottsville, VA 22906

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)
Office of Naval Research
Ballston Tower One
800 North Quincy Street, Arlington, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The view, opinion, and/or findings contained in the report are those of the authors and should not be construed as an official Department of the Navy position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Contains the text of 139 contributed papers and 4 invited plenary talks, documenting recent advances in the theory and applications of statistical signal and array processing. The center of focus is "SSAP for communications", with other sessions devoted to detection, estimation and spectral analysis, array processing, non-Gaussian signals, multichannel models, source separation, nonlinear models, and non-stationary signals.

14. SUBJECT TERMS: Signal Processing, Array Processing, Non-Gaussian Signals, NonLinear Models, Non-Stationary Signals, MultiChannel Models, Communications, Spectral Analysis, Source Separation, Estimation-Detection.

15. NUMBER OF PAGES
7

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

SSAP'96

Proceedings

8th IEEE Signal Processing Workshop on

Statistical Signal and Array Processing

June 24-26, 1996
Corfu, Greece

Sponsored by

The IEEE Signal Processing Society

With support from

U.S. Army Research Office
U.S. Office of Naval Research

IEEE Computer Society Press
Los Alamitos, California

Washington • Brussels • Tokyo

IEEE Computer Society Press
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314

IEEE Computer Society Press Order Number PR07576
IEEE Order Plan Catalog Number 96TB100040
ISBN 0-8186-7576-4
Library of Congress Number 96-75525

Additional copies may be ordered from

IEEE Computer Society Press
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: +1-714-821-8380
Fax: +1-714-821-4641
Email: cs.books@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: +1-908-981-1393
Fax: +1-908-981-9667
mis.custserv@computer.org

IEEE Computer Society
13, Avenue de l'Aquilon
B-1200 Brussels
BELGIUM
Tel: +32-2-770-2198
Fax: +32-2-770-8505
euro.ofc@computer.org

IEEE Computer Society
Ooshima Building
2-19-1 Minami-Aoyama
Minato-ku, Tokyo 107
JAPAN
Tel: +81-3-3408-3118
Fax: +81-3-3408-3553
tokyo.ofc@computer.org

Editorial production by Bob Werner
Cover art design and production by Joe Daigle/Studio Productions
Printed in the United States of America by KNI, Inc.

The Institute of Electrical and Electronics Engineers, Inc.

Table of Contents

IEEE Signal Processing Workshop on Statistical Signal and Array Processing

Plenary

MA-1: Detection, Estimation, and Spectral Analysis

Chairs: P. Djuric and A.O. Hero

MA-2: Array Processing I

Chairs: B. Friedlander and L. Swindlehurst

Plenary

MP-1: Signal Processing for Communications I

Chairs: P. Duhamel and L. Tong

MP-2: NonGaussian Signal Processing

Chairs: J.M. Mendel and C.L. Nikias

Plenary

TA-1: Applications

Chairs: T. Durrani and M. Lagunas

TA-2: Multichannel ID – Source Separation

Chairs: J.-F. Cardoso and E. Moulines

Plenary

TP-1: Nonlinear Signal Processing

Chairs: N. Kalouptsidis and E. Masry

TP-2: Nonstationary Signal Processing

Chairs: M. Amin and B. Porat

Plenary

WA-1: Signal Processing for Communications II

Chairs: G. Geraniotis and J. Proakis

WA-2: Array Processing II

Chairs: J. Bohme and M. Kaveh

Panel

Message from the Chairs

Welcome to the Eighth IEEE Statistical Signal and Array Processing (SSAP) Workshop. Corfu, in Greek, means apex and thus sounds most appropriate as a choice to host the biannual SSAP'96 summit of the Signal Processing Society. We look forward to an exciting and memorable meeting. The workshop venue is the Corfu Hilton in the heart of Corfu, featuring beaches amid cliffs and pines, and the atmosphere to promote the exchange of technical ideas while enjoying the Greco-Ionean ambiance.

Statistical signal and array processing continues to be the backbone of many real-world engineering applications, and consistent with previous meetings, we expect SSAP'96 to continue the tradition of excellence in the technical quality of presentations on state-of-the-art research. The international character of the workshop keeps growing, and this year's meeting, being the first one to move away from North America, is well attended by European participants. As with previous SSAP meetings, we have introduced some changes in the organization and the emphasis of the meeting. Correspondence with authors was primarily via e-mail, and for publicity and notifications we relied heavily on our regularly updated home page (http://watt.seas.virginia.edu/~spirit/ssap96/). Thanks to external support, we were able to offer bargain basement registration fees ($550 for regular and $450 for student attendees).

We received 270 summaries from 45 countries– a record number of submissions for SSAP. Each submission was scored by three reviewers, and in order to maintain the workshop's atmosphere we accepted only 139 papers which we expect to be of high quality. Our apologies to authors whose fine submissions we could not accommodate, and our sincere thanks to reviewer experts, mostly drawn from the SSAP Technical Committee, for their help with excellent and timely reviews. Signal Processing for Communications and Array Signal Processing were well represented in the number of submissions (and thus in the number of accepted papers). Applications, detection-estimation, non-Gaussian, non-stationary, and nonlinear formed other well-defined clusters, and all are represented in the ten poster sessions and five outstanding plenary talks. The center of focus for this year's research theme is SSAP for Communications.

Our warm thanks go to the volunteers of the international program committee, the European and Austral-Asian liaisons, the publication, publicity, and local arrangement chairs. The informative and creative home page prepared by Guotong Zhou contributed significantly to the workshop (its format is now being used as a template by other workshops). Maria Rangoussi's efforts in Greece are also greatly appreciated (she bridged the transatlantic distance with the organizers in the US). We finally wish to acknowledge support from the U.S. Army Research Office, the U.S. Office of Naval Research, the Greek General Secretariat for Research and Technology, and the Greek companies Intracom and Alpha.

We hope that your stay in Corfu will not only be technically enriching but will also give you the opportunity to meet new fellow researchers, renew old acquaintances, and to enjoy the Greek sea and sun. We look forward to meeting you in Corfu.

Georgios B. Giannakis and Ananthram Swami
Co-organizers and Co-Chairs

Conference Organization

Georgios B. Giannakis
Dept. of Electrical Engr.
University of Virginia
Charlottesville, VA 22903-2442, USA
Tel: (804) 924-3659
Fax: (804) 924-8818
e-mail: <georgios@virginia.edu>

Ananthram Swami
Malgudi Systems
P.O. Box 4640
Silver Spring, MD 20914-4640, USA
Tel: (301) 762-6796
Fax: (301) 762-2450
e-mail: <a.swami@ieee.org>

European Liaison

Mats Viberg, Chalmers Inst. Tech., Sweden

Australia-Asia Liaison

Yujiro Inouye, Osaka University, Japan

Technical Committee

Moeness Amin, Villanova University, USA
Sergio Barbarossa, University of Rome, Italy
Pierre Comon, Thomson-Sintra, France
Zhi Ding, Auburn University, USA
Petar Djuric, SUNY at Stony Brook, USA
Yingbo Hua Univ. of Melbourne, Australia
Fu Li, Portland State University, USA
Hagit Messer, Tel-Aviv University, Israel
Eric Moulines, Telecom-Paris, France
Boaz Porat, Technion, Israel
Surendra Prasad, I.I.T.-Delhi, India
Lang Tong, Univ. of Connecticut, USA
Michael Zoltowski, Purdue University, USA

Local Arrangements

Nicholas Kalouptsidis, Univ. of Athens, Greece
Maria Rangoussi, NTUA, Greece
National Tech. Univ. of Athens, Greece
e-mail: <maria@theseas.ntua.gr>

Publications

Brian Sadler, Army Research Lab, USA

Publicity

Guotong Zhou, Georgia Inst. Tech., USA

Treasurer

Michail Tsatsanis, Stevens Inst. Tech., USA

8th IEEE SIGNAL PROCESSING WORKSHOP
on
STATISTICAL SIGNAL AND ARRAY PROCESSING

June 24-26, 1996

The Corfu Hilton, Corfu, GREECE

CALL FOR PAPERS

This workshop is the eighth in a series of biannual meetings sponsored by the IEEE Signal Processing Society. Following the successful scheme of previous meetings, the workshop will feature keynote addresses by leading researchers in the area, and poster sessions consisting of both invited as well as contributed papers. Participation will be limited.

Authors are invited to submit contributions in the areas of, but not limited to :

- Power spectrum analysis
- Higher-order spectra in signal processing
- Detection and estimation theory
- Sensor array processing
- Performance analysis
- Nonlinear and chaotic signals and systems
- Non-stationary processes: Time-frequency and time-scale representations; evolutionary spectra
- Cyclo-stationary signal analysis
- Signal processing for communications
- Computational and implementation issues
- Applications in all areas

Prospective authors should submit four copies of a hundred word abstract and a two to four page extended summary to G.B. Giannakis; the summary should include affiliations, addresses, tel/fax numbers and e-mail addresses, and keywords identifying one of the above topics.

Important Dates:

Submission of summary	**December 1, 1995**
Notification of acceptance	**February 1, 1996**
Camera-ready paper	**March 15, 1996**

Organizers and Co-Chairs

Georgios B. Giannakis
Dept. of Electrical Engr.
University of Virginia
Charlottesville, VA 22903-2442, USA
Fax: (804) 924-8818
e-mail: georgios@virginia.edu

Ananthram Swami
Malgudi Systems
P.O. Box 4640
Silver Spring, MD 20914-4640, USA
Fax: (301) 762-2450
e-mail: a.swami@ieee.org

European Liaison
Mats Viberg, Chalmers Inst. Tech., Sweden

Austral-Asian Liaison
Yujiro Inouye, Osaka University, Japan

Technical Committee

Moeness Amin, Villanova University, USA
Sergio Barbarossa, University of Rome, Italy
Pierre Comon, Thomson-Sintra, France
Zhi Ding, Auburn University, USA
Petar Djuric, SUNY at Stony Brook, USA
Yingbo Hua, Univ. of Melbourne, Australia
Fu Li, Portland State University, USA
Hagit Messer, Tel-Aviv University, Israel
Eric Moulines, Telecom-Paris, France
Boaz Porat, Technion, Israel
Surendra Prasad, I.I.T.-Delhi, India
Lang Tong, Univ. of Connecticut, USA
Michael Zoltowski, Purdue University, USA

Local Arrangements
Nicholas Kalouptsidis, Univ. of Athens, Greece
Maria Rangoussi, NTUA, Greece
e-mail: maria@theseas.ntua.gr

Publications
Brian Sadler, Army Research Labs, USA

Publicity
Guotong Zhou, Georgia Inst. Tech., USA

Treasurer
Michail Tsatsanis, University of Virginia, USA

e-mail: ssap96@spirit.ee.virginia.edu Home Page: http://watt.seas.virginia.edu/~spirit/ssap96
Co-sponsored by: **The U.S. Army Research Office and The U.S. Office of Naval Research**

Plenary

Antenna Array Processing

H.L. Van Trees

MA-1: Detection, Estimation, and Spectral Analysis

Detecting Regularity in Minefields and Chaotic Signals using the Empty Boxes Test

Douglas E. Lake
Office of Naval Research, Code 313
800 N. Quincy St.
Arlington, VA 22217
laked@onrhq.onr.navy.mil

Abstract

A simple, flexible, and robust procedure to detect regularity in point processes versus the alternative of randomness (i.e., a poisson point process) is the empty boxes test (EBT). The EBT can be extended to a multivariate statistical test in several ways including an implementation of a skeptical likelihood test (SLT). These approaches have previously been used to detect the regularity of minefields, a two-dimensional point process, where the alternative is termed complete spatial randomness (CSR). In this paper, these methods are applied to the problem of detecting regularity in chaotic signals such as pseudo-random number generators.

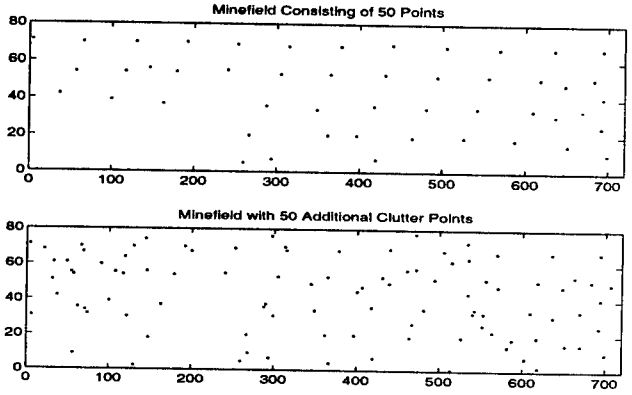


Figure 1. Examples of a minefield with 50 mines and 50 additional random clutter points

1. INTRODUCTION

Detecting minefields in the presence of clutter is an important challenge for the Navy. Minefields have point patterns that tend to exhibit regularity such as equal-spacing and collinearity that provide potentially valuable discriminants against natural occuring clutter which tends to exhibit *complete spatial randomness* (CSR). These tendencies arise because of a variety of compelling factors including strategic doctrine, safety, tactical and economic efficiency, and perhaps most intriguing the human element. In [4] and [5], several simple procedures were introduced to detect regularity in minefields and other point processes generated by humans (e.g., lottery numbers). Figure 1 shows an example of a minefield that is not so apparent with the addition of clutter points.

Another important problem where regularity is being detected as an alternative to randomness is the identification of chaotic signals. Chaos theory is being used to develop low probability of intercept (LPI) and spread spectrum communication signals where traditional detection methods would fail. In these cases, the EBT and its variants are alternative approaches to detection worth considering. A particular interesting example to illustrate this claim is a pseudo-random number generator (which is actually a deterministic, chaotic process) with a white spectrum.

2. TESTS TO DETECT REGULARITY

A variety of methods to detect deviations from CSR in point patterns have been developed for the most part on the alternative of tendency towards clustering rather than the tendency towards regularity. Cressie provides a comprehensive overview of these and other techniques with a demonstration on the longleaf pine data set [2]. Some alternative approaches are introduced below.

2.1. Empty boxes test

Consider a CSR process with n points on a set A in R^d that has been partitioned into N reqions of equal area

0-8186-7576-4/96 $5.00

to be referred to informally as *boxes*. A variety of tests to detect regularity can be based on $M_0, M_1, \ldots, M_n$ and $Y_1, Y_2, \ldots, Y_N$ where M_r and Y_i are random variables denoting respectively the number of boxes containing exactly r points and the number of points in box i.

A simple test to test regularity is based on M_0, that is, the number of empty boxes. The so-called *empty boxes test* (EBT) based on M_0 has been around for at least forty years [3], but has traditionally been used to detect the presence of too many empty boxes as an indication of lack of fit. In this context, regular point processes (and humans) tend to overfit and clustered point processes tend to not fit well. A disadvantage of the EBT for minefield detection is that there is no explicit modelling of collinearity and regular spacing, per se; the EBT is a generic regularity detector. However, the advantages of the EBT include its flexibility, lack of edge effects and its robustness.

Another advantage of the EBT is that the null moments of the test statistic can be calculated exactly without independence assumptions on $Y_1, Y_2, \ldots, Y_N$. The expected value and variance of M_0 under CSR is given by

$$\mu_0 = E[M_0] = Np_0 \quad (1)$$

$$\sigma_0^2 = Var[M_0] = \mu_0 + N(N-1)p_{00} - \mu_0^2 \quad (2)$$

where

$$p_0 = \Pr\{Y_i = 0\} = (1 - \frac{1}{N})^n \quad (3)$$

$$p_{00} = \Pr\{Y_i = 0, Y_j = 0\} = (1 - \frac{2}{N})^n \quad (4)$$

and $i \neq j$ in (4).

2.2. Generalizing the empty boxes test

The empty boxes test can be generalized by using $M_1, M_2, \ldots,$ in addition to M_0 to form test statistics. For general r and s the moments analagous to (1) and (2) are given by

$$\mu_r = Np_r \quad (5)$$

$$\sigma_r^2 = \mu_r + N(N-1)p_{rr} - \mu_r^2 \quad (6)$$

$$\sigma_{rs} = N(N-1)p_{rs} - \mu_r\mu_s \quad (7)$$

where $\sigma_{rs} = cov[M_r, M_s]$ for $r \neq s$ and

$$p_r = \binom{n}{r}(\frac{1}{N})^r(1 - \frac{1}{N})^{n-r} \quad (8)$$

$$p_{rs} = \binom{n}{r}\binom{n-r}{s}(\frac{1}{N})^{r+s}(1 - \frac{2}{N})^{n-r-s} \quad (9)$$

as in (3) and (4).

Let $\mathbf{M} = \mathbf{M}_k$ be the multivariate statistic vector $(M_0, M_1, \ldots, M_k)^T$ with mean μ and covariance Σ. Under appropriate mild conditions,the quadratic form

$$Q = Q_k = (\mathbf{M} - \mu)^T\Sigma^{-1}(\mathbf{M} - \mu) \quad (10)$$

is approximately χ^2 with $k+1$ degrees of freedom under CSR. By considering both the sign of $M_0 - \mu_0$ with the strictly nonnegative Q_k to form the real-valued statistic

$$D = D_k = sign(M_0 - \mu_0)Q_k \quad (11)$$

a one-sided test can be constructed . Positive values of D indicate clustering and negative values indicate regularity. Tests based on D_0 are equivalent to the EBT. Moreover, Q_1 is approximately exponential so that the test statistic D_1 is approximately double exponential. A one-sided test for regularity can be constructed using the approximation

$$\Pr\{D_1 < -d\} \doteq \frac{1}{2}e^{-\frac{d}{2}} \quad (12)$$

where $d \geq 0$.

2.3. Skeptical likelihood test

It can be shown that the most likely configuration under CSR would reject CSR under the empty boxes test. The reason for this apparent paradox is that the test is rejecting observations that are *too likely* under the null hypothesis suggesting some skepticism is in order. Generally, even distributions of the points among the regions are more likely than uneven distributions. Without specifying an alternative, a *skeptical likelihood test* (SLT) for a statistic T with null distribution f is to reject H_0 for high values of $f(T)$.

A skeptical likelihood test for minefield detection can be based on the test statistic

$$T = \sum_{r=2}^{n} M_r \log r! \quad (13)$$

where significantly small values of T indicate regularity. The mean and variances of (13) can be calculated directly using (5),(6), and (7). In practice, the summation in (13) can be truncated to simplify the computation.

2.4. Detection Performance Results

To demonstrate the EBT methods on the clutter example ($n = 100$) in Figure 1 a value of $N = 100$ was selected and the 80x720 region was divided into a 5x20 grid of rectangles of equal size (16x36). One could think of this example as having a SNR of 0 dB. The statistics for this partition are $M_0 = 30, M_1 = 45, M_2 = 20$, and $M_3 = 5$ leading to P-values are of .017 , .045 , and .015 respectively for the EBT, D_1, and SLT.

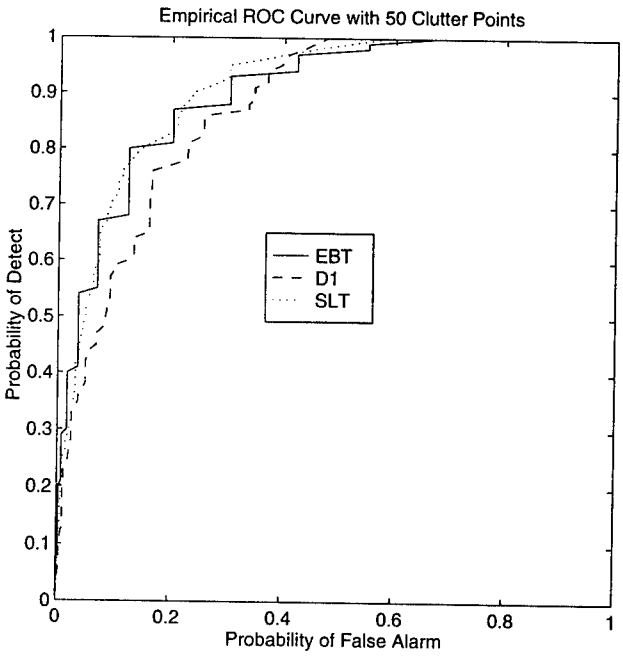


Figure 2. Empirical ROC curve for a point process with 50 mines and 50 additional random clutter points

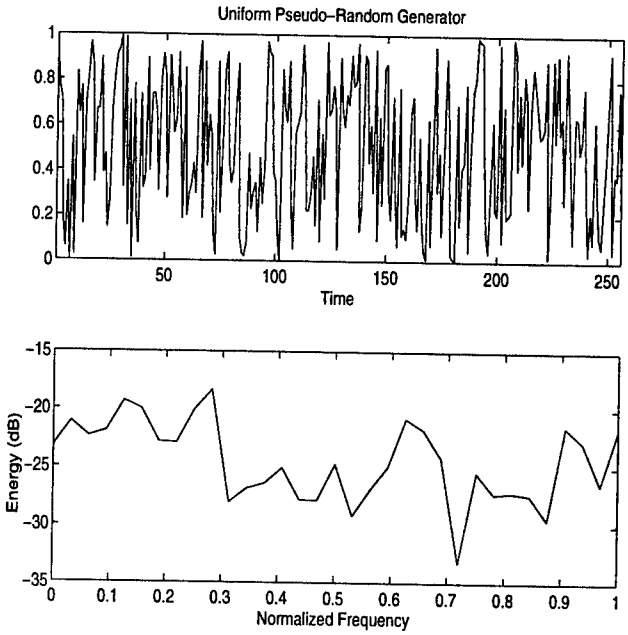


Figure 3. Example of a Pseudo-Random Process

In order to get a better understanding of the relative performance of these three methods, 100 realizations of the 50 random clutter points were simulated. Figure 2 shows the resulting empirical ROC curves. indicating that the three methods are fairly similar. For example, with a false alarm rate of .1 is approximately . the probability of detection is approximately .8. This performance is impressive considering that the patterns are not always visually obvious and these methods have no explicit modeling for collinearity and except, perhaps, for the selected dimensions of the regions no modelling of equal-spacing.

3. Examples of Chaotic Signals

Characterizing the difference between randomness and chaos is a fundamental question that is perhaps more philosophical in nature than mathematical, statistical, or physical. As is discussed recently in [1], a striking example to illustrate the fuzzy boundary of these concepts is pseudo-random number generators.

3.1. Pseudo-Random Number Generators

Uniform random variates $U_1, U_2, U_3, \ldots$ can be generated by multiplicative congruential methods of the form

$$U_{k+1} = aU_k \mathrm{mod} T \qquad (14)$$

along with some initial integer "seed" value U_0 (for example, see [8] pages 377-388). This method will necessarily repeat, but the constants a and T can be selected in such a manner to give a period on the order of T and a white spectrum.

The EBT will be demonstrated on an example with $n = 256$ samples from the chaotic process with parameters $a = 31623$ and $T = 2^{16} - 1 = 65535$. The time series realization ($U_0 = 14349$) normalized to give uniform deviates on the unit interval along with its spectrum is displayed in Figure 3 along with its spectrum. With $N = 256$ equally spaced intervals, there are $M_0 = 84$ empty boxes which gives a statistically significant z-value of -2.0024 ($P = .0226$). The other statistics have values of $M_1 = 106, M_2 = 49, M_3 = 16$, and $M_4 = 1$ which leads to less significant results of $D_1 = -4.4754$ ($P = .0673$) and $T = -65.8104$ ($P = .0881$) but still provide some evidence that the sequence is not random. Another pseudo-random generator that has been discussed extensively in [1] and [6] has the parameters $a = 16807$ and $T = 2^{31} - 1 = 2147483647$.

3.2. Kakutani-von Neuman Map

In this section we look at a minefield generated by a variant of the Kakutani / von Neuman map shown in Figure 4 which we will denote by the function K. The map K is an invertible, measure preserving map of the unit inter-

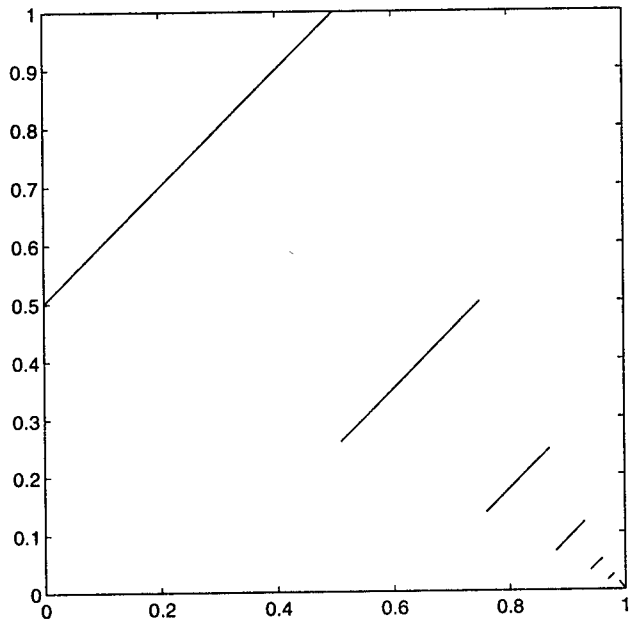


Figure 4. Kakutani / von Neuman Map

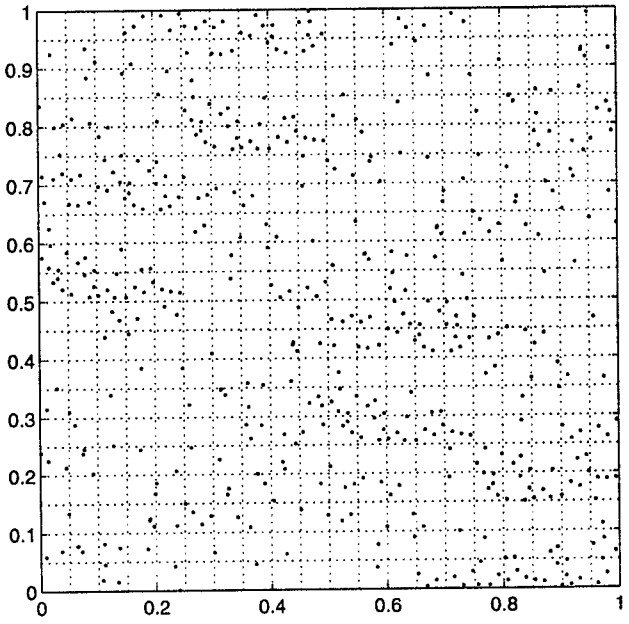


Figure 5. Chaotic minefield with 500 Points using a perturbation of the Kakutani / von Neuman Map.

val with a derivative of 1 (almost everywhere with respect to Lebesgue measure) that is weak mixing but not strong mixing (see [7] for details).

The (x, y) locations for the points in Figure 5 were generated by

$$x_{k+1} = K(x_k) \quad y_{k+1} = K(y_k + (y_k + x_k)^7/128) \quad (15)$$

The unit square was partitioned into N=400 sections to use the EBT. The $M_0 = 108$ empty boxes are significantly less than expected under CSR (z-value=-5). In this case, the CSR hypothesis is rejected for a tendency to cluster. However, there are clearly regularities and periodicities of this "minefield" that could be exploited as well.

4. Conclusions

The empty boxes test and its extensions offer simple, flexible, and robust approaches to detecting regularity in point processes. These methods are particularly applicable to the problem of characterizing the difference between random and chaotic processes as was demonstrated on some nontrivial examples.

5. Acknowledgements

Ray Brown of Applied Chaos Technologies provided the Kakutani / von Neumann example.

References

[1] R. Brown and L. Chua. Clarifying Chaos: Examples and Counterexamples. *to appear in International Journal of Bifurcation and Chaos*, 1996.

[2] N. Cressie. *Statistics for Spatial Data.* Wiley, New York, 1991.

[3] N. Johnson and S. Kotz. *Urn Models and their Applications.* Wiley, New York, 1977.

[4] D. E. Lake. Detecting regularity in point processes generated by humans. Proceedings 1994 IEEE-IMS Workshop on Information Theory and Statistics, IEEE Service Center, Piscataway, NJ, 1994.

[5] D. E. Lake and D. Keenan. Identifying minefields in clutter via collinearity and regularity detection. in Detection Technologies for Mines and Minelike Targets, Abinash C. Dubey, Ivan Cindrich, James M Ralson,Kelly Rigano, Editors, Proc. SPIE-2496, 519-530, 1995.

[6] S. Park and K. Miller. Random Number Generators: Good Ones and Hard to Find. /em Communications of the ACM 31(10):1192-1201, 1988.

[7] W. Parry. *Topics in Ergodic Theory.* Cambridge Unviversity Press, Cambridge, Great Britain, 1981.

[8] D. Phillips, A. Ravindran, and J. Solberg. *Operations Research: Principles and Practice.* John Wiley and Sons, New York, 1976.

An Iterative Solution to the Min-Max Simultaneous Detection and Estimation Problem

Bülent Baygün
Schlumberger-Doll Research
Old Quarry Road
Ridgefield, CT 06812
baygun@ridgefield.sdr.slb.com

Alfred O. Hero
University of Michigan
Department of EECS
Ann Arbor, MI 48109
hero@eecs.umich.edu

Abstract

Min-max simultaneous signal detection and parameter estimation requires the solution to a nonlinear optimization problem. Under certain conditions, the solution can be obtained by equalizing the probabilities of correctly estimating the signal parameter over the parameter range. We present an iterative algorithm based on Newton's root finding method to solve the nonlinear min-max optimization problem through explicit use of the equalization criterion. The proposed iterative algorithm does not require prior proof of whether an equalizer rule exists: convergence of the algorithm implies existence. A theoretical study of algorithm convergence is followed by an amplitude estimation example which shows that decoupling detection from estimation entails a very significant loss in estimation performance even when optimal decoupled decision rules rules are implemented.

1. Introduction

In practical applications, one frequently needs to design a signal detector or a signal parameter estimator without complete knowledge of the signal or noise model. Several approaches to detector and estimator design exist in the case of incompletely characterized models. Among these are invariance methods, Bayesian methods which use non-informative priors, and min-max methods. Min-max methods form an important solution category because they ensure optimal detector or estimator performance under worst case conditions. Furthermore, min-max solutions give rise to tight performance bounds which can be used to benchmark sub-optimal or ad hoc algorithms. Min-max methods have been applied to problems of adaptive array processing, harmonic retrieval, CFAR detection, and distributed detection.

Signal detection and signal parameter estimation are typically considered as separate problems. In other words, signal parameter estimation methods assume that there is no uncertainty about signal presence. However, there are many applications where signal parameter estimation has to be done under signal presence uncertainty, such as fault detection and estimation in dynamical system control and antenna array processing. Such problems are refered to as *simultaneous detection and estimation* problems. A min-max solution to simultaneous detection and estimation was recently given in [2]. The problem considered in [2] is estimation of a discrete parameter under a false alarm constraint. The statistical decision procedure which solves the problem is called the *constrained min-max classifier*. The constrained min-max classifier is characterized by a set of optimal weights. In Bayesian terminology, the optimal weights represent a *least favorable distribution* on the unknown parameter values. Numerical solutions to min-max detection or estimation problems involve nonlinear optimization to obtain the least favorable distribution [3, 1]. On the other hand, under certain assumptions, it is possible to formulate a min-max solution by making explicit use of a simplifying sufficient condition for min-max optimality. In the case of the constrained min-max classifier, this sufficient condition is the equalization of the correct classification probabilities. The purpose of the present work is to present an iterative algorithm for efficiently computing the constrained min-max classifier through the equalization condition. An important attribute of the proposed iterative algorithm is that it does not require prior proof of existence of an equalizer rule. Convergence of the algorithm proves existence, i.e. if we observe convergence, then the associated solution is the

0-8186-7576-4/96 $5.00

constrained min-max classifier.

The correct classification probability of the constrained min-max classifier provides a tight lower bound on the correct classification probability of any similarly constrained detection and classification procedure. By using the proposed algorithm, we can compute both this lower bound and the classification performance of sub-optimal simultaneous detection and classification procedures. Comparison of the performance of sub-optimal procedures with the lower bound allows us to assess the performance loss incurred by employing a sub-optimal approach to simultaneous detection and classification.

2. Problem Formulation

Consider the indexed probability space (Ω, σ, P_μ), where μ is a parameter that lies in a finite discrete parameter space Ξ, σ is a sigma algebra over Ω and P_μ is a probability measure defined on σ. Let $\mathbf{X}$ be a random variable taking values in a sample space Ω. Assume that $\mathbf{X}$ has a probability density function $f_\mu(x)$ with respect to a given measure. We will illustrate our approach for the case of a location parameter, i.e. $f_\mu(x) = f(x-\mu)$ for some fixed probability density function f. Applications of the location parameter case include modeling of a signal of unknown amplitude μ in additive noise whose probability density function is given by f.

Define the hypotheses $H_0, H_1, \ldots, H_n$ by:

$$H_i: \quad \mathbf{X} \sim f_{\mu_i}(x) = f(x-\mu_i)\ , \quad i = 0, \ldots, n \qquad (1)$$

Let $R_0, R_1, \ldots, R_n$ be the decision regions for hypotheses $H_0, H_1, \ldots, H_n$, respectively, i.e. the classifier declares $\mu = \mu_i$ if and only if $x \in R_i$, $i = 0, 1, \ldots, n$. The probability of a correct decision under hypothesis H_i, $i = 0, 1, \ldots, n$ is given by

$$\mathrm{P}_{\mu_i}(\text{decide } H_i) \;=\; \mathrm{P}_{\mu_i}(\mathbf{X} \in R_i) \qquad (2)$$

We will be interested in choosing the decision regions $R_0, R_1, \ldots, R_n$ such that the worst case correct classification probability $\min_i \mathrm{P}_{\mu_i}(\text{decide } H_i)$ is maximized subject to a given upper bound $\alpha \in (0, 1]$ on the false alarm probability $1 - \mathrm{P}_{\mu_0}(\text{decide } H_0)$. A decision rule which maximizes the worst case correct classification probability under a false alarm constraint is called a *constrained min-max classifier*. In [2] it was shown that the constrained min-max classifier is a weighted likelihood ratio test:

$$\max_{i>0} \left\{ c_i \frac{f_{\mu_i}(x)}{f_{\mu_0}(x)} \right\} \underset{H_0}{\overset{H_{i_{max}}}{\gtrless}} \gamma\ , \qquad (3)$$

i.e. if the maximum weighted likelihood ratio exceeds the threshold γ, then decide $H_{i_{max}}$, where $i_{max} = \arg\max_{i>0} \{c_i f_{\mu_i}(x)/f_{\mu_0}(x)\}$; otherwise decide H_0. The weights $c_1, \ldots, c_N$ are computed as the solution to a nonlinear optimization problem:

$$\min_{c_1,\ldots,c_n} \sum_{i=1}^{n} c_i \mathrm{P}_{\mu_i}(\text{decide } H_i)\ . \qquad (4)$$

The threshold γ is determined using the specified bound α. Solution of the nonlinear optimization problem (4) could be computationally expensive. We will outline an alternative solution scheme which characterizes the min-max optimal classifier by means of a sufficient condition.

Suppose that the parameterized density $f_\mu(x) = f(x-\mu)$ has infinite support ($f(x) > 0$ for all x) and has a monotone likelihood ratio. The infinite support assumption is made to simplify the discussion of algorithm convergence. Infinite support is not absolutely necessary for the algorithm to work. An important class of probability densities that satisfies the monotone likelihood property is the single parameter exponential family. Furthermore, a sufficient condition for $f(x-\mu)$ to have a monotone likelihood ratio is for the function $-\log f(x)$ to be convex in x [4, page 509]. The normal, the double exponential and the logistic distributions all satisfy the convexity condition. Under the monotone likelihood ratio assumption, it can be shown that the constrained min-max classifier (3) gives rise to the following decision regions $R_0, R_1, \ldots, R_n$:

$$\begin{aligned} R_0 &= (-\infty, x_0]; \\ R_i &= (x_{i-1}, x_i], \quad i = 1, \ldots, n-1; \\ R_n &= (x_{n-1}, \infty) \end{aligned} \qquad (5)$$

The correct decision probabilities are given by:

$$\begin{aligned} \mathrm{P}_{\mu_0}(\mathbf{X} \in R_0) &= F(x_0 - \mu_0) \\ \mathrm{P}_{\mu_1}(\mathbf{X} \in R_1) &= F(x_1 - \mu_1) - F(x_0 - \mu_1) \\ &\vdots \\ \mathrm{P}_{\mu_n}(\mathbf{X} \in R_n) &= 1 - F(x_{n-1} - \mu_n) \end{aligned} \qquad (6)$$

where F is the cumulative distribution function with density f. The acceptance region R_0 for the null hypothesis H_0 can be specified explicitly. For any given value of $\alpha \in (0, 1]$, there exists a value of x_0 that satisfies the false alarm constraint: $x_0 = F^{-1}(1-\alpha) + m_0$. The remaining decision boundary values $x_1, \ldots, x_{n-1}$ will be computed by an iterative procedure.

A sufficient condition for min-max optimality is the equalization of the correct classification probabilities $\mathrm{P}_{\mu_i}(\text{decide } H_i)$ for $i = 1, \ldots, n$ [2, Corollary 2]. The

equalization condition is represented by the set of equations

$$\mathrm{P}_{\mu_i}(\text{decide } H_i) = p, \quad i = 1, \ldots, n \tag{7}$$

where $p \in (0,1)$ is the unknown common value of the correct classification probabilities. Let $\underline{y} = [x_1, \ldots, x_{n-1}, p]^T$ ("T" denotes matrix transpose) and define the function $G(\underline{y})$ as follows.

$$G(\underline{y}) \stackrel{def}{=} \begin{bmatrix} F(x_1 - \mu_1) - F(x_0 - \mu_1) - p \\ F(x_2 - \mu_2) - F(x_1 - \mu_2) - p \\ \vdots \\ F(x_{n-1} - \mu_{n-1}) - F(x_{n-2} - \mu_{n-1}) - p \\ 1 - F(x_{n-1} - \mu_n) - p \end{bmatrix} . \tag{8}$$

Then the set of equations (7) is equivalent to

$$G(\underline{y}) = [0, \ldots, 0]^T \tag{9}$$

We propose to solve (9) iteratively using Newton's root finding method. More specifically, we consider the sequence $\underline{y}(k)$ generated through the iterations

$$\underline{y}(k+1) = \underline{y}(k) - J^{-1}(\underline{y}(k))G(\underline{y}(k)) , \tag{10}$$

where $J(\underline{y})$ is the Jacobian of the function $G(\underline{y})$, i.e.

$$[J(\underline{y})]_{ij} \stackrel{def}{=} \frac{\partial [G(\underline{y})]_i}{\partial y_j} . \tag{11}$$

For $j = 1, \ldots, n-1$, $y_j = x_j$ and $y_n = p$. Therefore, the elements in the first $n-1$ columns of $J(\underline{y})$ are found from (8):

$$[J(\underline{y})]_{ij} = \begin{cases} f(x_j - \mu_j) & \text{, if } i = j \\ -f(x_{j+1} - \mu_j) & \text{, if } i = j+1 \\ -1 & \text{, } j = n \\ 0 & \text{, otherwise} \end{cases} \tag{12}$$

A few words about the convergence of the iterative algorithm (10) are in order. Assume that there exists a solution $\underline{y}^*$ to the equation (9). If

1. $J^{-1}(\underline{y}^*)$ exists (the Jacobian is invertible); and
2. $||J(\underline{y}^* + \underline{\delta y}) - J(\underline{y}^*)|| \leq \gamma ||\underline{\delta y}||$ for some $\gamma > 0$ and for all sufficiently small perturbations $\underline{\delta y}$ (J is Lipschitz continuous); and
3. $||J^{-1}(\underline{y}^*)|| \leq \beta$ for some $\beta > 0$ (the norm of the Jacobian inverse is bounded from above);

then for any starting point $\underline{y}(0)$ that is sufficiently close to $\underline{y}^*$, the sequence $\underline{y}(k)$ generated through (10) is well-defined, converges locally to $\underline{y}^*$ and has a quadratic rate of convergence with coefficient $\gamma\beta$ [5, Theorem 5.2.1]. Next we provide a sketch of the proof that the three conditions are satisfied in the present problem.

Condition 1: Since $f(x) > 0$ for all x, the columns of J are linearly independent.

Condition 2: The non-zero elements of the difference δJ of two Jacobians evaluated at points $\underline{y} + \underline{\delta y}$ and $\underline{y}$, respectively, are of the form $\pm(f(x_i + \delta x_i - \mu_j) - f(x_i - \mu_j))$. But $f(x_i + \delta x_i - \mu_j) - f(x_i - \mu_j) = \int_{x_i}^{x_i + \delta x_i} f'(t - \mu_j)dt$. Assuming that the derivative f' of the probability density function f is bounded, i.e. $\sup_x |f(x)| \leq M$ for some $M > 0$, it follows that $|f(x_i + \delta x_i - \mu_j) - f(x_i - \mu_j)| \leq M|\delta x_i|$. It can then be shown that the Frobenius norm of δJ, denoted by $||\delta J||_F$ is bounded above by a multiple of the l_2 norm of the vector $\underline{\delta y}$. Since the l_2-induced norm of δJ is smaller than the Frobenius norm of δJ [5, Theorem 3.1.3], Lipschitz continuity is satisfied.

Condition 3: For arbitrary $\underline{z} = [z_1, \ldots, z_n]^T$, consider the linear equation

$$J(\underline{y}(k))\underline{y}(k+1) = \underline{z} . \tag{13}$$

For notational simplicity, we will write the Jacobian as J and suppress its dependence on $\underline{y}$. After Gaussian elimination, the equation (13) can be re-written in terms of an upper triangular matrix $\tilde{J}$: $\tilde{J}\underline{y}(k+1) = \tilde{\underline{z}}$. The matrices $\tilde{J}$ and J are related by a non-singular transformation T, i.e. $\tilde{J} = TJ$. It suffices to establish an upper bound on the Frobenius norm $||\tilde{J}^{-1}||_F$ of $\tilde{J}^{-1}$ because $||J^{-1}||_F$ and $||\tilde{J}^{-1}||_F$ are related by $||J^{-1}||_F \leq ||T||_F||\tilde{J}^{-1}||_F$ and $||T||_F$ is bounded. Suppose that the last column of $\tilde{J}$ is the vector $[-a_1, \ldots, -a_n]^T$, i.e. $[\tilde{J}]_{in} = -a_i$, $i = 1, \ldots, n$. It can be shown that $a_1 = 1$ and $a_i = 1 + a_{i-1}\frac{f(x_{i-1} - \mu_i)}{f(x_{i-1} - \mu_{i-1})}$, $i = 2, \ldots, n$. The Frobenius norm of $\tilde{J}^{-1}$ can be expressed as: $||\tilde{J}^{-1}||_F = [\mathrm{tr}((\tilde{J}^{-1})^T \tilde{J}^{-1})]^{1/2}$, where "tr" denotes matrix trace. After some algebra, we obtain an upper bound:

$$\begin{aligned} ||\tilde{J}^{-1}||_F &= \left(\sum_{i=1}^{n-1} (1 + \frac{a_i^2}{a_n^2}) \frac{1}{f^2(x_i - \mu_i)} + \frac{1}{a_n^2} \right)^{1/2} \\ &\leq ((n-1)L + 1)^{1/2} , \end{aligned} \tag{14}$$

where $L = \max_i \{(a_i^2 + a_n^2)/f^2(x_i - \mu_i)\}$, $i = 1, \ldots, n-1$. In finite dimensional spaces all norms are equivalent, therefore there exists some $\beta > 0$ such that $||J|| \leq \beta$.

3. Applications on Simultaneous Detection and Classification in Gaussian Noise

We will illustrate the iterative algorithm (10) for the case of normal densities. Let $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2}{2\sigma^2})$ and $\mu_i = i$ for $i = 0, 1, \ldots, n$. We consider three different simultaneous detection and estimation rules. One of the rules is the constrained min-max classifier described earlier, which maximizes the worst case classification performance under a given false alarm constraint. One can also perform simultaneous detection and estimation by combining a classifier with a separately designed detector. With this strategy, the data are not presented to the classifier unless the detector declares "signal present". In other words, the classifier is gated by the detector.

We consider two gated classifiers and compare their performance to the performance of the constrained min-max classifier. Both of the gated classifiers use a min-max optimal detector for detection, but they differ in the design of their classifier structures. One of them uses an unconstrained min-max classifier designed independently of any detection objective. An unconstrained min-max classifier maximizes the worst case correct classification probability as if signal presence is certain. This classifier is obtained by removing the false alarm constraint ($\alpha = 1$) in the constrained min-max classifier. The other gated classifier uses a conditionally min-max classifier designed with explicit knowledge of the detector decision regions. A conditionally min-max optimal classifier maximizes the worst case correct classification probability conditioned on the detector having declared signal present. The conditionally min-max classifier is obtained by replacing all the densities $f_{\mu_i}(x)$ under the alternative hypotheses $H_1, \ldots, H_n$ with the conditional densities $f_{\mu_i}(x|\mathbf{X} \notin R_0)$ in the analysis of Section 2. Since we are using the min-max detector, $R_0 = (-\infty, x_0]$ as before, and x_0 is specified by the false alarm probability α.

Figure 1 shows the variation of the worst case correct classification probability $\min_i \mathrm{P}_{\mu_i}(\text{decide } H_i)$ for the three simultaneous detection and estimation rules as a function of the false alarm probability α. In this example $\sigma = 0.6$, and there are five alternative hypotheses ($n = 5$). In general, the constrained min-max classifier (solid line) performs best, while the unconstrained min-max classifier gated by the min-max detector (dashed line) gives rise to the lowest performance. The conditionally min-max classifier gated by the min-max detector (dashdot line), although better than the unconstrained min-max classifier, still

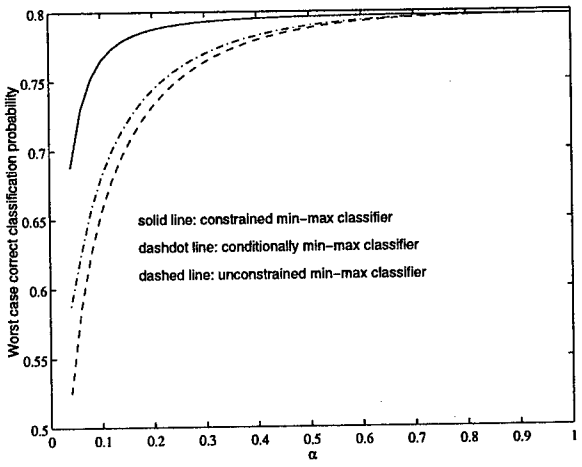


Figure 1. Worst case correct classification probability as a function of α.

falls significantly short of the performance of the constrained min-max classifier for small α. On the other hand, as α increases all three curves come together as expected. This is because for high α, the three simultaneous detection and estimation rules degenerate to an unconstrained min-max classifier for the alternative hypotheses $H_1, \ldots, H_n$.

References

[1] J. C. Preisig. A minmax approach to adaptive matched field processing in an uncertain propagation environment. *IEEE Trans. on Signal Proc.*, 42(6):1305–1316, June 1994.

[2] B. Baygün and A. O. Hero. Optimal simultaneous detection and estimation under a false alarm constraint. *IEEE Trans. Info. Theory*, 41(3):688-703, May 1995.

[3] C.-I. Chang and L. D. Davisson. Two iterative algorithms for finding minimax solutions. *IEEE Trans. Info. Theory*, 36(1):126–140, January 1990.

[4] E. L. Lehmann. *Testing Statistical Hypotheses.* 2nd ed., Wadsworth & Brooks/Cole, Pacific Grove, 1991.

[5] J. E. Dennis, Jr. and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice Hall, Englewood Cliffs, 1983.

An Optimum Permutation Test for Nonparametric Radar Detection

Francisco Alvarez-Vaquero and José L. Sanz-González
Departamento de Señales, Sistemas y Radiocomunicaciones,
ETSI de Telecomunicación, Universidad Politécnica de Madrid,
Ciudad Universitaria s/n, 28040 Madrid, Spain
E-Mail: fav@gc.ssr.upm.es

Abstract

A hypothesis H is parametric if every distribution from the process defined by H belongs to a family of distributions characterized by a finite number of parameters; on the other hand, if the distribution can not be definided by a finite number of parameters, the hypothesis is nonparametric.

In this paper, we analyze a detector based on the optimum permutation test, in the Neyman-Pearson sense, and under Gaussian noise conditions, which operates on radar video signal. The computational complexity of the detector is high and its implementation in real time is difficult, due to the number of operations increases with the factorial of the number of samples. Also, we present an algorithm that reduces the computational work required.

We also present the characteristic of detectability of the optimum permutation test under Gaussian noise environments and different types of target models (nonfluctuating, Swerling I and Swerling II). The detection probability versus signal-to-noise ratio is estimated by Monte-Carlo simulations for different parameter values (N pulse, M reference samples and false alarm probability P_{fa}).

1.-Introduction.

There are many posibilities to solve radar detection problems by means of nonparametric tests, which do not have a global solution. We are interested in the class of binary nonparametric tests called permutation tests, which are distribution-free under independent and identically distributed (IID) samples.

The distribution of a block of IID samples is invariant under the permutation of its sample components. That is, consider a IID sample vector $(x_1,x_2,...,x_n)$ of n samples where $F_0(x)$ is the distribution function of a sample, if $F(x_1,x_2,...,x_n)= F_0(x_1)\cdot F_0(x_2)...F_0(x_n)$, then
$F(x_1,x_2,...,x_n)=F(x_2,x_1,...,x_n)=..^{(n!)}..=F(x_n,...,x_2,x_1)$

To generate a permutation test the sample space R^n is partitioned into $n!$ regions D_i $(i=1,2,...,n!)$ where

$$D_i = \{ \boldsymbol{x}=(x_1, ..., x_n) : if \boldsymbol{x} \in D_i , \boldsymbol{x}\ (permutation) \notin D_i\}$$

in such a way that

$$D_i \cap D_j=\varnothing \quad i,j=1,2,...,n! \quad i\neq j, \text{ and } \bigcup_{i=1}^{n!} D_i=R^n$$

Each sample vector $\boldsymbol{x}$ belongs to one of these regions D_i $(i=1,2,...,n!)$, and we can get a different vector by permuting their components, each one belonging to one different region D_i. It is possible to partition R^n-space in different ways in order to fulfil D_i-conditions. A particular case is the well known rank test [1,2,3], whose regions D_i are

$$D_i=\left\{ \boldsymbol{x}=(x_1,x_2,...,x_n): \quad x_{i_1}<x_{i_2}<\cdots<x_{i_n} \right\}$$

with $i_j \in \{1,2,...,n\}$, $i_j \neq i_k$ when $j\neq k$, $j,k=1,2,...,n$

Under the *null hypothesis* H_0 (target absent), the probability that the sample vector $\boldsymbol{x}$, belongs to one of the regions D_i is $1/n!$, i.e. $Prob\{\boldsymbol{x}\in D_i\} = 1/n!$

Under the *alternative hypothesis* H_1 (target present), there are D_i-regions with more probability measure than other ones and now the probability that $\boldsymbol{x} \in D_i$ $(i=1,...,n!)$ is not uniform.

Given a D_i-partition, we define the decision region as the union of K regions D_i. In order to get the maximum probability of detection, we select the D_i-regions with largest probabilities. Just under H_0, the false alarm probability P_{fa} *is* $K/n!$, where K is the number of D_i-regions. The optimum permutation test would be the partition that achieves a maximum detection probability.

In radar applications, we have N sample vectors $\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_N$ where N is the number of pulses per antenna beamwidth. Each sample vector $\boldsymbol{x}_i$ has M noise reference samples x_{ij}, $j=1,2,...,M$ and the sample under test x_i, i.e. $\boldsymbol{x}_i = (x_{i1},x_{i2},...x_{iM},x_i)$. Under the null hypothesis H_0 (target absent) we suppose that the components of $\boldsymbol{x}_i$ are IID, but under the alternative hypothesis H_1 (target present) they are not IID (reference samples $x_{i1},x_{i2},...,x_{iM}$ are IID and x_i has different distribution of x_{ij}, $j=1,2,...,M$).

0-8186-7576-4/96 $5.00

Now, the distributions associated with H_0 are

$$f(\mathbf{x}/H_0)=\prod_{i=1}^{N}\prod_{j=1}^{M}[f_{0i}(x_{ij})]\cdot f_{0i}(x_i) \qquad (1a)$$

where $f_{oi}(\cdot)$ is the probability density function of a noise sample in the *ith*-pulse.

Under H_1, we have

$$f(\mathbf{x}/H_1)=\prod_{i=1}^{N}[\prod_{j=1}^{M}f_{0i}(x_{ij})]\cdot f_{1i}(x_i) \qquad (1b)$$

where $f_{1i}(x_i)$ is the probability density function of a sample under test x_i (signal + noise) in the *ith*-pulse.

2.-Permutation Test Algorithm.

In order to test H_o against H_1 in Neyman-Pearson sense, we take the likelihood ratio

$$\frac{f(\mathbf{x}/H_1)}{f(\mathbf{x}/H_0)}=\frac{\prod_{i=1}^{N}[\prod_{j=1}^{M}f_{0i}(x_{ij})]\,f_{1i}(x_i)}{\prod_{i=1}^{N}[\prod_{j=1}^{M}f_{0i}(x_{ij})]\,f_{0i}(x_i)}=\prod_{i=1}^{N}\frac{f_{1i}(x_i)}{f_{oi}(x_i)} \qquad (2)$$

In case of Gaussian noise conditions and nonfluctuation target models, applying (2) at the output of linear envelope detector, we have (after taking Neperian logarithm):

$$Ln\frac{f(\mathbf{x}/H_1)}{f(\mathbf{x}/H_0)}=\sum_{i=1}^{N}LnI_0(x\sqrt{2S})+(-NS) \qquad (3)$$

where S is the signal-to-noise ratio (SNR), and $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero.

(a) If signal-to-noise ratio (SNR) is low

$$Ln\frac{f(\mathbf{x}/H_1)}{f(\mathbf{x}/H_0)}\sim\sum_{i=1}^{N}x_i^2 \qquad (4a)$$

(b) If SNR is high

$$Ln\frac{f(\mathbf{x}/H_1)}{f(\mathbf{x}/H_0)}\sim\sum_{i=1}^{N}|x_i| \qquad (4b)$$

We optimize the permutation test using (2) or (3), by permuting all the samples in each vector $\mathbf{x}_i=(x_{i1},x_{i2},...x_{iM},x_i)$, $i=1,2,...,N$ and selecting the upper results in (4). The number of K higher results selected depends on the false alarm probability P_{fa}, i.e. $P_{fa}=K/(M+1)^N$, where K is the number of D_i-regions associated with upper results of (3) after doing permutations.

We optimize the permutation test using (3) or (4) in the following way, from $i=1$ *to* N we have the matrix (for application of (4a)).

$$\begin{bmatrix} x_{11}^2 & x_{12}^2 & \cdots & x_{1M}^2 & x_1^2 \\ x_{21}^2 & x_{22}^2 & \cdots & x_{2M}^2 & x_2^2 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ x_{i1}^2 & x_{i2}^2 & \cdots & x_{iM}^2 & x_i^2 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ x_{N1}^2 & x_{N2}^2 & \cdots & x_{NM}^2 & x_N^2 \end{bmatrix} \qquad (5)$$

adding the elements of the right column, we have

$$y=\sum_{i=1}^{N}x_i^2 \qquad (6)$$

Now permuting the components in each vector (row vector) in (5) and summing by columns, and ordering these $(M+1)^N$ summs from the lower to the upper, we get the set of Kth-greatest summs. If (6) is in this set, it is supposed target present (hypothesis H_1); otherwise, it is supposed that target is absent (hypothesis H_0).

An efficient algorithm is as follows. First, in (5) we order from the lowest to the highest the components of each row vector, obtaining the matrix (7):

$$\begin{bmatrix} z_{11}^2 & z_{12}^2 & \cdots & z_{1M}^2 & z_{1M+1}^2 \\ z_{21}^2 & z_{22}^2 & \cdots & z_{2M}^2 & z_{2M+1}^2 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ z_{i1}^2 & x_{i2}^2 & \cdots & z_{iM}^2 & z_{iM+1}^2 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ z_{N1}^2 & z_{N2}^2 & \cdots & z_{NM}^2 & z_{NM+1}^2 \end{bmatrix} \qquad (7)$$

$$where \quad z_{i1}^2<z_{i2}^2<\cdots<z_{iM}^2<z_{iM+1}^2$$

we get

$$y_1^{(M+1)} = \sum_{i=1}^{N} z_{i\,M+1}^2 \qquad (8)$$

Note that (8) is the upper value. Now, swapping z_{1M} and z_{1M+1}, and summing again the new right column, we get the next value $y_1^{(M)}$, and so on in order to obtain the K upper values; so we have $y_1^{(C)}$ $(C=M+1,M,\cdots,1)$ In each step we compared the $y_1^{(C)}$ with the y value of (6), if $y_1^{(C)} < y$, stop the process with the first row and go to the second row-vector of (7). So in (7) we swapp z_{2M} and z_{2M+1} in order to get $y_2^{(M)}$, and so on. We repeat this algoritm in order to know if the y belongs or not to the K uppers values. If we get K upper values $y_r^{(C)} > y$ where ($1<C<M$ and $1<r<N$), it is not necessary continue the process, testing the N rows and doing the M+1 swapping in each row; in this case we supposed that the target is absent (hypothesis H_0).

3.- Computer results

We have analyzed the detection performance of permutation tests in terms of detection probability P_d with constant false alarm probability P_{fa}, considering (4a), as the statistic for the implementation of algorithm described.

For a particular target model, the detection probability P_d is a function of SNR, P_{fa}, N, and M. We have considered $P_{fa}=10^{-6}$ and 10^{-8} as practical radar values. We present in the Figures 1, 2 and 3, P_d-curves with M =6 and N=10 and 12 for different types of targets (Swerling II, Swerling I and nonfluctuating). As it can be seen, we obtain a important variation in P_d for a low difference in N-values. Also, it is observed that as P_{fa} decreases then the diference between P_d curves increase.

The Figures 4, 5 and 6 show P_d-curves for N=8 and M= 10 and 16. The variation in P_d with N is more important than the variation with M, and this fact is because the integrate pulses convey more information than the noise reference samples. Also, from Figures 3 and 6, we can see very large differences in SNR for P_d-curves of N=10 and 12 when $P_{fa}=10^{-8}$. More reseach worh is required about this fact.

Finally, due to $P_{fa} = K/(M+1)^N$, the computational complexity of the permutation test algorithm increases with M and N values for a specific P_{fa}-values, because K increases. Consequently an optimization process is required for the best determination of N, M and SNR in practical applications.

Other results about optimum parametric and rank detectors against permutation test will be published elsewhere. Differences up to 1 dB in SNR are found between rank test and permutation test for the same P_d, P_{fa}, N and M.

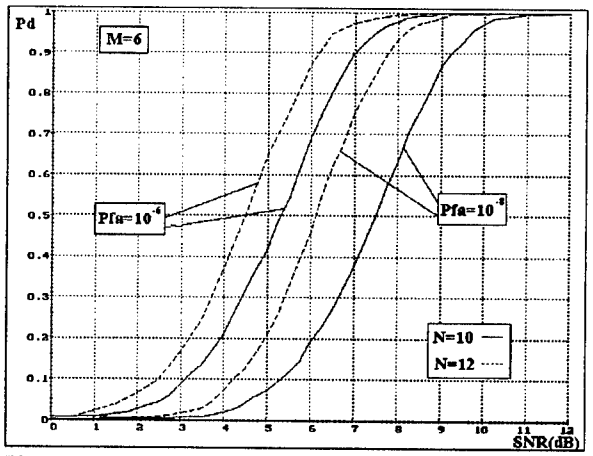


Fig.1: Detection Probablity P_d versus Signal-to-Noise (SNR) for permutation test with, M=6, N=10 and N=12 with false alarma probaility $P_{fa}=10^{-6}$ and $P_{fa}=10^{-8}$, for Nonfluctuating target model.

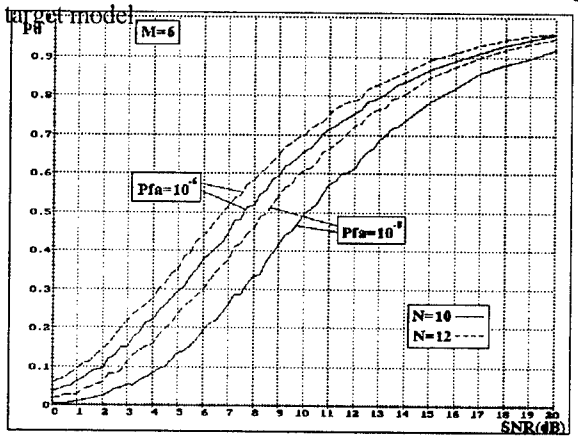


Fig.2: Detection probability P_d versus Signal-to-Noise (SNR) for permutation test, with M=6 ,N=10 and N=12 with false alarma probability $P_{fa}= 10^{-6}$ and $P_{fa} = 10^{-8}$, for Swerling I target model.

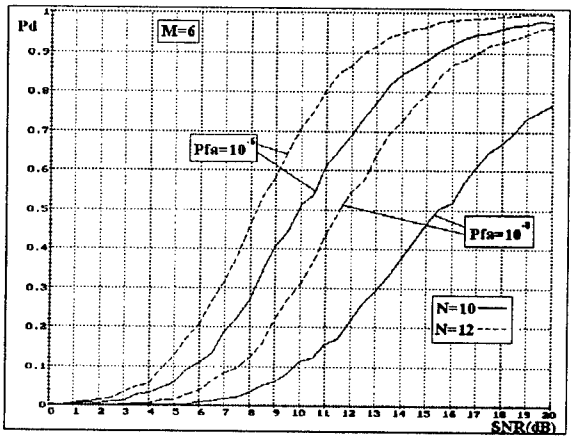


Fig.3: Detection probability P_d versus Signal-to-Noise (SNR) for permutation test. with M=6,N=10 and N=12 with false alarm probability $P_{fa}=10^{-6}$ and $P_{fa}=10^{-8}$, for Swerling II target model.

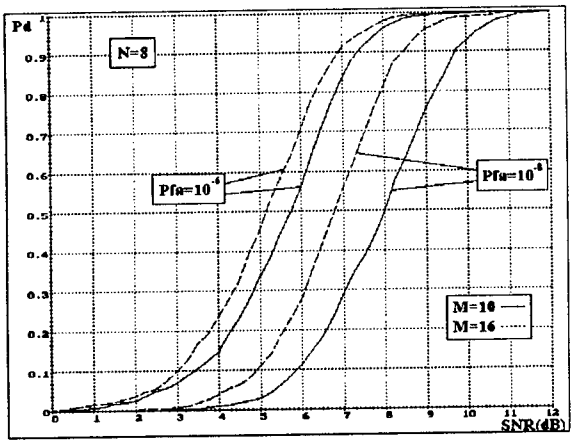


Fig.4: Detection probability P_d versus Signal-to-Noise (SNR) for permutation test, with N=8, M =10 and M=16 with false alarm probability $P_{fa}=10^{-6}$ and $P_{fa}=10^{-8}$, for Nonfluctuating target model..

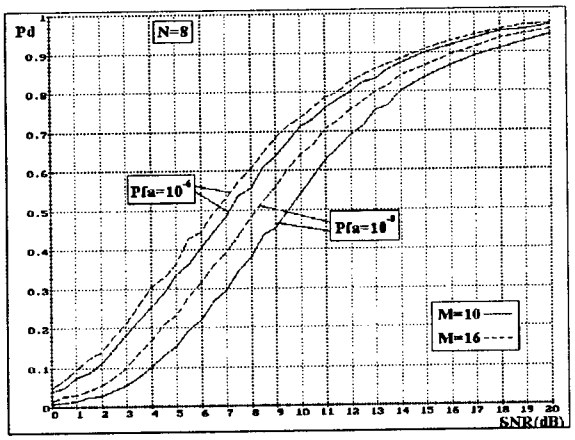


Fig.5: Detection probability P_d versus Signal-to-Noise (SNR) for permutation test, wiht N=8, M=10 and M=16 with false alarm probability $P_{fa}=10^{-6}$ and $P_{fa}=10^{-8}$, for Swerling I target model.

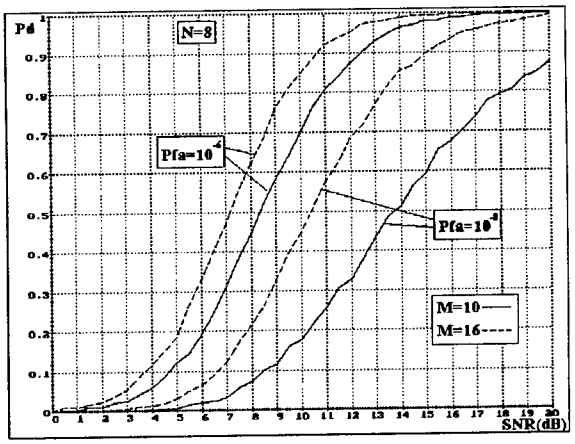


Fig.6: Detection probability P_d versus Signal-to-Noise (SNR) for permutation test, with N=8, M =10 and M= 16 with false alarm probability $P_{fa}=10^{-6}$ and $P_{fa}=10^{-8}$, for Swerling II target model.

REFERENCES.

[1] J. L. Sanz-González and A. R. Figueiras-Vidal, "A Suboptimum Rank Test for Nonparametric Radar Detection", IEEE Trans. Aerosp. Electronic Syst. Vol. AES-22, pp. 670-680, Nov. 1986.

[2] J. L. Sanz-González, "Nonparametric Rank Detectors on Quantized Radar Video Signals", IEEE Trans. Aerosp. Electronic Syst. Vol. AES-39, pp. 969-975, Nov. 1990.

[3] S. Y. Kim and I. Song, "On the Score Functions of the Two-Sample Locally Optimum Rank Test Statistic for Random Signals in Additive Noise", IEEE Trans. Inform. Theory, Vol. IT-41, pp. 842-846, May 1995.

A GENERALIZED LIKELIHOOD RATIO TEST DETECTOR FOR MOVING TARGETS IN CLUTTER

Ivars P. Kirsteins
Naval Undersea Warfare Center
Code 3124
New London, CT 06320 USA

Nikitas V. Nikitakos
National Technical University of Athens
Dept. of ECE - ICCS , 42, Patission Str.
10682 Athens, GREECE

ABSTRACT

A generalized likelihood-ratio test (GLRT) detector is derived for detecting a space-time signal in the presence of unknown subspace interference and unknown target doppler. The near optimality and constant false alarm rate (CFAR) property of the GLRT is shown by the relationship to the uniformly most powerful invariant (UMPI) test using a simple approximation. Examples are presented comparing the performance of the proposed detector against the UMPI test. The ROC curves indicate that the GLRT detector compares favorably to the UMPI detector.

1. INTRODUCTION

We start by reviewing the subspace interference model. Suppose we have an array of m sensors that are simultaneously sampled at time t_k and the outputs stacked into the vector $\mathbf{x}(t_k) = [x_1(t_k)x_2(t_k)\cdots x_m(t_k)]^T$. We say the interference is subspace if at any instant of time, it can be represented as

$$\mathbf{x}(t_k) = H\Theta_{t_k} = \sum_{n=1}^{r} \mathbf{h}_n \theta_n^{t_k} \tag{1}$$

where H is a $m \times r$ matrix whose columns generate the interference space and Θ_{t_k} is a $r \times 1$ vector of scale factors. The data vector $\mathbf{x}(t_k)$ is a linear combination of the columns of H, which remain fixed, i.e., do not change as a function of time. The only dependence on time is through Θ_{t_k}. The subspace model has wide application. Many type of interference components, e.g., clutter, can be represented using a subspace model (see Scharf [1] for an extensive treatise on subspace or reduced-rank modeling and [2]). For example, if the sensor outputs $\mathbf{x}(t_k) = [g_1 I(t-\tau_1)\ g_2 I(t-\tau_2)\ \cdots\ g_m I(t-\tau_m)]^T$ (where $I(t)$ is some interference time series) are time delay steered to align the interference wavefronts (output of the kth sensor is delayed by $\tau_k - \tau_1$), then $\mathbf{x}(t_k) = I(t_k)[g_1 g_2 \cdots g_m]^T$ where g_k is the gain of the kth sensor. Also, colored noise can be modeled as subspace where the subspace dimension is proportional to bandwith [1, 2]. Thus narrowband components can be represented using a low order subspace model. Another aspect of the subspace model is it inherently accounts for array calibration error, eg., gain errors.

1.1. Signal and interference model

Usually the received signal has undergone multipath distortion or time dispersion from the channel. Therefore the waveform received by the nth sensor is modeled as

$$\tilde{s}_n(t) = \sum_{k=1}^{L_n} c_k s\left(\frac{t - \tau_k - \delta\tau_k}{a}\right) \tag{2}$$

where $s(t)$ is the signal replica, τ_k is the time it takes for the signal to travel from the source to sensor 1 over path k, $\delta\tau_k$ is the inter-sensor propagation time delay measured relative to sensor 1, $a = (c+v)/(c-v)$ is the contraction/dilation of the signal(s) due to target/platform motion (where c and v are the propagation and relative target velocities, assumed to be the same for each multipath), and c_k is a scalar corresponding to the attenuation from the kth path. The snapshot of sensor outputs at time t_j is then

$$\begin{aligned}\tilde{\mathbf{s}}_j &= \sum_{k=1}^{L} c_k \left[s(t_j/a - \tau_k/a)\ \ s(t_j/a - \tau_k/a - \delta\tau_2/a) \right. \\ &\qquad \left. \cdots\ s(t_j/a - \tau_k/a - \delta\tau_m/a)\ \right]^T \end{aligned} \tag{3}$$

or, after substituting ζ_k^j for the vector of data samples due to the kth path,

$$\tilde{\mathbf{s}}_j = \sum_{k=1}^{L} c_k\, \zeta_k^j \tag{4}$$

A total of K snapshots of data are collected at times $\{t_1, t_2, \ldots, t_K\}$ and stacked into a matrix. The matrix corresponding to the signal component in the data is then

$$\mathcal{D}^a = [\tilde{\mathbf{s}}_1 \mid \tilde{\mathbf{s}}_2 \mid \cdots \mid \tilde{\mathbf{s}}_K] \tag{5}$$

and has the equivalent form

$$\mathcal{D}^a = \sum_{k=1}^{L} c_k\, \mathcal{D}_k^a \tag{6}$$

where $\mathcal{D}_k^a = [\zeta_k^1 \mid \zeta_k^2 \mid \cdots \mid \zeta_k^K]$.

Using the above signal representation, the received space-time data matrices are modeled as

$$\mathcal{H}_0 : X = H\theta + N \tag{7}$$

$$\mathcal{H}_1 : X = H\theta + \sum_{k=1}^{L} c_k\, \mathcal{D}_k^a + N \tag{8}$$

0-8186-7576-4/96 $5.00

under the signal absent and signal present hypotheses respectively where $\theta = [\Theta_1 \Theta_2 \cdots \Theta_2]^T$. The elements of the background noise matrix N are modeled as IID complex Gaussian distributed with zero-mean and variance σ^2. We now discuss the uniformly most powerful invariant (UMPI) test for this hypothesis testing problem.

1.2. UMPI test

We want to test the hypothesis that $\sum_k |c_k|^2 = 0$ (signal absent) or $\sum_k |c_k|^2 > 0$ (signal present). We assume that the interference subspace H and doppler a are known, but that the parameters θ, c_k, and σ^2 in (7) and (8) are unknown and deterministic, ie., can take on a range of values. In sonar and radar, it is usually difficult or impossible to determine distributions for the interference and signal parameters since the relevant scattering and channel physics are usually not known or at best, partially known. This type of detection problem is called a composite hypothesis testing problem [8].

It is difficult to find an optimum test when no probability density function is available for the unknown parameters [8, 4]. Ideally, we would like to construct an uniformly most powerful (UMP) test [4]. A problem is that UMP tests usually do not exist [8, 4].

In [3] it is argued that principles of invariance should be used to find the UMP test which is invariant to the unknown nuisance parameters (eg., noise variance, signal phase), known as the UMPI test. The motivation is that nuisance parameters are probably responsible for the nonexistence of the UMP test in the first place [3]. Also, in many applications the test should be invariant to nuisance parameters such as the background noise level, ie., a CFAR test. However, the UMPI test is also difficult to find and may not exist. An alternative approach frequently used is to form the likelihood-ratio and replace the unknown parameters by their maximum likelihood estimates [8]. This is called the generalized likelihood ratio test (GLRT).

Scharf [6] derived the GLRT for the related problem of detecting in a single data snapshot a subspace signal in the presence of subspace interference (when the subspace is known) and showed that it is the UMPI test. The space-time signal and interference models we have are analogous to the data model used by Scharf [6] if the matrices in (7) and (8) are vectorized (by stacking the matrix columns into a vector). Vectorizing (7) and (8) and applying the results of [6], the UMPI test is

$$\frac{\| P_{S'} \mathbf{x}' \|_F^2}{\| P_{S'}^{\perp} \mathbf{x}' \|_F^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda \qquad (9)$$

where the projection operators are given by $P_H^{\perp} = I - H(H^H H)^{-1} H^H$, $P_{S'} = S'(S'^H S')^{-1} S'^H$, and $P_{S'}^{\perp} = I - P_{S'}$. The vectors are $S' = [vec(P_H^{\perp} \mathcal{D}_1^a) \,|\, vec(P_H^{\perp} \mathcal{D}_2^a) \,|\, \cdots \,|\, vec(P_H^{\perp} \mathcal{D}_L^a)]$ and $\mathbf{x}' = vec(P_H^{\perp} X)$. The scalar λ is some threshold. The operator $vec(\cdot)$ takes a matrix and converts it to a vector representation by stacking the columns. The numerator of (9) can be intrepreted as the magnitude-squared output of a space-time matched filter (beamforming-matched filter processing) using as the replica the part of the signal(s) which remain after the interference has been nulled. This is then normalized by an estimate of the background noise variance given by the denominator of (9). Because (9) is invariant to scalings of the data matrix and rotations in the column space of H, it is the best possible CFAR detector.

Unfortunately, test (9) usually can not be implemented because the interference subspace matrix H is not known (e.g., H is a function of such things as the channel, direction of arrival, array geometry and sensor characteristics which are either unknown or at best, partially known) and the target multipath structure and doppler are also unknown. In previous work Kirsteins [5] proposed a GLRT detector for the above problem given that doppler is known. The intent here is to extend those results to the case when doppler is not known and determine the effect on performance. In the remainder of this paper we derive the GLRT for the above hypothesis testing problem assuming the interference subspace H and target doppler are both unknown and then discuss its relationship to the UMPI test and determine analytically the effect on performance when doppler is unknown. Finally, some numerical examples are presented comparing the performances of the GLRT and UMPI tests. We start by deriving the GLRT.

2. GLRT

A GLRT statistic for choosing between hypotheses (7) and (8) is

$$z_1 = \frac{\min_{H_0,\theta_0} \|X - H_0\theta_0\|_F^2}{\min_{H_1,\theta_1,a,c_1,\ldots,c_L} \|X - H_1\theta_1 - \sum_{k=1}^{L} c_k \mathcal{D}_k^a\|_F^2} \qquad (10)$$

where H, θ, a, c_k, and σ^2 have been treated as unknown. The GLRT statistic (10) is simply a ratio of fitting errors. The numerator is the error in fitting the matrix X by a rank r matrix and the denominator is the error in jointly fitting X by a rank r matrix and $\sum_{k=1}^{L} c_k \mathcal{D}_k^a$.

The numerator in (10) is easily evaluated using the singular value decomposition (SVD) of X as $\min_{H_0,\theta_0} \|X - H_0\theta_0\|_F^2 = \sum_{k=r+1}^{m} \gamma_k^2$ where γ_k^2 are the singular values of X. We need to evaluate the denominator of (10). Unfortunately, a direct solution is not available. We propose an iterative scheme to perform the minimization based on the criss-cross regressions method of Gabriel [7] for solving the weighted low rank approximation problem. Basically the idea is to linearize the optimization problem, for each hypothesized doppler a, by holding H constant and then minimizing with respect to only θ and the c_k. This is a standard linear least-squares fitting problem and is easy to solve. The procedure is then repeated, except that this time θ is replaced with its estimate from the previous step and the minimization now done with respect to H and the c_k. These steps are continued until convergence. The algorithm steps are summarized below:

a. *Initialization. Iteration counter k is set to zero $k = 0$. Select initial guess H_0.*

b. $k = k + 1$

c. *Holding H_{k-1} fixed, minimize with respect to only θ and c_k:*

$$\theta_k, \hat{c}_k = arg \min_{\theta, c_1, \ldots, c_L} \| X - H_{k-1}\theta - \sum_{k=1}^{L} c_k \mathcal{D}_k^a \|_F^2$$

d. *Holding θ_k fixed, minimize with respect to only H and c_k:*

$$H_k, \hat{c}_k = arg \min_{H, c_1, \ldots, c_L} \| X - H\theta_k - \sum_{k=1}^{L} c_k \mathcal{D}_k^a \|_F^2$$

e. *Check if converged. If not converged, go back to step b.*

The operator *arg* here means the solution to the minimization problem.

2.1. Relationship to UMPI test

We now discuss the relationship of the proposed GLRT to the optimum UMPI test. It was shown in [5] that when the signal $\sum_{k=1}^{L} c_k \mathcal{D}_k^a$ and background noise N are much weaker than the subspace interference $H\theta$ with doppler a known, the GLRT has the approximate form

$$z_1 - 1 \approx \frac{\| P_{S_a''} \mathbf{x}'' \|_F^2}{\| P_{S_a''}^{\perp} \mathbf{x}'' \|_F^2} \tag{11}$$

where $P_{S_a''} = S''(S''^H S'')^{-1} S''^H$, $P_{S_a''}^{\perp} = I - P_{S_a''}$, $P_\theta^{\perp} = I - \theta^H(\theta\theta^H)^{-1}\theta$, $\mathbf{x}'' = vec(P_H^{\perp} X P_\theta^{\perp})$, and $S'' = [vec(P_H^{\perp}\mathcal{D}_1^a P_\theta^{\perp}) \,|\, vec(P_H^{\perp}\mathcal{D}_2^a P_\theta^{\perp}) \,|\, \cdots \,|\, vec(P_H^{\perp}\mathcal{D}_L^a P_\theta^{\perp})]$. The approximation (11) and UMPI test (9) are nearly the same except for a post-multiplication of X and the signal by the projection operator $P_\theta^{\perp}$. The post-multiplication of X by $P_\theta^{\perp}$ corresponds to an additional temporal nulling of the data ($P_\theta^{\perp}$ projects onto the orthogonal complement of the complex conjugate of the row space of the subspace interference matrix $H\theta$, which corresponds to the time series observed by each sensor due to this interference). The extra temporal nulling can be intrepreted as a loss due to estimation. Test (11) is also CFAR since it is invariant to scalings of the data matrix.

The distribution of (11) has been derived in [5] and was shown to be central and non-central F distributed under $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively.

2.2. Performance degradation

We now discuss the loss in performance when estimating doppler. When the signal is present and not too weak compared to the background noise N, we expect the test statistic (10) to be nearly the same as when doppler is known (at high signal-to-noise ratios doppler should be estimated accurately). However, when the signal is not present (noise only case) the value of (10) will clearly increase, resulting in an increased false alarm rate for the same threshold. We now determine the extent of the increase using (11). An exact analysis is difficult since it involves determining order statistics. Here we present an approximate analysis using (11) given that the possible range of dopplers is restricted to some small interval (often times we know the feasible target velocities).

Approximation (11) can be rewritten as

$$z_1 \approx \frac{1}{1 - \| P_{S_a''} \mathbf{x}'' \|_F^2 / \| \mathbf{x}'' \|_F^2} \tag{12}$$

When doppler is being estimated, the above approximation becomes

$$z_1 \approx \max_a \frac{1}{1 - \| P_{S_a''} \mathbf{x}'' \|_F^2 / \| \mathbf{x}'' \|_F^2} \tag{13}$$

An equivalent test statistic is

$$z_1' = \max_a \frac{\mathbf{x}''^H P_{S_a''} \mathbf{x}''}{\mathbf{x}''^H \mathbf{x}''} \tag{14}$$

Next linearize (14) about a_0 by keeping the first-order terms of its Taylor series expansion. This results in

$$z_1'(a) \approx \frac{\mathbf{x}''^H P_{S_{a_0}''} \mathbf{x}''}{\mathbf{x}''^H \mathbf{x}''} + (a - a_0) \left(\frac{\partial}{\partial a} \frac{\mathbf{x}''^H P_{S_a''} \mathbf{x}''}{\mathbf{x}''^H \mathbf{x}''} \Big|_{a=a_0} \right) \tag{15}$$

If a is restricted to some small interval $[a_0 - \Delta, a_0 + \Delta]$, the maximum of (15) must occur at one of the end points of the interval. Therefore the maximum of (14) is approximately

$$z_1' \approx \frac{\mathbf{x}''^H P_{S_{a_0}''} \mathbf{x}''}{\mathbf{x}''^H \mathbf{x}''} + \Delta \left| \frac{\partial}{\partial a} \frac{\mathbf{x}''^H P_{S_a''} \mathbf{x}''}{\mathbf{x}''^H \mathbf{x}''} \Big|_{a=a_0} \right| \tag{16}$$

where the first term in (16) is the GLRT when a is known and the last term is the perturbation due to estimating doppler.

We now calculate the second moment of the last term in (16) (the first moment is difficult because of the absolute value), that is, the expected value of

$$e = \frac{\Delta^2}{(\mathbf{x}''^H \mathbf{x}'')^2} \left| \frac{\partial}{\partial a} \mathbf{x}''^H P_{S_a''} \mathbf{x}'' \Big|_{a=a_0} \right|^2 \tag{17}$$

Replacing $(\mathbf{x}''^H \mathbf{x}'')^2$ in (17) by its expected value $(\sigma^2)^2((m-r)^2(K-r)^2 + 2(m-r)(K-r))$ and using some results in [9] for the moments of complex Wishart distributed matrices, the expected value of e is found to be

$$\mathcal{E}[e] \approx \frac{\Delta^2}{(m-r)^2(K-r)^2 + 2(m-r)(K-r))} \cdot \left(trace[\left(\frac{\partial}{\partial a} \hat{P}_{S_a''} \,|_{a=a_0} \right)^2] + trace^2[\frac{\partial}{\partial a} \hat{P}_{S_a''} \,|_{a=a_0}] \right) \tag{18}$$

where $\hat{P}_{\hat{S}_a''}$ is obtained applying the previous formulas using $\hat{S}'' = [vec(U_0^H \mathcal{D}_1^a V_0) \,|\, vec(U_0^H \mathcal{D}_2^a V_0) \,|\, \cdots \,|\, vec(U_0^H \mathcal{D}_L^a V_0)]$ in place of S'' and the orthonormal columns of matrices U_0 and V_0 span the column spaces of $P_H^{\perp}$ and $P_\theta^{\perp}$ respectively.

Discussion

As expected, to first-order the magnitude of the perturbation (relative to the detector when doppler is known) is related to the doppler resolution of the waveform. To adjust detector thresholds, we can approximately determine the expected value of the perturbation using $\sqrt{e}$ and then (13) to determine the increase of z_1.

3. NUMERICAL EXAMPLES

In order to evaluate the performance of the GLRT detector a number of studies were made. We simulated an active sonar system with an array of 10 hydrophones with half-wavelength spacing. The reverberation component was modeled as arising from IID Gaussian point scatterers (Rayleigh distributed amplitudes and uniformly distributed phases) along a line perpindicular to the center of the array. The per sample reverbation power is normalized to unity. The ambient noise component is modeled as white Gaussian with variance 3.125×10^{-3}. The transmitted pulse is a .6 second 400-425 Hz LFM waveform. The received target echo is modeled as Rayleigh fading with variance 1.95×10^{-5}. In all simulations the signal is arriving 1/2 of beamwidth from broadside, noting that the reverberation is arriving from broadside.

The target velocity was set at 4 m/sec. A total of 200 independent trials were performed. The UMPI test, GLRT when target velocity is known, and GLRT when target doppler is not known (doppler search is restricted to the interval 0-5 m/sec) were evaluated for each trial using the same realizations of interference and signal.

The measured ROC curves are plotted in figure 1. Note that the unknown target doppler GLRT is close in performance to the GLRT using the correct target doppler and also the UMPI test. Next, the square root of (18) (second moment of the increase of the approximate test statistic (14)) vs. Δ is plotted in figure 2 (note that $\Delta = .004$ corresponds to a velocity change of about 30 m/sec). This is compared with the experimentally measured second moment. The plots indicate the approximations are accurate over a wide range.

4. CONCLUSION

The theoretical and experimental analysis indicates that the proposed GLRT detectors perform well. Furthermore, formulas are provided relating the GLRT's to the UMPI test and allowing the approximate calculation of the expected increase of the test statistic when doppler is estimated.

REFERENCES

[1] L.L. Scharf, "The SVD and reduced-rank signal processing," Signal Processing, Vol. 25, pp. 113-133, 1991.

[2] L.L. Scharf and D.W. Tufts, "Rank reduction for modeling stationary signals," IEEE Trans. on Acoustics, Speech, and Signal Proc., Vol. ASSP-35, No. 3, pp. 350-354, March 1987.

[3] L.L. Scharf, *Statistical Signal Processing: Detection, estimation, and time series analysis*, Addison-Wesley Publishing Co., 1991.

[4] T.S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York, 1967.

[5] I.P. Kirsteins, "A Class of Generalized Likelihood-Ratio Tests for Detecting a Signal in Unknown Subspace Interference," in preparation.

[6] L.L. Scharf and B. Friedlander, "Matched subspace detectors," IEEE Trans. on Sig. Proc., Vol. 42, No. 8, pp. 2146-2156, Aug. 1994.

[7] K.R. Gabreil, "Lower Rank Approximation of Matrices by Least Squares Methods With Any Choice of Weights," Technometrics, Vol. 21, No. 4, pp. 489-498, Nov. 1979.

[8] H.V. Poor, *An introduction to signal detection*, Springer-Verlag, New York, 1988.

[9] J.A. Tague and C.I. Caldwell, "Expectations of useful complex Wishart forms," Ohio University technical report, Athens, OH, 1993.

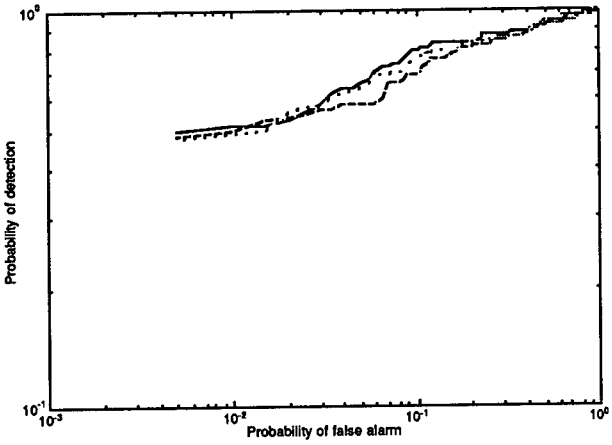


Figure 1. Experimentally measured ROC curves based on 200 trials. The curves are labeled as follows: solid - UMPI, widely dotted – GLRT using true doppler, dashed - GLRT when doppler is unknown.

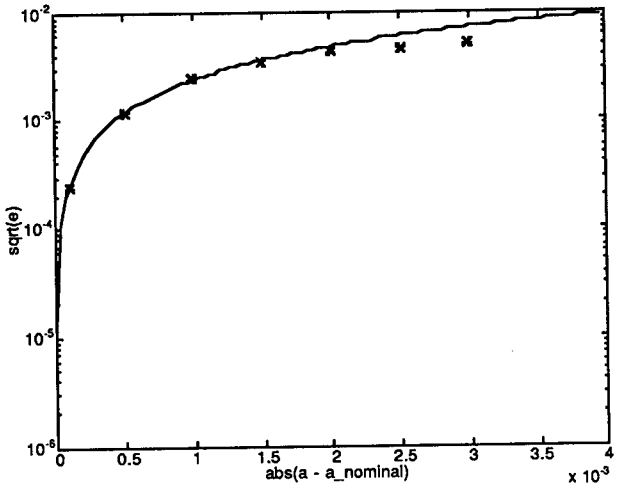


Figure 2. Expected increase of test statistic (14). Solid line is theoretical and * is experimental.

DETECTION AND ESTIMATION OF MULTIPLICATIVE JUMPS

Jean-Yves TOURNERET, Marie CHABERT and Mounir GHOGHO

National Polytechnics Institute of Toulouse
ENSEEIHT/GASPE, 2 rue Camichel, 31071 Toulouse, France
Tel: (33) 61 58 83 14 email: tournere@len7.enseeiht.fr

ABSTRACT

Muliplicative jumps have been considered in many applications. These applications include speckle signal in radar images, mechanical vibrations, non-linear time series and random communication models. The problem addressed here is the detection of multiplicative jumps using the Neyman-Pearson test. This test constitutes a reference to which suboptimal detectors can be compared. In practical applications, the parameters of the noise and of the jump have to be estimated. The Maximum Likelihood Estimator and the Cramer Rao bound for these parameters are then studied.

1. INTRODUCTION

This paper studies the performance of a multiplicative jump detector based on the Neyman-Pearson test. For the sake of simplicity, we consider the case of a shifted step embedded in a multiplicative non zero mean white Gaussian process, which leads to simultaneous mean value and variance jumps. This kind of signal has been considered in many applications. These applications include speckle signal on piecewise constant backgrounds in radar images, mechanical vibrations, non-linear time series and random communication models. In the first section, we formulate the problem and develop the optimal Neyman-Pearson test [1]. This test is optimal in the sense that it minimizes the probability of false alarm (Pfa) for fixed probability of non detection (Pnd). The second section is devoted to the estimation of the multiplicative jump parameters which leads to a suboptimal detector.

2. NEYMAN PEARSON TEST

The problem addressed here is the detection of multiplicative jumps using the Neyman-Pearson test.

Under hypothesis H_1, the signal is a stationary white Gaussian process $x(n)$ with mean m and variance σ^2.

Under hypothesis H_0, the process $x(n)$ is multiplied by a step of amplitude A at time n_0:

$$y(n) = x(n)\left[1 + A.U\left(n - n_0\right)\right]$$

where $U(n)$ is the Heaveside step. The Neyman-Pearson test is then defined by:

$$H_0 \text{ rejected if } \quad \frac{L\left(Y\,|H_1\right)}{L\left(Y\,|H_0\right)} > k(Pnd) \qquad (1)$$

In (1), $L\left(Y\,|H_i\right)$ is the Likelihood function for the vector $Y = [y(1), ..., y(N)]^t$ under hypothesis H_i. Using the normality of vector Y, H_0 is rejected if:

$$Z = \sum_{i=n_0+1}^{N} \left[y(i) - m\left(\frac{1+A}{2+A}\right)\right]^2 < S(Pnd) \qquad (2)$$

Introducing the unit normal n-dimensional variable $W = [w(1), ..., w(N)]^t$

$$W = \frac{Y - m\left(1+A\right)}{\sigma\left(1+A\right)} \text{ under } H_0 \qquad (3)$$

and

$$W = \frac{Y - m}{\sigma} \text{ under } H_1$$

we can express Z as the sum of $N - n_0$ independent and identically distributed (i.i.d.) variables:

$$Z = d_j \sum_{i=n_0+1}^{N} \left(w(i) + M_j\right)^2 \text{ under } H_j \qquad (4)$$

with

$$M_0 = \frac{m\left(1+A\right)}{\sigma\left(2+A\right)}, d_0 = \sigma^2\left(1+A\right)^2 \text{ under } H_0 \qquad (5)$$

and

$$M_1 = \frac{m}{\sigma\left(2+A\right)}, d_1 = \sigma^2 \text{ under } H_1 \qquad (6)$$

Eq. 4 shows that, under hypothesis H_j, the distribution of Z/d_j is a non-central χ^2 distribution with $N - n_0$

0-8186-7576-4/96 $5.00

degrees of freedom and with non-centrality parameter $\lambda_j = (N - n_0) M_j$ [4]. The probabilities of false alarm and of non detection can then be expressed as functions of the cumulative distribution function of a non-central χ^2 distribution :

$$Pnd = \int_{-\infty}^{S(Pnd)/d_0} f_0(t)\, dt \tag{7}$$

$$Pfa = \int_{S(Pnd)/d_1}^{+\infty} f_1(t)\, dt \tag{8}$$

In these equations, $f_j(t)$ denotes the probability density function of the χ^2 distribution with $N - n_0$ degrees of freedom and with non-centrality parameter $\lambda_j = (N - n_0) M_j$. As an example, we consider $N = 2048$ samples of a Gaussian distributed random sequence with $m = 1$ and $\sigma^2 = 1$. The multiplicative jump occurs at time $n_0 = 1024$. The variations of pfa and pnd as functions of the threshold S are plotted in Fig. 1 for different jump amplitude A:

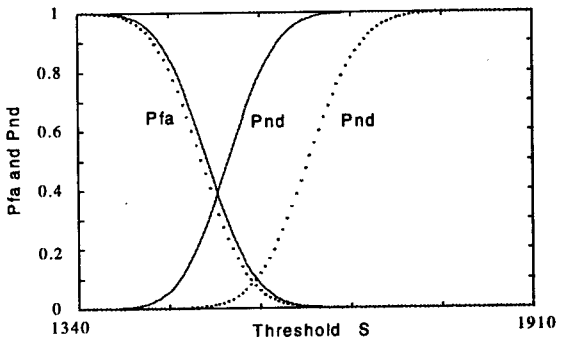


Figure 1: Pfa and Pnd as functions of the threshold S (Dashed Line: $A = 0.05$, Continuous Line: $A = 0.01$).

As it can be seen, the Neyman-Pearson test shows good performance. This test constitutes a reference to which suboptimal detectors can be compared. To study the sensitivity of the test as a function of the jump amplitude A, we have plotted in Fig. 2 the variations of the probability of false alarm as a function of A for a fixed probability of non detection ($Pnd = 0.01$). As can be seen, a multiplicative jump with amplitude $A \geq 0.1$ can be detected with low probabilities of non detection and of false alarm. However, in practical applications, the parameters m, σ^2, A and n_0 are unknown and have to be estimated. In the next part of the paper, we derive the Maximum Likelihood Estimator (MLE) and the Cramer Rao Bound for these parameters.

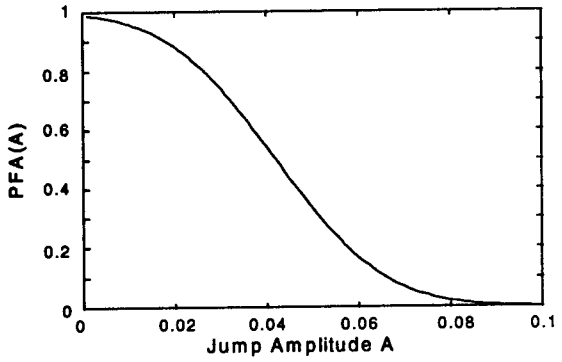


Figure 2: Probability of False alarm as a function of the jump amplitude for fixed probability of non detection $P_{ND} = 0.01$.

3. MAXIMUM LIKELIHOOD ESTIMATOR

The Maximum Likelihood principle [1] provides a method to estimate the parameter vector $\theta = \left(m, \sigma^2, A, n_0\right)^t$ from a finite length data record $Y = (y_1, ..., y_N)^t$. When a jump occurs at time n_0, the likelihood function of the Gaussian vector Y is defined by:

$$L(Y;\theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}} (1+A)^{N-n_0}} \exp -\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^{N} \left(\frac{y_i}{d_i} - m \right)^2 \right\} \tag{9}$$

with $d_i = 1$ for $i \in \{1, ..., n_0\}$ and $d_i = 1 + A$ for $i \in \{n_0 + 1, ..., N\}$. The MLE of the vector θ denoted by $\widehat{\theta}_{ML}$ is the one which maximizes the likelihood function over a subset Θ of $\mathbf{R}^3 \times E$ with $E = \{1, ..., N\}$. When the vector $(A, n_0)^t$ is known, the MLE of (m, σ^2) is obtained by setting to 0 the partial derivatives of $L(Y;\theta)$ with respect to m and σ^2:

$$\widehat{m}_{ML} = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i}{d_i} \tag{10}$$

$$\widehat{\sigma}^2_{ML} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{Y_i}{d_i} - \widehat{m}_{ML} \right)^2 \tag{11}$$

When the vector $(A, n_0)^t$ is unknown, we substitute the expression of $\widehat{m}_{ML}$ and $\widehat{\sigma}^2_{ML}$ in (9) and drop the constant terms. We need now to maximize:

$$J_1(Y; A, n_0) = -(N - n_0)\, Ln\,|1 + A| - \frac{N}{2} \ln \frac{1}{N} \sum_{i=1}^{N} \left(\frac{y_i}{d_i} - f(Y; A, n_0) \right)^2 \tag{12}$$

with:

$$f(Y;A,n_0)=\frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{d_i} \tag{13}$$

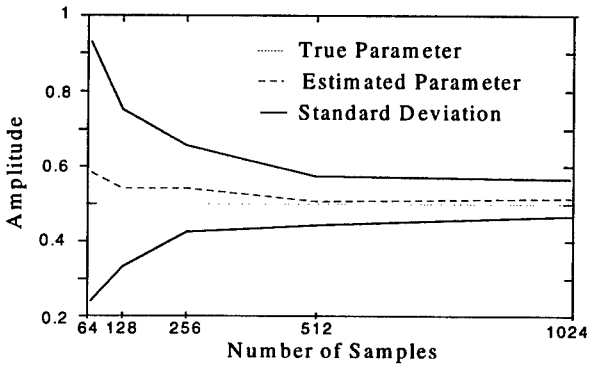
Setting to 0 the derivative of $J_1(Y;A,n_0)$ with respect to A leads to a second degree equation with respect to A. The solution of this equation gives us an expression of A as a function of the jump time n_0 and of the observations y_i denoted by $J_2(Y;n_0)$. When n_0 is known, the MLE of A is then given by:

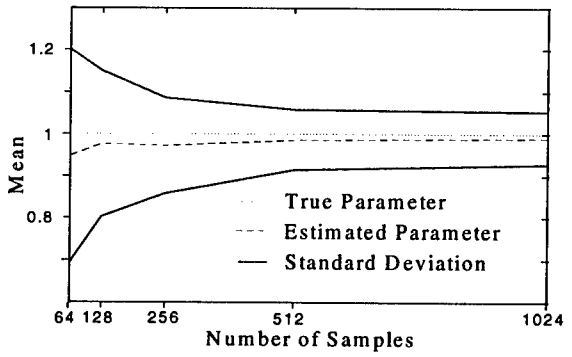
$$\widehat{A}_{ML}=J_2(Y;n_0) \tag{14}$$

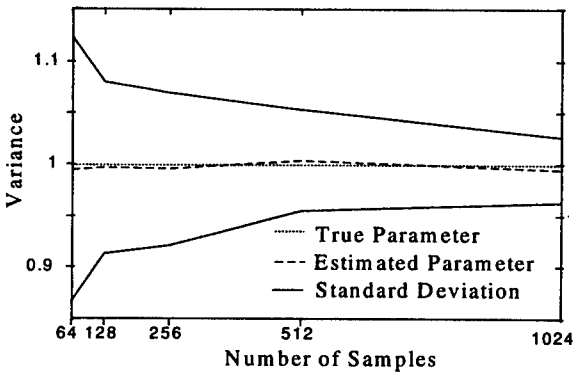
When n_0 is unknown, its MLE is obtained by the argument of the maximum of the criterion $J_3(Y;n_0)=J_1\left(Y;\widehat{A}_{ML},n_0\right)$ such that:

$$J_3(Y;\widehat{n_0}_{ML})=\underset{n=1,...,N}{Max}\,J_3(Y;n) \tag{15}$$

In other words, the maximization of $L(Y;\theta)$ over the whole parameter vector θ is equivalent to the maximization of $J_3(Y;n_0)$ with respect to n_0 only. The MLE of m,σ^2 and A are then given by replacing n_0 by $\widehat{n_0}_{ML}$ in (10), (11) and (14). Note that the maximization of $J_3(Y;n_0)$ with respect to n_0 is a discrete maximization which is very simple to implement. The mean and standard deviation of the MLE are shown in Fig. 2.a), b) , c) and d) and compared to the true parameters $m=1,\sigma^2=1,A=0.5$ and $n_0=\frac{N}{2}$ for different number of samples N.







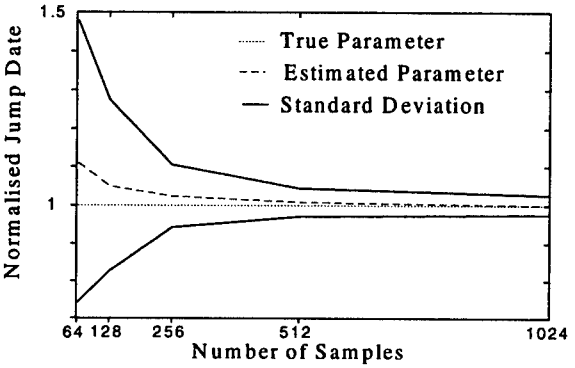


Fig. 2. Mean and Standard Deviation of the estimated parameters for different numbers of samples N (a) $\widehat{A}_{ML}$ *(b)* $\widehat{m}_{ML}$ (c) $\widehat{\sigma}^2_{ML}$ (d) $\widehat{n_0}_{ML}/n_0$.

4. CRAMER RAO BOUND

It is well known that the covariance matrix of any unbiased estimator cannot be smaller than the inverse of the Fisher information matrix known as the Cramer-Rao bound (CRB). For a parameter vector $\theta=(\theta_1,...,\theta_p)^t$, the elements of the Fisher information matrix are given by:

$$[I(\theta)]_{k,l}=E\left\{-\frac{\partial^2 \ln L(Y;\theta)}{\partial\theta_k\partial\theta_l}\right\}\qquad 1\le k,l\le p \tag{16}$$

where $L(y;\theta)$ is the probability density function of the vector $y=(y_1,...,y_N)^t$. For Gaussian time series, many equivalent expressions for the Fisher information matrix can be found in the literature [3]. For instance, we have:

$$\begin{aligned}[I(\theta)]_{k,l} &= \frac{1}{2}tr\left\{R^{-1}(\theta)\frac{\partial R(\theta)}{\partial\theta_k}R^{-1}(\theta)\frac{\partial R(\theta)}{\partial\theta_l}\right\}\\ &\quad+\left[\frac{\partial m(\theta)}{\partial\theta_k}\right]^t R^{-1}(\theta)\left[\frac{\partial m(\theta)}{\partial\theta_l}\right]^t\end{aligned} \tag{17}$$

where tr{.} denotes the trace operator and $R(\theta),m(\theta)$ are the covariance matrix and the mean of the vector $Y=(Y_1,...,Y_N)^t$.

When n_0 is known, using eq. (16) or (17), the Fisher's information matrix I_α corresponding to the vector $\alpha = (m, \sigma^2, A)^t$ can be computed. The following results can be obtained:

$$I_\alpha = \begin{pmatrix} \frac{N}{\sigma^2} & 0 & \frac{m(N-n_0)}{(1+A)\sigma^2} \\ 0 & \frac{N}{2\sigma^4} & \frac{N-n_0}{(1+A)\sigma^2} \\ \frac{m(N-n_0)}{(1+A)\sigma^2} & \frac{N-n_0}{(1+A)\sigma^2} & \frac{N-n_0}{(1+A)^2}\left(\frac{m^2}{\sigma^2}+2\right) \end{pmatrix}$$

The determinant of this matrix is of the form $\det(I_\alpha) = Cn_0(N-n_0)$ with $C > 0$. Thus, when the jump occurs at time $n_0 = N$ or $n_0 = 0$, I_α is singular and we cannot estimate A with (14).

When n_0 is unknown, the parameter vector is $\theta = (m, \sigma^2, A, n_0)^t$. The problem of finding a bound for the covariance matrix of $\widehat{\theta}_{ML}$ becomes difficult because n_0 is a discrete parameter (belonging to the set $\{1, ..., N\}$). If we consider that the jump occurs at time $t_0 \in [0, T]$, the MLE of t_0 is obtained by maximizing $J_3(Y; n_0 + 1)$, $n_0 = int(t_0)$ being the integer part of t_0. In this case, we cannot derivate the likelihood function with respect to t_0 which prevents us to compute the terms $[I_\theta]_{i,4}$ of the Fisher's information matrix. Thus, the CRB for the vector θ cannot be computed.

For a known parameter n_0, a comparison between the mean square error (MSE) of $\widehat{\theta}_{ML}$ estimated with $N_r = 500$ Monte-Carlo runs and the CRB is presented in Fig. 3. In this figure, the MSE of $\widehat{\theta}_{ML}$ in the case of known and unknown parameter n_0 is also compared.

5. CONCLUSION

The optimal Neyman-Pearson multiplicative jump detector is derived. For fixed probability of non detection, the threshold minimizing the probability of false alarm can be determined. This test constitutes a reference to which suboptimal detectors can be compared. In practical applications, the parameters of the noise and of the jump are unknown. The Maximum Likelihood Estimator and the Cramer Rao Bound for these parameters are then studied. The next step in our study will be to compare our results with a wavelet based detection strategy [2].

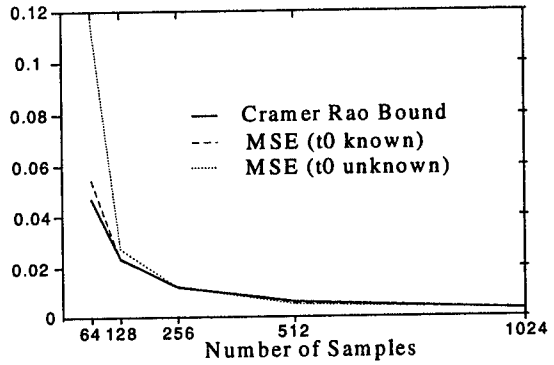
6. REFERENCES

[1] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, John Wiley & Sons, 1968.

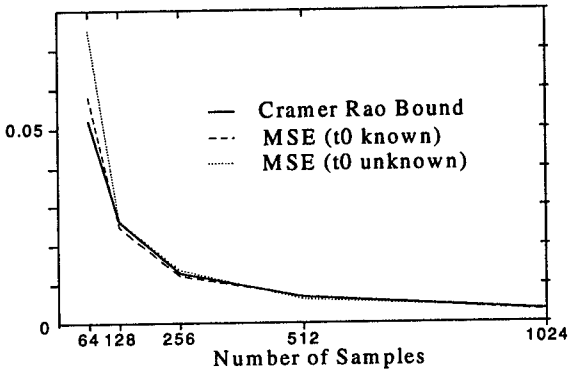
[2] M. Chabert, J-Yves Tourneret and F. Castanié, "*Additive and Multiplicative Jump Detection using the Continuous Wavelet Transform*", Proc. of ICASSP'96, Atlanta, to appear.

[3] B. Porat and B. Friedlander, "Computation of the Exact Information Matrix of Gaussian Time Series with Stationary random Components, " *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 118-130, Feb. 1986.

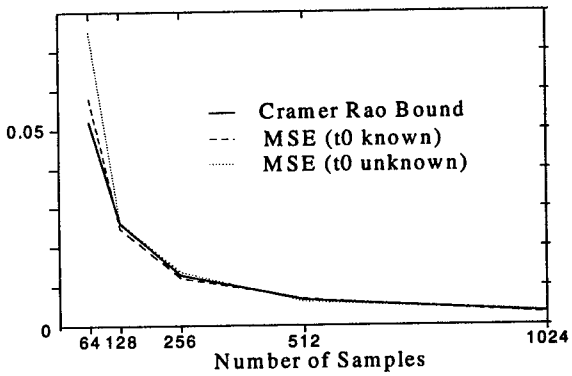
[4] N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions-2, 2nd Edition*, John Wiley, 1994.



(*a*) Amplitude $\widehat{A}_{ML}$.



(*b*) Mean $\widehat{m}_{ML}$.



(*c*) Variance $\widehat{\sigma}^2_{ML}$.

Fig. 3. Cramer Rao Bound (t_0 known) and MSE of parameters (t_0 known and unknown) for different number of samples N (a) $\widehat{A}_{ML}$ *(b)* $\widehat{m}_{ML}$ (c) $\widehat{\sigma}^2_{ML}$.

Bayesian Estimation of the Variance of a Jitter Using MCMC

†Christophe Andrieu- †‡Arnaud Doucet- †Patrick Duvaut
†ENSEA-ETIS Groupe Signal - 6 avenue du Ponceau, 95014 Cergy Pontoise France
‡CEN Saclay LETI/DEIN/SPE Bat. 451, 91191 Gif-sur Yvette cedex France
andrieu@ensea.fr - douceta@ensea.fr - duvaut@ensea.fr

Abstract

The problem treated in this paper is the Bayesian estimation of the variance of the sampling jitter occuring when a process is irregularly observed. This problem is often met in practice [2], and has already received treatment in [1][5] using higher order statistics. The Bayesian solution to this problem is performed using powerful stochastic algorithms, the MCMC (Markov Chain Monte Carlo) methods.

1. Statement of the problem

1.1. Motivations

The problem addressed in this paper is the estimation of the variance of the jitter occuring while sampling irregularly a process whose a priori density is known. This problem has already received treatment in [5] for example, and [1] in the case of a Gaussian process, using higher order statistics: the second method is mainly based on the fact that a continuous Gaussian process does not give birth to a discrete Gaussian process when irregularly observed. In this paper we propose a Bayesian statistical approach for estimating this quantity, in a wider framework, because we remove the Gaussian assumption.

The main interests of the approach we develope in this paper are that:

- it does not require a lot of observations (as in the case of higher order statistics),
- we remove the assumption of Gaussianity of the continuous process which is sampled,
- we estimate the a posteriori probability density of the jitter, (thus allowing the calculus of conditional expectations, confidence intervals...)
- we use stochastic algorithms, the MCMCs (Markov Chain Monte Carlo), which have been very popular for fifty years in statistical physics, and more recently in image processing and statistics, but are not yet popular in signal processing despite their power.

1.2. Assumptions-Notations

- $(x(t))_{t\in\mathbb{R}}$ is a continuous time process,
- this process is sampled at times:

$$t_n = n + \varepsilon_n \tag{1}$$

where the $(\varepsilon_n)_{n\in\mathbb{Z}}$ are zero mean iid, and we note $y_n = x(t_n)$.

- $x_{1\to N}$ is a useful notation for $(x_n)_{n=1,\dots,N}$.

2. Bayesian solution to the problem

We wish to estimate the following density:

$$p(\sigma / y_{1\longrightarrow N}) \tag{2}$$

where σ is the variance of the sampling jitter, and $y_{1\longrightarrow N}$ the observations.

Remark 1 *it is worth noticing that in this paper, we restrict ourselves to σ as an unknown parameter of the distribution of the time perturbations. Obviously σ could be replaced in all what follows by the complete finite dimensional parameter vector $\underline{\theta}$, characterizing the distribution.*

This problem can be thought as a missing data problem, where the $y_{1\longrightarrow N}$ are the incomplete data, which can be completed by the $\varepsilon_{1\to N}$ to form the complete data set $y_{1\to N}, \varepsilon_{1\to N}$.

We thus use a stochastic algorithm that allows us to estimate the following joint density:

$$p(\sigma, \varepsilon_{1\longrightarrow N} / y_{1\longrightarrow N}) \tag{3}$$

and thus the required density, $p(\sigma / y_{1\longrightarrow N})$.

A natural method would be here to use the Gibbs sampler [4], which consists in drawing iteratively and alternatively subsets of the parameters, according to

0-8186-7576-4/96 $5.00

others, thus building a chain of samples. Under mild conditions [4], the joint density of the samples drawn as described above will converge to $p(\sigma, \varepsilon_{1\longrightarrow N}/y_{1\longrightarrow N})$ thus providing a representation of this joint density.

In our case, sampling from the Gibbs sampler amounts to draw iteratively and alternatively with the following laws:

$$\begin{array}{l} p(\sigma/y_{1\longrightarrow N}, \varepsilon_{1\longrightarrow N}) \\ p(\varepsilon_{1\longrightarrow N}/\sigma, y_{1\longrightarrow N}) \end{array} \quad (4)$$

which using the Bayes's rule yields:

$$\begin{array}{lll} p(\sigma/y_{1\longrightarrow N}, \varepsilon_{1\longrightarrow N}) & \propto & p(\varepsilon_{1\longrightarrow N}/\sigma)p(\sigma) \\ p(\varepsilon_{1\longrightarrow N}/\sigma, y_{1\longrightarrow N}) & \propto & p(y_{1\longrightarrow N}/\varepsilon_{1\longrightarrow N})p(\varepsilon_{1\longrightarrow N}/\sigma) \end{array} \quad (5)$$

Where:

- $\propto$ means here "is proportional to",
- we note that $p(\varepsilon_{1\longrightarrow N}/\sigma, y_{1\longrightarrow N}) \propto p(\varepsilon_{1\longrightarrow N}/\sigma)$,
- and that $p(y_{1\longrightarrow N}/\sigma, \varepsilon_{1\longrightarrow N}) = p(y_{1\longrightarrow N}/\varepsilon_{1\longrightarrow N})$.

Remark 2 *it is worth noticing that without any additional difficulty it is possible to suppose that the process is embedded in noise, thus taking into account thermal noise, quantification noise, which was not so simple with previous methods. In that case, one would only write:*

$$\begin{aligned} p(y_{1\to N}/\sigma, \varepsilon_{1\to N}, x_{1\to N}) &\propto p(x_{1\to N}/y_{1\to N}) \\ &\quad \times p(y_{1\longrightarrow N}/\varepsilon_{1\longrightarrow N}) \end{aligned}$$

$$\begin{array}{l} p(\sigma/y_{1\longrightarrow N}, \varepsilon_{1\longrightarrow N}, x_{1\longrightarrow N}) = p(\sigma/y_{1\longrightarrow N}, \varepsilon_{1\longrightarrow N}) \\ p(\varepsilon_{1\longrightarrow N}/\sigma, x_{1\longrightarrow N}, y_{1\longrightarrow N}) \propto p(x_{1\longrightarrow N}/y_{1\longrightarrow N}) \\ \qquad \times p(y_{1\longrightarrow N}/\varepsilon_{1\longrightarrow N}) \\ \qquad \times p(\varepsilon_{1\longrightarrow N}/\sigma) \end{array} \quad (6)$$

where the $x_{1\longrightarrow N}$ are the observations including noise, which is assumed to be stationary for sake of convinience. We notice that it requires to be able to sample from $p(y_{1\longrightarrow N}/\varepsilon_{1\longrightarrow N})$.

The simulation algorithm could thus be summarized as follows [4]:

1. Draw $\sigma^{(k+1)}$ according to $p\left(\sigma \Big/ y_{1\longrightarrow N}, \varepsilon_{1\longrightarrow N}^{(k)}\right)$.
2. Draw $\varepsilon_{1\longrightarrow N}^{(k+1)}$ according to $p\left(\varepsilon_{1\longrightarrow N}/\sigma^{(k+1)}, y_{1\longrightarrow N}\right)$.
3. Go to 1.

Nevertheless in practice it is generaly impossible to perform such an algorithm directly:

- in most cases, it is impossible to apply an accept/reject procedure [4] using (5), as it would require to determine an efficient generating density, so as to ensure a good acceptation rate, and to evaluate analytically the normalization constant of the likelihood.

- in most cases, one can not determine prior conjugate densities, which would simplify the drawing procedure. For example in the case where the $\varepsilon_{1\longrightarrow N}$ are iid Gaussian, one could choose $\sigma^2 \sim \text{Inv}-\chi^2(\alpha, \beta)$.

- due to the limited precision of computers small values are rounded to zero, and the algorithm may not converge. This often happens when we deal with joint densities of large size vectors.

In order to circumvent all those problems, we propose a combination of two MCMC, more precisely a product of two Metropolis-Hastings (M-H) kernels [4], whose respective invariant densities are the full conditional densities, $p(\sigma/y_{1\longrightarrow N}, \varepsilon_{1\longrightarrow N})$ and $p(\varepsilon_{1\longrightarrow N}/\sigma, y_{1\longrightarrow N})$. This algorithm is no more a Gibbs sampler (which nevertheless is a particular case of product of specific M-H kernels). We use M-H based on random walk [4] (ie a simple Metropolis algorithm), that is we make the parameters evolve with random increments: in what will follow, the scalar increment for σ will be named z and will be distributed according to q_z which must be symetric [4]. The vector increment for $\varepsilon_{1\longrightarrow N}$ will be $U_{1\longrightarrow N}$, distributed according to q_U which must be also symetric. (We notice that we could make the 'q' depend on the preceeding step). This provides a general algorithm which is ensured to work in all situations (under convergence conditions), but which may not be computationaly efficient in particular cases, and thus would have to be adapted.

The algorithm can thus be summarized as follows:

1. Draw σ_1^* according to p(σ) and $\varepsilon_{1\longrightarrow N}^{(1)}$ conditionally to σ_1^*.
2. while(i<M)
   - (Initialization: $i = 0$)
   - $i = i + 1$
   - Draw z^* according to q_z
   - Set $\sigma_2^* = \sigma_1^* + z^*$
   - Calculate α
   - $= \min\left[1, \frac{p\left(\varepsilon_{1\longrightarrow N}^{(k)}/\sigma_2^*\right)p(\sigma_2^*)}{p\left(\varepsilon_{1\longrightarrow N}^{(k)}/\sigma_1^*\right)p(\sigma_1^*)}\right]$
   - Accept the value σ_2^* with probability α.
   - If σ_2^* is accepted then set
   - $\sigma_1^* = \sigma_2^*$.
3. $\sigma^{(k+1)} = \sigma_1^*$
4. while(i<M')

$$\left\{\begin{array}{c}\text{(Initialization: } i=0)\\ i=i+1\\ \text{Draw } U_{1\to N}^{*} \text{ according to } q_U\\ \varepsilon_{1\longrightarrow N}^{**}=\varepsilon_{1\longrightarrow N}^{*}+U_{1\to N}^{*}\\ \text{Calculate } \alpha\\ \min\left[1,\frac{p(y_{1\longrightarrow N}/\varepsilon_{1\longrightarrow N}^{**})p(\varepsilon_{1\longrightarrow N}^{**}/\sigma^{(k+1)})}{p(y_{1\longrightarrow N}/\varepsilon_{1\longrightarrow N}^{*})p(\varepsilon_{1\longrightarrow N}^{*}/\sigma^{(k+1)})}\right]\\ \text{Accept } \varepsilon_{1\longrightarrow N}^{**} \text{ with probability } \alpha.\\ \text{If } \varepsilon_{1\longrightarrow N}^{**} \text{ is accepted then set}\\ \varepsilon_{1\longrightarrow N}^{*}=\varepsilon_{1\longrightarrow N}^{**}\end{array}\right.$$

5. $\varepsilon_{1\longrightarrow N}^{(k+1)}=\varepsilon_{1\longrightarrow N}^{*}$

6. go to 2.

One of the advantages of using the M-H kernels, is that ratios of densities appear, which allow to avoid the numerical problems discussed above.

Remark 3 *one notices that in the case where M, M' are sufficiently high, so that the stationary regime is reached, this would lead to a Gibbs sampler. Nevertheless as it will be shown in the next section, this is not required, but the algorithm obtained in such a case is no more a Gibbs sampler.*

3. Convergence of the algorithm

We give sufficient conditions to ensure convergence of this stochastic algorithm using Markov chains theory. Let E be the state space of the Markov chain. We assume $E = E_1 \times E_2$ is an open connected subset of $\mathbb{R}^N \times \mathbb{R}$. Furthermore we use x_1, y_1 to denote elements of E_1 and x_2, y_2 to denote elements of E_2. Let $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$ be the Borel σ-field on E.

We make the following assumptions:

- $\pi(dx)$ admits a stricly positive density $\pi(dx)$ on E with respect to the Lebesgue measure.

- $\pi_{1|2}(dx_1|x_2)$ and $\pi_{2|1}(dx_2|x_1)$ admit densities $\pi_{1|2}(x_1|x_2)$ and $\pi_{2|1}(x_2|x_1)$ on their space with respect to the Lebesgue measure.

- $\forall y_1 \in E_1$ $q(x_1, y_1|x_2)$ (resp. $\forall x_2 \in E_2$ $q(x_2, y_2|y_1)$) is stricly positive on $E_1 \times E_1$(resp. $E_2 \times E_2$)

The transition probability kernel $P : E \times \mathcal{F} \to [0,1]$ is in our case defined as follows. Let us first consider the 'local' transition kernels:

$$P_1(x_1, dy_1|x_2) = q(x_1, y_1|x_2)\,\alpha(x_1, y_1|x_2)\,dy_1 + \underbrace{\left[1 - \int q(x_1, y_1|x_2)\,\alpha(x_1, y_1|x_2)\,dy_1\right]}_{r(x_1/x_2)} \delta_{x_1}(dy_1)$$

where $\alpha(x_1, y_1|x_2)$ is the probability of accepting the candidate y_1 sampled according to $q(x_1, y_1|x_2)$. In a zero-mean random walk setup $\alpha(x_1, y_1|x_2) = \min\left(\frac{\pi_{1|2}(y_1|x_2)}{\pi_{1|2}(x_1|x_2)}, 1\right)$ if $\pi_{1|2}(x_1|x_2)\,q(x_1, y_1|x_2) > 0$ and 1 elsewhere. Similarly

$$P_2(x_2, dy_2|y_1) = p(x_2, y_2|y_1)\,dy_2 + r(x_2|y_1)\,\delta_{x_2}(dy_2)$$

The total transition kernel is thus equal to

$$P((x_1, x_2), (A_1, A_2)) = \int_{A_1}\int_{A_2} P_1(x_1, dy_1|x_2)\,P_2(x_2, dy_2|y_1)$$

To establish that this Markov chain converges towards the required posterior density $\pi(x_1, x_2)$, it is sufficient to show that P admits π as invariant density and that P is π-irreducible and aperiodic.

- Under the above assumptions, $\forall (x, A) \in E \times \mathcal{F}$ $P(x, A) > 0$, thus P is π-irreducible and aperiodic.

- π is invariant for P. Indeed,

$\int\int P_1(x_1, dy_1|x_2)\,P_2(x_2, dy_2|y_1)\,\pi(x_1, x_2)\,dx_1 dx_2$
$= \int P_2(x_2, dy_2|y_1)$
$\times\left[\int P_1(x_1, dy_1|x_2)\,\pi(x_1|x_2)\,dx_1\right]\pi(x_2)\,dx_2$
$= \int P_2(x_2, dy_2|y_1)\,\pi_{1|2}(dy_1|x_2)\,dx_2$
$= \int P_2(x_2, dy_2|y_1)\,\frac{\pi_{2|1}(x_2|y_1)\pi_1(dy_1)}{\pi(x_2)}\pi(x_2)\,dx_2$
$= \pi_1(dy_1)\int P_2(x_2, dy_2|y_1)\,\pi_{2|1}(x_2|y_1)\,dx_2$
$= \pi_1(dy_1)\,\pi_{2|1}(dy_2|y_1) = \pi(dy_1, dy_2)$

From π-irreducibility and aperiodicity one deduces that,

$$\|P^n(x, .) - \pi\| \to 0 \ \forall x \in E$$

where $\|.\|$ is the total variation norm. On the contrary of what is claimed in many papers, one can not conclude about ergodicity of the Markov chain, because it would require to establish in addition that P is Harris reccurent. In many cases - when there is no measure-theoretic pathology - irreducibility implies Harris recurence. However there exists yet no general results on the convergence of hybrid samplers, and we have not been able to establish this property rigorously. If that property was true, then this Markov chain would be ergodic, i.e. for any real-valued $\pi-$integrable function f one would have

$$\frac{1}{n}\sum_{k=1}^{n} f(x(k)) \to \int f(x)\,\pi(dx) \text{ almost surely}$$

4. Simulation

In order to compare the performance of our method with performance of previous methods we have chosen to follow [1]. The continuous process is thus assumed to be Gaussian, but we notice that normality is not required to perform the algorithm. One must only be able to evaluate $p(y_{1\longrightarrow N}/\varepsilon_{1\longrightarrow N})$.

- The correlation matrix of the Gaussian process is:

$$[R]_{(i,j)\in[1,n]^2} = \exp\left(\frac{-(t_i - t_j)^2}{2}\right) \qquad (7)$$

where the $(t_i)_{i=1,\ldots,n}$ are the sampling times.

- The perturbations are iid centered normal, (in the simulation $\mathcal{N}(0, .07)$) restricted to $[-.5, +.5]$ in order to place ourselves in a case similar to [1] and to ensure the convergence of the algorithm.

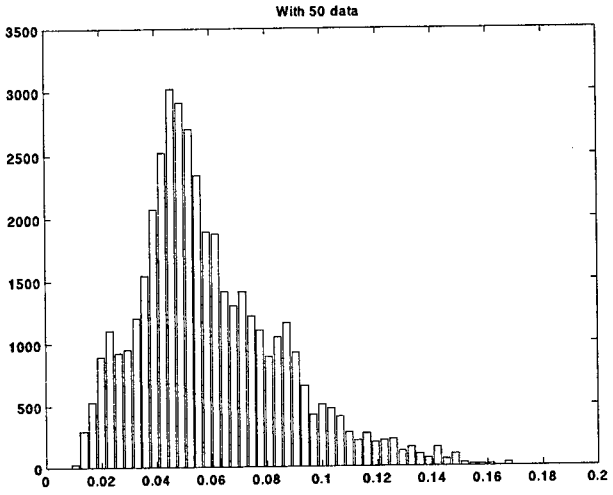
- σ has a noninformative prior, ie is distributed according to the uniform law on $[0, .3]$.

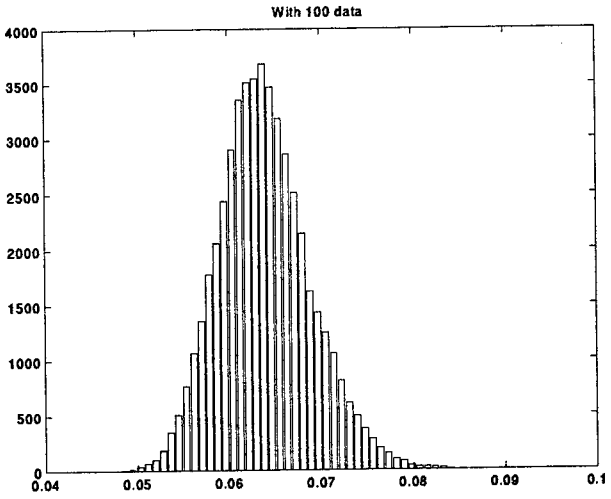
- the increment laws are $\mathcal{N}(0, .3)$ and multivariate $\mathcal{N}(0, .3I_N)$.

- the number of iterations for both Metropolis-Hastings algorithms are 200.

- we performed the algorithm in two cases: 50 and 100 data available.

The results are good, and even better than those obtained by [1] althought a small amount of data was available.





5. Conclusion

In our paper, we propose an original and efficient solution to a problem which is of interest in many applications, when data are not sampled at regular intervals. This is a typical missing data problem that we have solved in a Bayesian framework using MCMC: this avoids complex expectation evalution and often untractable global optimisation encountered when using the E-M algorithm and related versions. Our solution allows one to estimate not only the density of the variance of the jitter, but also the densities of each of the perturbation. We believe that the procedure is sufficiently flexible to be applied to many situations met in practice. Of course, MCMC techniques have many other applications in statistical signal processing, see [3] for an application to deconvolution.

References

[1] N. D. Aakvaag and B. Lacaze, "Estimation of Sampling Jitter Variance Using Fourth Order Cumulants", IEEE Signal Processing Athos Workshop on Higher Order Statistics, Spain, June 1995.

[2] D. Brillinger, S. Fiensberg, J. Gani, J. Hartigan, K. Krickeberg, Time Series of Irregularly Observed Data, Lecture Notes in Statistics, Proceedings, College Station, Springer Verlag (1983).

[3] A. Doucet and P. Duvaut, "Fully Bayesian Analysis of Conditionally Linear Gaussian State Space Models", Proceedings IEEE Conference ICASSP, Atlanta, May 1996.

[4] A.F.M. Smith, G.O. Roberts, "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods", J.R. Statist. Soc. B (1993) 55, No. 1, pp. 3-23.

[5] I. Sharfer and H. Messer, "A Suboptimal Estimator of the Sampling Jitter Variance Using the Bispectrum", Signal Processing, July 1994, Vol. 38, pp. 169-186.

Modeling and Estimation for Doppler-shifted Gaussian Random Processes

Christophe Couvreur*
Service de Physique générale
Faculté Polytechnique de Mons
B-7000 Mons (Belgium)
couvreur@thor.fpms.ac.be

Yoram Bresler†
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801
ybresler@uiuc.edu

Abstract

We address the problems of modeling Doppler-shifted wide-band Gaussian random processes and of estimating the Doppler parameter from a finite series of discrete-time samples. Relations between the continuous-time process, the Doppler shift parameter, and the discrete-time process obtained by sampling are established. Approximate rational models are proposed. Various estimators are proposed for Doppler parameter when the second-order statistics of the original continuous-time random process are known. The Cramér-Rao bound is derived. The estimators are compared experimentally on synthetic Doppler-shifted data. We also hint at some extensions of the method to non-stationary processes and time-varying Doppler shifts.

1. Introduction

The Doppler-shift effect is well-known for narrow-band signals emitted by moving sources (Fig. 1). In that case, freshman's physics tells us that a harmonic wave of frequency ω_0 emitted by a point source moving toward a fixed receiver with speed v is observed with apparent frequency $\omega = (1 + M)\omega_0$, where $M = v/c$ is the Mach number. In this paper, we consider wide-band moving sources that are modeled by continuous-time Gaussian random processes with known statistics. We address the problems involved in estimating the Doppler shift from samples of signals emitted by such sources. The estimators proposed here are motivated by applications in acoustics, e.g., in environmental sound monitoring, where wide-band moving sources are often encountered. For example, the maximum-likelihood estimator, or one of its approximations, can be used in a GLRT for detection and classification of Doppler-shifted wide-band processes given a dictionary of possible non-shifted spectra.

*Christophe Couvreur is a Research Assistant of the Belgian National Fund for Scientific Research at Faculté Polytechnique de Mons, Mons, Belgium.

†The work of Yoram Bresler is supported in part by a National Science Foundation grant No. MIP 91-57377.

Figure 1. Source moving toward the receiver.

2. Mathematical formulation

Let $x_c(t)$ be a continuous-time Gaussian zero-mean stationary random process modeling the signal emitted by the moving point source S and let $y_c(t)$ be the signal observed at the fixed receiver O (see. Fig. 1). For simplicity, we consider the one-dimensional case of a source moving at a constant speed v toward O. It can be shown that in the far field $y_c(t)$ is related to $x_c(t)$ by [5]

$$y_c(t) \approx \sigma x_c(\alpha t - \delta), \tag{1}$$

where $\alpha = 1 + v/c$ is the Doppler shift factor, σ is some attenuation factor due to the propagation of the acoustic wave, δ is a propagation delay, and c is the wave propagation speed. Equation (1) is also valid for sources moving *away* from the receiver: in that case $v < 0$ and, consequently, $\alpha < 1$. It is straightforward to see that $y_c(t)$ is also a Gaussian zero-mean stationary random process with covariance function

$$R_{y_c}(\tau) = E[y_c(t)y_c(t-\tau)] = \sigma^2 R_{x_c}(\alpha\tau), \tag{2}$$

where $R_{x_c}(\tau)$ denotes the covariances function of the stationary process $x_c(t)$. Equivalently, the power spectral densities (PSD) of the processes $x_c(t)$ and $y_c(t)$ are related by

$$S_{y_c}(\omega) = \frac{\sigma^2}{\alpha} S_{x_c}(\frac{\omega}{\alpha}) \tag{3}$$

where $S_{x_c}(\omega)$ and $S_{y_c}(\omega)$ denote the PSD of $x_c(t)$ and $y_c(t)$, respectively.

In practice, we only have access to sampled versions, $x[n] = x_c(nT_s)$ and $y[n] = y_c(nT_s + \kappa)$, $n \in \mathbb{N}$, of the continuous-time processes $x_c(t)$ and $y_c(t)$; κ is some arbitrary time instant, and $\omega_s = 2\pi/T_s$ is the radian sampling

0-8186-7576-4/96 $5.00 © 1996 IEEE

frequency. Both discrete-time random processes are Gaussian and zero-mean, with covariance sequences given by

$$R_x[k] = R_{x_c}(kT_s), \qquad R_y[k] = \sigma^2 R_{x_c}(\alpha k T_s). \quad (4)$$

Their PSD's are related to the PSD's of their continuous-time counterparts by

$$S_i(\Omega) = \frac{1}{T_s}\sum_{k=-\infty}^{\infty} S_{i_c}(\frac{\Omega + 2k\pi}{T_s}), \qquad i = x, y\,, \quad (5)$$

where $\Omega \in [-\pi, \pi]$ is the normalized radian frequency.

The effect of the Doppler shift can be viewed as a change of sampling period with respect to the original signal $x(t)$ from T_s to αT_s. It is clear from (5) that the Nyquist condition (non-aliasing) for $y[n]$ is that $x_c(t)$ must be band-limited to $W < \omega_s/2\alpha$. If this condition holds, then we have

$$S_y(\Omega) = \frac{\sigma^2}{\alpha} S_x(\frac{\Omega}{\alpha}) = \frac{\sigma^2}{\alpha T_s} S_{x_c}(\frac{\Omega}{\alpha T_s}). \quad (6)$$

3. Rational Modeling

Let us assume that the PSD of $x_c(t)$ is rational (i.e., $x_c(t)$ is the output of a linear filter excited by white Gaussian noise),

$$S_{x_c}(\omega) = \eta \left|H_{x_c}(j\omega)\right|^2 = \eta \left|\frac{P(j\omega)}{Q(j\omega)}\right|^2, \quad (7)$$

where $P(s)$ and $Q(s)$ are polynomials in s. Let

$$H_{x_c}(s) = \sum_{k=1}^{p} \frac{A_k}{s - s_k} \quad (8)$$

be the partial fraction expansion $H_{x_c}(s)$, where we have assumed without loss of generality that all the poles of $H_{x_c}(s)$ are simple. Let us further assume that $x_c(t)$ is essentially band-limited[1] to frequency $W < \omega_s/2\alpha$. Then, by the impulse invariant method [6], it is easy to show that

$$S_y(\Omega) \approx \eta\sigma^2 T_s \left|H_\alpha(e^{j\Omega})\right|^2 \quad (9)$$

with

$$H_\alpha(z) = \sum_{k=1}^{p} \frac{A_k}{1 - e^{\alpha s_k T_s} z^{-1}}. \quad (10)$$

Equation (10) leads to an heuristically appealing interpretation of the Doppler effect for rational random processes in terms of pole displacements. Consider, for example, the degenerate case where the poles in the s-plane are purely imaginary, $s_k = j\omega_k$, then, in the z-plane, the Doppler effect rotate the poles on the unit circle by a factor α. This interpretation is consistent with the analysis of the Doppler effect for deterministic harmonic signals. It is interesting to note that, in the general case, the poles in the z-plane move on logarithmic spirals, not on circles.

The rational modeling approach just introduced also suggests an efficient way to artificially synthetize Doppler-shifted processes with rational spectra. If a white Gaussian sequence is used the input of a digital filter with transfert fuction (10), it is straighforward to see that the output sequence of the filter will be Gaussian and have the desired spectral shape. The scaling by σ is trivial. This method allows easy implementation of the synthetizer, even for varying shift α, via the parallel form of the filter defined by (10).

[1] That is, all but a negligible fraction of its energy is contained in the band $[-W, W]$.

4. Estimators of Doppler parameters

Let $\mathbf{y} = (y[0], y[1], \ldots, y[N-1])^t$ be a length N sample of the process $y[n]$. We address the problem of estimating the Doppler parameters α and σ from $\mathbf{y}$ when the statistics of $x_c(t)$ are known (i.e., either $R_{x_c}(\tau)$, $S_{x_c}(\omega)$, or a rational model of $S_x(\Omega)$ is known).

Clearly, there are situations in which it will not be possible to identify the Doppler parameters, a trivial example being the white noise case. An identifiability condition for α and σ from $y_c(t)$ is the $\sigma^2 R_{x_c}(\alpha\tau) = \sigma'^2 R_{x_c}(\alpha'\tau)$ implies implies $\alpha = \alpha'$ and $\sigma = \sigma'$. It can be shown that if $x_c(t)$ is of finite power the identifiability condition holds. For the discrete process $y[k]$, the identifiability condition becomes $R_y(\alpha, \sigma)[k] = R_y(\alpha', \sigma')[k], \forall k \in \mathbb{N}$, or, alternately, $S_y(\Omega; \alpha, \sigma) \overset{\text{a.e.}}{=} S_y(\Omega; \alpha', \sigma')$, implies $\alpha = \alpha'$ and $\sigma = \sigma'$.

4.1. Maximum-likelihood estimator

A finite length vector $\mathbf{y}$ is realization of a Gaussian zero-mean random variable with covariance matrix $\mathbf{\Sigma} = \mathbf{\Sigma}_y(\theta)$

$$\mathbf{\Sigma}_y(\theta) = \sigma^2 \mathbf{\Sigma}_{x_c}(\alpha) = \sigma^2 \left(R_{x_c}(\alpha|i-j|T_s)\right)_{\substack{1\le i\le N \\ 1\le j\le N}}, \quad (11)$$

with $\theta = \{\alpha, \sigma\}$. Thus, estimating θ given $\mathbf{y}$ can be viewed as a structured covariance estimation problem. The covariance matrix $\mathbf{\Sigma}_y(\theta)$ is linear in σ^2, but will generally be non-linear in α.

The maximum-likelihood estimator $\hat{\theta}_{\text{ML}}$ minimizes [2]

$$L(\theta) = \log|\mathbf{\Sigma}_y(\theta)| + \text{tr}[\mathbf{\Sigma}_y^{-1}(\theta)\mathbf{S}], \quad (12)$$

where $\mathbf{S} = \mathbf{y}\mathbf{y}^t$ is the sample covariance matrix. In general, the minimization of $L(\theta)$ is a non-linear problem, and it is necessary to resort to iterative methods. However, for large N, the computational load involved by the matrix operations in the computation of (12) can be greatly reduced by taking advantage of the Toeplitz structure of $\Sigma_y(\theta)$. It can be shown [3] (see also [4]) that the minimization of $L(\theta)$ is asymptotically equivalent to the minimization of

$$\sum_{k=0}^{N-1} \left[\log S_y(\frac{2\pi k}{N}; \theta) + \frac{I_N(\frac{2\pi k}{N})}{S_y(\frac{2\pi k}{N}; \theta)}\right], \quad (13)$$

where $S_y(\Omega;\theta)$ is the PSD of $y[n]$ considered as a function of θ, $I_N(\Omega) = \frac{1}{2\pi}\sum_{k=0}^{N-1} y[k]e^{j\Omega k}$ is the periodogram spectral estimate computed from $\mathbf{y}$. Once $I_N(\frac{2\pi k}{N})$ has been computed (by FFT if N is power of 2), the evalution of $L(\theta)$ by (13) requires only $O(N)$ operations. The minimization of (13) can be viewed as the minimization of the Itakura-Saito distance between the empirical spectrum and the parametric model of the data. The ML estimator can thus be interpreted as a "spectral matching" estimator based on the Itakura-Saito distance.

If an approximate value or a range for the Doppler shift α is known, it is possible to obtain an approximate ML estimator with a lower complexity. Consider the Taylor series expansion of $\Sigma(\theta)$ around some value of $\alpha = \alpha_0$. For example, for $v \ll c$ we have $\alpha \approx 1$, justifying an expansion about $\alpha_0 = 1$. If we restrict the Taylor series expansion to the first two terms, it is trivial algebra to show that $\mathbf{\Sigma}_y(\theta)$ can be expressed as a linear combination $\mathbf{\Sigma}_y(\theta) \approx \gamma\mathbf{A} + \phi\mathbf{B}$ of two $N \times N$ Toeplitz matrices $\mathbf{A} = (a[i-j])$ and $\mathbf{B} = (b[i-j])$, with

$$\begin{aligned} a[k] &= R_{x_c}(\alpha_0 k T_s) - \alpha_0 b[k], \\ b[k] &= \left.\frac{\partial R_{x_c}(\alpha k T_s)}{\partial \alpha}\right|_{\alpha=\alpha_0}, \end{aligned}$$

and the reparameterization $\theta = \{\gamma, \phi\}$ where $\gamma = \sigma^2$ and $\phi = \sigma^2\alpha$. Even with this linear approximation, the minimization of (12) is still a non-linear problem and still requires an iterative solution, but a simpler one than the original (see [1]).

4.2. Method of moments estimator

The iterative maximum-likelihood methods requires an initial estimate of the parameters. The method of moments (MoM) method can be used to provide such estimate. The MoM estimate $\hat{\theta}_{\text{MoM}}$ is obtained by equating sample moments $\hat{R}_y[k]$ of $\mathbf{y}$ with their theoretical values expressed as functions of θ. Using the first two covariance lags yields

$$\hat{\sigma}_{\text{MoM}} = \sqrt{\frac{\hat{R}_y[0]}{R_{x_c}(0)}}, \quad (14)$$

$$0 = \hat{R}_y[1] - \hat{\sigma}^2_{\text{MoM}} R_{x_c}(\hat{\alpha}_{\text{MoM}} T_s). \quad (15)$$

Equation (15) needs to be solved numerically, unless a Taylor expension similar to that of the previous section is used for $R_{x_c}(\alpha T_s)$. Note that using such a linear expansion is equivalent to performing only the first iteration of a Newton-Raphson algorithm for the solution of (15). The existence of a solution to (15) is guaranteed by the "maximization at 0" property of covariance functions and of covariances sequences, if $R_{x_c}(\tau)$ is continuous and in L_2. Note that the MoM method requires a stronger identifiability condition than the one introduced earlier, i.e., that $R_{x_c}(\alpha T_s) = R_{x_c}(\alpha' T_s)$ implies $\alpha = \alpha'$. If, in addition, $R_{x_c}(\tau)$ (and, hence, $R_y[k]$ viewed as a function of α) is continuous and differentiable, the MoM estimate can be shown to be consistent by Theorem 3.14. in [7].

To relax this strict identifiability condition, an alternate moment-based estimate can be obtained by equating a set of K moments, $K < N-1$, and then solving in the least-square sense:

$$\hat{\theta}_{\text{LSMoM}} = \arg\min_{\theta} \sum_{k=0}^{K-1} \left\| \hat{R}_y[k] - R_y(\theta)[k] \right\|^2. \quad (16)$$

Under an identifiability condition for the first K lags, the LSMoM estimator can be shown to be consistent (by Theorem 3.14 in [7] again). From Parseval's theorem, it is not difficult to see that, for large K, (16) is equivalent to minimization the L_2 distance between the theoretical PSD $S_y(\Omega;\theta)$ and a windowed periodogram spectral estimate. If the linear approximation of the covariance

$$R_y(\theta)[k] \approx \gamma a[k] + \phi b[k] \quad (17)$$

is used, we obtain a linear least-square problem. Equivalently, linearization of the PSD $S_y(\Omega;\theta)$ could be used with a L_2 "spectral matching" criterion.

4.3. Rational modeling estimator

Assume that a rational model (AR or ARMA) is available for $S_x(\Omega)$, let p_k, $k = 1, \ldots, p$ be its poles. Further assume that $x[k]$ arise from a $x_c(t)$ that is essentially bandlimited. An ad-hoc estimator based on the rational modeling approach can be obtained as follows. An AR and ARMA model can be fitted to $\mathbf{y}$ with classical methods (e.g., Yule-Walker). For ARMA models, only the AR part (poles) is of importance. Let $\hat{p}_k$, $k = 1, \ldots, p$, be these poles. From (10), it follows that they are related to the original poles by

$$\hat{p}_k \approx e^{\alpha \log p_k}, \quad k = 1, \ldots, p. \quad (18)$$

Minimization of

$$\sum_{k=1}^{p} \left\| \log \hat{p}_k - \alpha \log p_k \right\|^2 \quad (19)$$

with respect to α, which is a trivial linear problem, provides the rational modeling estimator of the Doppler shift $\hat{\alpha}_{\text{RM}}$, and $\hat{\sigma}_{\text{RM}}$ can be taken equal to $\hat{\sigma}_{\text{MoM}}$. Note that, like the ML and LSMoM estimates, the RM estimates can be interpreted as a "spectral matching" estimator with the particular spectral distance defined by (19).

4.4. Cramér-Rao bound

It can be shown that the Cramér-Rao bound (CRB) for the estimation of α and σ is given by

$$\text{CRB}(\alpha) = 2\left[\|\mathbf{\Psi}\|_F^2 - N^{-1}(\operatorname{tr}\mathbf{\Psi})^2\right]^{-1} \quad (20)$$

$$\text{CRB}(\sigma) = \frac{\sigma^2}{4N}\|\mathbf{\Psi}\|_F^2 \,\text{CRB}(\alpha), \quad (21)$$

with $\Psi = \Sigma_{x_c}^{-1}(\alpha)\frac{\partial \Sigma_{x_c}(\alpha)}{\partial \alpha}$.

If σ is known, Whittle's asymptotic version of the CRB for α takes the form

$$\mathrm{CRB}(\alpha) \asymp N^{-1}\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{\frac{\partial S_y(\Omega)}{\partial \alpha}}{S_y(\Omega)}\right)^2 d\Omega \qquad (22)$$

which has a nice intuitive interpretation. The CRB depends on the "sensitivity" of the PSD of $y[n]$ to variations in α, this sensitivity beeing weighted by the PSD itself (i.e., the energy repartition in the spectral domain).

5. Preliminary results

In order to evaluate the performance of the proposed estimators, we conducted several Monte-Carlo simulations. We used a 4th order linear model for the continuous-time process $x(t)$, i.e.,

$$H_{x_c}(s) = \frac{1}{10^{-17}s^4+3\times10^{-14}s^3+7.02\times10^{-9}s^2+1.2\times10^{-5}s+1}.$$

The sampling frequency F_s was set equal to 40860 Hz. The process was simulated via the rational-modeling approach described in Sec. 3.

Monte-Carlo simulations have been performed in Matlab, with 1000 independent runs for each estimator. Figure 2 summarizes the results obtained for the MoM estimator, the rational-modeling estimator, and the LSMoM estimator, of the Doppler shift α. For the rational-modeling method, the least-square modified Yule-Walker algorithm [7] has been used to compute the poles of the rational model. Five covariances lags were used in the LSMoM method with the linearization of the covariance (17) about $\alpha_0 = 1$. The ML estimator suffered from numerical convergence problems and its results are not included here.

Of the three estimators, the linearized LSMoM estimator has the lowest computational cost, followed by the rational-modeling estimator, and the MoM estimator (because of the numerical resolution of (15)). For samples of moderate and large size N, the rational modeling estimator of the Dopple shift *alpha* outperforms the two moment-based estimators which have very close variances. For the estimation of the gain σ, all three methods offered very similar performances. More complete results, including performance on real Doppler-shifted acoustic data will be presented at the conference.

6. Concluding remarks

Various estimators of the Doppler parameters for Doppler-shifted Gaussian random processes have been proposed. All these estimator rely on a stationarity assumption of the Gaussian process. In the context of Doppler-shift for moving sources, this amounts to assuming that the wave source is moving at a constant radial toward (or away from) the fixed receiver. If the source is not moving at constant radial speed, the random process $y(t)$ is no longer stationary. The approach proposed here can be extended to treat this case. For example, an adaptive ARMA modeling technique could be easily combined with the rational modeling approach to track the evolution of the Doppler parameters in time. Likewise, the ML and moment-based methods (or their "spectral matching" equivalent) could be applied to a sliding window of adequate length so as to insure quasi-stationary of the signal. The resulting estimates of the Doppler shift $\alpha(t)$ and gain $\sigma(t)$ could be further smoothed by using a priori knowledge on the movement of the source, if is such knowledge is available.

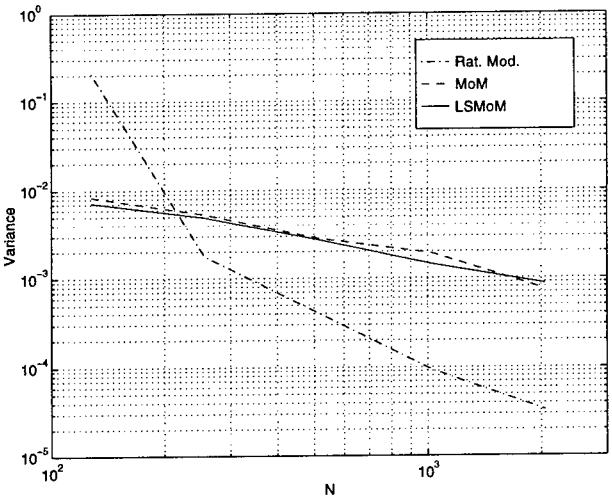


Figure 2. Variance of the estimators of the Doppler shift α.

Another approach would be to use a non-stationary representation of the signal. An appealing candidate for this non-stationary representation is the wavelet transform (as suggested in [8]): moving noise sources result in dilatations of the time axis and propagation delays, which are exactly the operations (translation and dilatation) used in the definition of the wavelet transform.

References

[1] T. W. Anderson. Asymptotically efficient estimation of covariance matrices with linear structure. *The Annals of Statistics*, 1(1):135–141, 1973.

[2] J. P. Burg, D. G. Luenberger, and D. L. Wenger. Estimation of structured covariance matrices. *Proceedings of the IEEE*, 70(9):963–974, Sept. 1982.

[3] K. Dzhaparidze. *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Times Series*. Springer-Verlag, New York, 1986.

[4] R. M. Gray. Toeplitz and circulant matrices: A review. Technical report, Stanford University, Stanford, CA, Apr. 1993.

[5] P. Morse and K. Ingard. *Theoretical Acoustics*. McGraw-Hill, New York, 1968.

[6] A. V. Oppenheim and R. W. Schaffer. *Discrete Time Signal Processing*. Prentice-Hall, Englewood Cliff, N.J., 1989.

[7] B. Porat. *Digital Processing of Random Signals, Theory and Methods*. Prentice-Hall, Englewood Cliff, N.J., 1994.

[8] L. H. Sibul. Application of wavelet transform techniques for imaging wideband acoustic moving sources. In *Proceedings INTER-NOISE '95*, pages 1157–1162, Newport Beach, CA, July 1995.

Estimation of Multiple Target Trajectories with Time Varying Amplitudes*

S.M. Tonissen and A. Logothetis

Dept. of Electrical and Electronic Eng.
University of Melbourne
Parkville, Victoria 3052
Australia
s.tonissen@ee.mu.oz.au a.logothetis@ee.mu.oz.au

Abstract

This paper is concerned with the estimation of multiple constant velocity target trajectories in a low SNR environment. Each target trajectory is characterised by the initial position and velocity, which are to be estimated. A major difficulty is that the target amplitudes are unknown and will in general be time varying. The approach in this paper is to model the target amplitude as an autoregressive (AR) process. A maximum likelihood estimator is derived for the parameters of the AR process and the unknown target position and velocity using the expectation conditional maximisation (ECM) algorithm.

1. Introduction

This paper is concerned with the estimation of the trajectories of multiple constant velocity targets observed using an optical sensor, recording 2-D images. In a low SNR environment, target locations cannot be estimated using a single image, so a number of frames must be recorded and processed. At the end of this observation interval estimates of all target trajectories are obtained.

Previously the single target problem has been formulated in [1, 2, 3, 4] as either a maximum likelihood estimation problem, or frequency domain matched filter problem. The amplitude is either assumed constant [1, 2, 3], or completely unknown and therefore uncorrelated from pixel to pixel [4]. In addition, a discrete set of candidate target velocities is tested, resulting in performance loss in the presence of a mismatch between assumed and actual target velocity.

In general the target amplitude is time varying. Recently in [5] a first order model is proposed for the time varying amplitude, but the mean, variance and correlation from one time to the next are assumed known *a-priori*. A numerical optimisation procedure is used to obtain continuous estimates of target position and velocity, overcoming the problem of mismatch.

This paper extends the formulation in [5] to account for multiple targets. The time varying amplitudes of the targets are modeled as independent first order autoregressive (AR) process, similar to [5], but the AR process parameters are assumed unknown *a-priori*. A maximum likelihood estimator, using the expectation conditional maximisation (ECM) algorithm, is derived to simultaneously estimate the parameters of the AR processes and the unknown target positions and velocities.

2. Problem Formulation

2.1. Signal Model

There are N constant velocity targets, with time varying amplitude. For simplicity, N is assumed known *a-priori*, but in practice would also need to be estimated. The time varying amplitude for the n^{th} target, denoted $\xi^{(n)}$ is modeled as a first order AR series, given by

$$\xi^{(n)}(k)-a^{(n)}(k) = d^{(n)}(\xi^{(n)}(k-1)-a^{(n)}(k-1))+v^{(n)}(k) \quad (1)$$

*This work was supported by the Co-operative Research Centre for Sensor Signal and Information Processing (CSSIP). A. Logothetis is supported by the Australian Telecommunications and Electronics Research Board (ATERB)

0-8186-7576-4/96 $5.00 © 1996 IEEE

where $a^{(n)}(k) = a^{(n)}, \forall k$ is the constant mean amplitude of the n^{th} source (target). The $v^{(n)}, n \in \{1,\ldots,N\}$ are statistically independent zero-mean white Gaussian sequences with variances $\sigma^2_{v^{(n)}}$, so the target amplitude series are independent. The inter-frame amplitude correlation is given by $d^{(n)}$.

The measurements are images recorded using an optical sensor array consisting of $C \times D$ resolution cells (pixels) of dimension $\Delta x \times \Delta y$. The measurement at time k is $Y(k) \in \mathrm{R}^{C\times D}$, the set of observations recorded in all cells at time k. The signal from a point source is spread according to a point spread function, approximated here by a 2-D Gaussian. If $(x_0^{(n)}, y_0^{(n)}, v_x^{(n)}, v_y^{(n)})$ represents the initial position (in x and y) and initial velocity (in x and y) for the n^{th} source, then the signal in cell i,j at time k due to source n is $h_{ij}^{(n)}(k)\xi^{(n)}(k)$, where

$$h_{ij}^{(n)}(k) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{(x_0^{(n)} + kTv_x^{(n)} - i\Delta x)^2}{2\sigma_x^2}\right)\cdot \exp\left(-\frac{(y_0^{(n)} + kTv_y^{(n)} - j\Delta y)^2}{2\sigma_y^2}\right) \quad (2)$$

and σ_x and σ_y are characteristic for the sensor and assumed known, and T is the sampling interval.

The measurement in cell i,j at time k is given by the weighted sum of the time varying amplitudes embedded in white noise, given by

$$y_{ij}(k) = \sum_{n=1}^{N} h_{ij}^{(n)}(k)\xi^{(n)}(k) + e_{ij}(k) \quad (3)$$

where $e_{ij}(k)$ are statistically independent zero-mean white Gaussian processes with known variance σ_e^2. Eqs (1) and (3) can be written in a state space form as follows

$$\begin{aligned} x(k) &= Ax(k-1) + Bv(k) \\ z(k) &= G(k)x(k) + w(k) \end{aligned} \quad (4)$$

where

$$\begin{aligned} x(k) &= \mathrm{vec}\left(x^{(1)}(k),\ldots,x^{(N)}(k)\right) \\ x^{(n)}(k) &= \begin{pmatrix} \xi^{(n)}(k) \\ \xi^{(n)}(k-1) \\ a^{(n)}(k) \end{pmatrix} \\ A &= \mathrm{diag}(A^{(1)},\ldots,A^{(N)}) \\ A^{(n)} &= \begin{pmatrix} d^{(n)} & 0 & 1-d^{(n)} & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\ B &= \mathrm{diag}(B^{(1)},\ldots,B^{(N)}) \\ B^{(n)} &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ v(k) &= \mathrm{vec}\left(v^{(1)}(k),\ldots,v^{(N)}(k)\right) \\ v^{(n)}(k) &\sim N(0,\sigma^2_{v^{(n)}}) \\ z(k) &= \mathrm{vec}(Y(k)) \\ Y(k) &= [y_{ij}(k)] \\ G(k) &= \left(\mathrm{vec}(H^{(1)}(k)),\ldots,\mathrm{vec}(H^{(N)}(k))\right)B' \\ H^{(n)}(k) &= [h_{ij}^{(n)}(k)] \\ w(k) &= \mathrm{vec}(E(k)) \\ E(k) &= [e_{ij}(k)] \end{aligned} \quad (5)$$

and $x(k)$ is of dimension $3N \times 1$, A is $3N \times 3N$, B is $3N \times N$, $v(k)$ is $N \times 1$, $z(k)$ is $CD \times 1$, $G(k)$ is $CD \times 3N$ and $w(k)$ is $CD \times 1$.

The total number of images (observations) recorded is K, and the observation sequence $(z(1),\ldots,z(K))$ is denoted Z_K. The state sequence $(x(1),\ldots,x(K))$ is denoted X_K.

2.2. Estimation Objective

Let $\theta_0 = (\theta_1, \theta_2, \ldots, \theta_S) \in \Theta$ denote the true model, such that $\theta_1 = (\sigma^2_{v^{(1)}},\ldots,\sigma^2_{v^{(N)}})$, $\theta_2 = (d^{(1)},\ldots,d^{(N)})$ and θ_3 to θ_S represent $(x_0^{(1)}, y_0^{(1)}, v_x^{(1)}, v_y^{(1)})$ to $(x_0^{(N)}, y_0^{(N)}, v_x^{(N)}, v_y^{(N)})$. The value of S depends on how the parameter space $\Theta \subseteq \mathrm{R}^{6N}$ is partitioned, and $S \in \{3,\ldots,4N+2\}$.

Given a realization $\{z(1),\ldots,z(K)\}$, of the stochastic model of (4), the objective is to obtain the maximum likelihood (ML) estimate

$$\hat{\theta}^{ML} \triangleq \arg\max_{\theta} p(Z_K|\theta) \quad (6)$$

where $p(Z_K|\theta)$ denotes the marginal density of the observations Z_K conditioned on the model θ.

3. Proposed Algorithm

The ECM algorithm [6] is used to obtain $\hat{\theta}^{ML}$, meeting the objective in Section 2.2. As a by-product of the ECM, MAP estimates of the states $x(k)$ are also obtained.

The ECM algorithm is an extension of the expectation maximisation (EM) algorithm [7] and is an iterative method of extracting the mode of the likelihood function. From an initial estimate $\theta^{\{0\}} \in \Theta$ a sequence of estimates $\{\theta^{\{l\}}\}$ are generated. Let $F = \{f_s(\theta) : s = 1,\ldots,S\}$ be the set of S constraint

functions of θ, with $f_s(\theta) = (\theta_1, \ldots, \theta_{s-1}, \theta_{s+1}, \ldots, \theta_S)$. At the $(l+1)^{\text{st}}$ iteration ("pass") of the ECM algorithm we perform the following steps:

E-step: Just like the EM, we evaluate the conditional expectation of the log likelihood of the complete data:

$$Q(\theta, \theta^{\{l\}}) = E\{\ln p(Z_K, X_K|\theta)|Z_K, \theta^{\{l\}}\}. \quad (7)$$

Here p is the density function of the complete data $M_K = (Z_K, X_K)$ and $\theta^{\{l\}}$ is the parameter estimate at the l^{th} iteration.

CM-steps: For $s = 1, \ldots, S$ find $\theta^{\{l+s/S\}}$ that maximises $Q(\theta, \theta^{\{l\}})$ (as a function of θ) subject to the constraint $f_s(\theta) = f_s(\theta^{\{l+(s-1)/S\}})$, ie

$$Q(\theta^{\{l+s/S\}}, \theta^{\{l\}}) \geq Q(\theta, \theta^{\{l\}}),$$
$$\text{for all } \theta \in \Theta \text{ for which } f_s(\theta) = f_s(\theta^{\{l+(s-1)/S\}}) \quad (8)$$

where

$$\theta^{\{l+s/S\}} = (\theta_1^{\{l+1\}}, \ldots, \theta_{s-1}^{\{l+1\}}, \theta_s^{\{l+1\}}, \theta_{s+1}^{\{l\}}, \ldots, \theta_S^{\{l\}})$$

The appealing property of the ECM algorithm is that likelihoods increase monotonically, i.e., $p(Z_K|\theta^{\{l+1\}}) \geq p(Z_K|\theta^{\{l\}})$ with equality holding at the ML estimate, provided the set F of constraints spans the parameter space [6]. The rate of convergence of this algorithm is studied in [8].

3.1. E-Step

The evaluation of $Q(\theta, \theta^{\{l\}})$ requires the density function of the complete data, $p(Z_K, X_K|\theta)$. From the model in (4)

$$\begin{aligned} p(Z_K, X_K|\theta) &= \prod_{k=1}^{K} p(z(k)|x(k), \theta) \cdot \\ &\quad \prod_{k=1}^{K} p(x(k)|x(k-1), \theta) \\ &= \prod_{k=1}^{K} \prod_{i=1}^{C} \prod_{j=1}^{D} p(y_{ij}(k)|x(k), \theta) \cdot \\ &\quad \prod_{k=1}^{K} \prod_{n=1}^{N} p(x^{(n)}(k)|x^{(n)}(k-1), \theta) \end{aligned} \quad (9)$$

where the conditional densities of $x^{(n)}(k)$ and $y_{ij}(k)$ are Gaussian and obtained from (1) and (3), and the definition of $x^{(n)}(k)$. Taking the conditional expectation of the log of (9) gives $Q(\theta, \theta^{\{l\}})$. This requires the computation of $\overline{x(k)}^{\{l\}}$ and $\overline{x(k)x(k)'}^{\{l\}}$ (where $\overline{(\cdot)}^{\{l\}} \triangleq E\{\cdot|Z_K, \theta^{\{l\}}\}$) which are obtained using a Kalman smoother [9]. Due to space constraints, $Q(\theta, \theta^{\{l\}})$ is not shown in full in this summary.

3.2. CM-Steps

The constrained maximisation of $Q(\theta, \theta^{\{l\}})$ consists of the following S steps:

CM-Step 1: Calculate $\theta^{\{l+1/S\}}$ using (8). This step determines ${\sigma^2_{v^{(n)}}}^{\{l+1\}}$, $n = 1, \ldots, N$. A closed form solution is obtained for each of the ${\sigma^2_{v^{(n)}}}^{\{l+1\}}$ by differentiating $Q(\theta^{\{l+1/S\}}, \theta^{\{l\}})$ with respect to $\sigma^2_{v^{(n)}}$ and setting the derivative equal to zero.

CM-Step 2: Calculate $\theta^{\{l+2/S\}}$ using (8). This step determines ${d^{(n)}}^{\{l+1\}}$, $n = 1, \ldots, N$. A closed form solution for each ${d^{(n)}}^{\{l+1\}}$ is obtained in the same way as for CM-Step 1.

CM-Steps 3 to S: Calculate $\theta^{\{l+s/S\}}$, $s = 3, \ldots, S$ using (8). In this step the target position and velocity parameters are updated. In general there will be no closed form solution, so the s^{th} CM-step will involve some form of iterative search in a parameter space $\Theta' \subset \Theta$. The dimensionality of Θ' is determined by the dimensionality of θ_s, and is less than that of Θ. In this paper the parameter space is partitioned such that $S = N + 2$ and each $\theta_s, s = 3, \ldots, S$ corresponds to the initial position and velocity parameters for one target. This definition means each parameter update requires a 4-dimensional search, which is implemented using a gradient descent technique. This results in more complex implementation, but faster convergence, than if each θ_s corresponded to a single position or velocity parameter.

4. Simulation

The procedure has been implemented for the following scenario: The measurement sensor consists of $C \times D = 20 \times 20$ resolution cells of dimension $\Delta x = \Delta y = 1$ cell and $\sigma_x = \sigma_y = 0.6$, with a measurement noise variance of $\sigma_e^2 = 1.0$. The total number of frames processed is $K = 100$. There are $N = 3$ targets moving through the field of view, with amplitude statistics, inital positions and constant velocities as given in Table 1. For these trajectories there will be a number of frames for which there is signal from more than one target in some cells. The approximate average SNRs for the cells in which the targets are located are 1.5 dB, 0.6 dB, and 1.8 dB respectively.

The ECM algorithm was initialised randomly around the true position and velocity parameters, and with $a^{(n)} = 0.0$, $\sigma^2_{v^{(n)}} = 0.25$ and $d^{(n)} = 0.0$. The estimated amplitude statistics and target initial positions

n	$a^{(n)}$	$\sigma^2_{v^{(n)}}$	$d^{(n)}$	$x_0^{(n)}$	$y_0^{(n)}$	$v_x^{(n)}$	$v_y^{(n)}$
1	2.5	5.00	0.7	1.25	1.35	0.17	0.16
2	3.0	0.00	1.0	3.05	12.55	0.14	-0.05
3	2.0	8.00	0.0	10.00	16.50	0.0	-0.15

Table 1. True target parameters

and velocities obtained from a single run are given in Table 2. The recovered trajectories, along with the true trajectories are shown in Figure 1.

n	$\hat{a}^{(n)}$	$\hat{\sigma}^2_{v^{(n)}}$	$\hat{d}^{(n)}$	$\hat{x_0}^{(n)}$	$\hat{y_0}^{(n)}$	$\hat{v_x}^{(n)}$	$\hat{v_y}^{(n)}$
1	2.78	5.34	0.76	1.17	1.30	0.17	0.16
2	2.98	0.30	0.15	3.21	12.35	0.14	-0.05
3	1.67	9.45	0.05	9.96	16.51	0.00	-0.15

Table 2. Estimated target parameters

The target amplitude parameters and the position and velocity parameters are recovered accurately, even for low SNR targets with crossing trajectories. The correlation (d) for target 2 is difficult to recover since a constant amplitude target can be represented by any d, with $\sigma_v^2 = 0$.

The rate of convergence of the ECM algorithm deteriorates as the initialisation of position and velocity parameters moves further from the true values. Some form of grid search is required to provide adequate initialisation. The rate of convergence also deteriorates as the SNR decreases. For this example about 30 passes of the ECM algorithm were required.

5. Conclusion

This paper presents a technique to estimate the trajectories of multiple constant velocity targets with time varying amplitudes, observed with an optical sensor. The ECM algorithm is used to obtain the ML estimate of the target trajectories and MAP estimates of target amplitudes. The technique has been applied successfully to crossing targets in low SNR conditions.

References

[1] N.C. Mohanty. Computer tracking of moving point targets in space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp 606–611, Sep. 1981.

[2] I. Reed, R. Gagliardi, and L. Stotts. Optical moving target detection with 3-D matched filtering. *IEEE Transactions on Aerospace and Electronic Systems*, pp 327–336, Jul. 1988.

[3] I. Reed, R. Gagliardi, and L. Stotts. A recursive moving-target-indication algorithm for optical image sequences. *IEEE Transactions on Aerospace and Electronic Systems*, pp 434–440, May 1990.

[4] Stephen C. Pohlig. Spatial-temporal detection of electro-optic moving targets. *IEEE Transactions on Aerospace and Electronic Systems*, pp 608–616, Apr. 1995.

[5] S.M. Tonissen and Y. Bar-Shalom. Maximum likelihood track-before-detect with fluctuating target amplitude. *Submitted to IEEE Transactions on Aerospace and Electronic Systems*

[6] X.L. Meng and D.B. Rubin. Maximum likelihood estimation via the ECM algorithm: a general framework, *Biometrica*, Vol 80, pp. 267-278, 1993.

[7] A.P. Dempster, N.M Laird, D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm, *J. Royal Stat. Soc.*, Ser 39, Vol 6, pp. 1-38, 1977.

[8] X.L. Meng. On the rate of convergence of the ECM algorithm, *The Annals of Statistics*, Vol 22, No 1, pp. 326-339, 1994.

[9] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*, Englewood Cliffs, NJ, Prentice-Hall 1979.

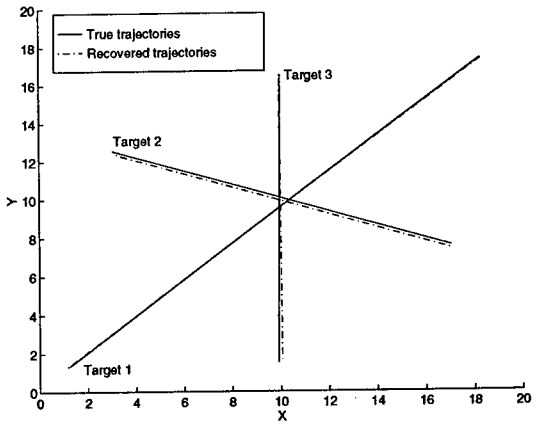


Figure 1. True and recovered trajectories

A Frequency-Domain Solution to Unbiased Equation Error System Identification and Reduced Order Approximation

Jitendra K. Tugnait
Dept. of Electrical Engineering
Auburn University
Auburn, Alabama 36849, USA
tugnait@eng.auburn.edu

ABSTRACT

Estimation of parametric input-output (IO) infinite impulse response (IIR) transfer function is considered. Some of the desirable properties of any approach to this problem are: unimodality of the performance surface, consistent identification in the sufficient-order case, and stability of the fitted model under undermodeling. Some of the well-known approaches fail to satisfy one or more of these properties. The time-domain equation error method (EEM) yields a unimodal performance surface, biased estimates in colored noise and sufficient-order case, and stable fitted models under undermodeling if the input is autoregressive. In this paper we propose a frequency-domain solution to the least-squares equation error identification problem using the power spectrum and the cross-spectrum of the IO data to estimate the IO parametric transfer function. The proposed approach is shown to yield a unimodal performance surface, consistent identification in colored noise and sufficient-order case, and stable fitted models under undermodeling for arbitrary stationary inputs so long as they are persistently exciting of sufficiently high order.

1 Introduction

Consider the following widely used input-output linear system model:

$$y(t) = H(q^{-1})u(t) + v(t) \qquad (1-1)$$

where $\{u(t)\}$ is the measured input sequence, t is discrete-time, $\{y(t)\}$ is the noisy output, and $\{v(t)\}$ is a measurement noise (disturbance) sequence. With q^{-1} denoting the backward-shift operator (i.e. $q^{-1}u(t) = u(t-1)$), the linear system $H(q^{-1})$ represents an IIR (infinite impulse response) system:

$$H(q^{-1}) = \sum_{i=0}^{\infty} h(i)q^{-i}. \qquad (1-2)$$

Given an input-output record $\{u(t), y(t), \quad t = 1, 2, \cdots\}$, but the underlying true system model $H(q^{-1})$ *unknown*, it is of much interest in signal processing, communications and control applications to fit a rational function model

$$G(q^{-1}) := \frac{B(q^{-1})}{A(q^{-1})} = \frac{\sum_{i=0}^{n_b} b_i q^{-i}}{1 + \sum_{i=1}^{n_a} a_i q^{-i}} \qquad (1-3)$$

This work was supported by the National Science Foundation under Grant ECS-9504878.

to given input-output record [1]-[6],[8]. A wide variety of approaches exist [1],[4],[5],[8].

In any model fitting and parameter estimation problem, key issues influencing the choice of the approach are [1],[4],[5],[8]:

(i) *Global Convergence:* Unimodality of the cost function. Does there exist a unique global asymptotic convergence point? For instance, the prediction error method (PEM) and the output error method (OEM) [4],[5] do not have a unimodal cost function, in general, whereas the equation error method (EEM) [1], the Steiglitz-McBride method (SMM), and the instrumental variable method (IVM) [4],[5],[8] all have a unique global asymptotic convergence point.

(ii) *Consistency:* If the *model set* (i.e. the set from which the fitted model is selected) contains the true system (the so-called *sufficient order* case), does the fitted model asymptotically converge to the true model? Ignoring the lack of unimodality, PEM is consistent under a broad class of conditions [4],[5] and so is IVM, but SMM and OEM are so only for white measurement noise and EEM (as modified in [1]) has similar limitations.

(iii) *Statistical Efficiency:* What is the variance (and bias) of the fitted parameters? If it converges to the correct solution, PEM is known to yield the smallest variance [4],[5].

(iv) *Reduced-Order Modeling (Undermodeling):* When the true system does not belong to the model set (for instance, suppose that $H(q^{-1})$ is not a rational function), it is meaningless to talk of consistency. From a practical viewpoint, a key issue now is if the fitted model is stable like the underlying true model? It turns out that only EEM leads to a reduced-order stable model provided that the input $\{u(t)\}$ is an AR process. Indeed, it is noted in [1] that "... if the input can not be ascertained autoregressive, equation error methods should perhaps not be used."

The main objective of this paper is to provide a frequency-domain solution to the problem of equation error (least-squares) system identification using spectral analysis. The proposed method is shown to lead to a unimodal performance surface, consistent identification in col-

0-8186-7576-4/96 $5.00

ored noise and sufficient-order case, and stable fitted models under undermodeling for arbitrary stationary inputs so long as they are persistently exciting of sufficiently high order.

2 Model Assumptions

We impose the following conditions on (1-1):

(AS1) $\{u(t)\}$ and $\{y(t)\}$ are zero-mean and jointly stationary. The power spectral density $S_{uu}(e^{j\omega})$ of $\{u(t)\}$ is > 0 for almost all $\omega \in [0, \pi]$.

(AS2) The true system transfer function $H(q^{-1})$ is causal and stable. Therefore, $\sum_{i=0}^{\infty} h^2(i) < \infty$.

(AS3) The noise sequence $\{v(t)\}$ is zero-mean, stationary and independent of $\{u(t)\}$.

(AS4) The following summability conditions hold true:

$$\sum_{\tau_1,\cdots,\tau_{k-1}=-\infty}^{\infty} [1+|\tau_j|]\,|C_{z_1 z_2 \cdots z_k}(\tau_1,\cdots,\tau_{k-1})| < \infty,$$

for each $j = 2,\cdots,k-1$ and each $k = 2,3,\cdots$ where $z_i(t) \in \{y(t), u(t), v(t)\}$ and $C_{z_1 z_2 \cdots z_k}(\tau_1,\cdots,\tau_{k-1})$ is the k-th order joint cumulant of the random variables $\{z_1(t+\tau_1),\cdots,z_{k-1}(t+\tau_{k-1}), z_k(t)\}$.

Let the vector of unknown parameter be given by

$$\theta = [a_1 \ \cdots \ a_{n_a} \ b_0 \ \cdots \ b_{n_b}]^T. \qquad (2-1)$$

3 A Frequency-Domain Solution

Consider the cross-spectral density

$$S_{yu}(\omega) = \sum_{k=-\infty}^{\infty} E\{y(t+k)u(t)\}e^{-j\omega k}. \qquad (3-1)$$

It then follows easily that

$$H(e^{j\omega}) = H(q^{-1})|_{q=e^{-j\omega}} = \frac{S_{yu}(\omega)}{S_{uu}(\omega)}. \qquad (3-2)$$

The basic approach to model parameter estimation consists of two steps. First obtain a consistent estimator $\widehat{H}(e^{j\omega})$ of $H(e^{j\omega})$ via consistent estimators $\widehat{S}_{yu}(\omega)$ and $\widehat{S}_{uu}(\omega)$ of $S_{yu}(\omega)$ and $S_{uu}(\omega)$, respectively, based upon the input-output record $\{u(t), y(t),\ t = 1,2,\cdots,T\}$. Next estimate the system parameters using the estimated transfer function matrix as "data."

3.1 Transfer Function Estimator and Its Statistics

This involves little more than estimation of cross-spectrum between $\{y(t)\}$ and $\{u(t)\}$, and of power spectrum of $\{u(t)\}$. Numerous techniques are available for this purpose; see [7] and references therein. We will follow the approach of smoothing in frequency domain [7, Sec. 7.4]. Given a record of length T, let $Y(\omega)$ denote the DFT of $\{y(t),\ 1 \le t \le T\}$ given by

$$Y(\omega_k) = \sum_{t=0}^{T-1} y(t+1)\exp(-j\omega_k t) \qquad (3-3)$$

where

$$\omega_k = \frac{2\pi}{T}k, \quad k = 0,1,\cdots,T-1. \qquad (3-4)$$

Similarly define $U(\omega_k)$.

Given the above DFT's, following [7, Sec. 7.4] we define the cross- and auto-spectrum estimators as

$$\widehat{S}_{yu}(k) = \frac{2\pi}{T^2}\sum_{s=1}^{T-1} W^{(T)}(\omega_{k-s})Y(\omega_{k-s})U^*(\omega_{k-s}) \quad (3-5)$$

and

$$\widehat{S}_{uu}(k) = \frac{2\pi}{T^2}\sum_{s=1}^{T-1} W^{(T)}(\omega_{k-s})U(\omega_{k-s})U^*(\omega_{k-s}) \quad (3-6)$$

for $1 \le k \le T-1$, where the scalar weighting function $W^{(T)}(\alpha)$ is given by

$$W^{(T)}(\alpha) = B_T^{-1}\sum_{i=-\infty}^{\infty} W\left(B_T^{-1}[\alpha + 2\pi i]\right) \qquad (3-7)$$

such that $W(\beta)$, $-\infty < \beta < \infty$, is real-valued, even, of bounded variation satisfying $\int_{-\infty}^{\infty} W(\beta)\,d\beta = 1$ and $\int_{-\infty}^{\infty} |W(\beta)|\,d\beta < \infty$ [7, Secs. 5.6 and 7.4]. It is convenient to take $W(\beta) = 0$ for $|\beta| > 2\pi$ and $W(\beta) = (4\pi)^{-1}$ for $|\beta| \le 2\pi$. In this case (3-5) involves uniform weighting of the $2B_T T + 1$ cross-periodogram ordinates whose frequencies fall in the interval $(\omega_k - 2\pi B_T, \omega_k + 2\pi B_T)$. Thus (3-5) reduces to

$$\widehat{S}_{yu}(k) = \frac{1}{2m_T+1}\sum_{i=-m_T}^{m_T} Y(\omega_{k-i})U^*(\omega_{k-i}), \quad (3-8)$$

where $m_T = B_T T$. Similar modification holds for (3-6). Let us choose B_T to be such that as $T \to \infty$, we have $B_T \to 0$ and $B_T T \to \infty$. Let $k_l(T)$ with $T = 1,2,\cdots$ be a sequence of integers such that $\lim_{T\to\infty} k_l(T)/T = f_l$, a fixed frequency (in Hz).

In light of (3-8) define a coarser frequency grid:

$$\omega_l = \frac{2\pi l}{L_T} = \frac{2\pi l(2m_T+1)}{T} = \frac{2\pi l(2B_T T+1)}{T} \quad (3-9)$$

with $l = 0,1,\cdots,L_T - 1$ where $L_T = \lfloor \frac{T}{2m_T+1} \rfloor$. Using the estimated spectra we have an estimator of the system transfer function at frequency ω_k (as in [7, Chapter 8])

$$\widehat{H}(e^{j\omega_k}) = \widehat{S}_{uu}^{-1}(k)\widehat{S}_{yu}(k) \qquad (3-10)$$

provided that $\widehat{S}_{uu}^{-1}(k)$ exists. If $S_{uu}^{-1}(\omega_k)$ exists, then it follows from [7, Thm. 8.11.1] that

$$\lim_{T\to\infty}\widehat{H}(e^{j2\pi f}) = \lim_{T\to\infty}\widehat{S}_{uu}^{-1}(k(T))\widehat{S}_{yu}(k(T)) = H(e^{j2\pi f}) \ \text{ w.p.1} \qquad (3-11)$$

where $\lim_{T\to\infty} k(T)/T = f$. Convergence in (3-11) is uniform in f. Finally, by the asymptotic independence of the periodogram and cross-periodogram on the grid (3-4) for

$0 \le k \le T/2$ (see also [7, Chap. 7]), it follows that $\widehat{H}(e^{j\omega_k})$ for $k = l(2m_T+1)$, $l = 0, 1, \cdots, (L_T/2)-1$, are (asymptotically) independent. It follows from [7, Thm. 8.8.1] that $\widehat{H}(e^{j\omega_k})$ for $k = l(2m_T+1)$, $l = 0, 1, \cdots, (L_T/2)-1$, are (asymptotically) jointly complex (circularly symmetric) Gaussian such that

$$\lim_{T\to\infty} \Delta_T \operatorname{cov}\left(\widehat{H}(e^{j\omega_k}), \widehat{H}(e^{j\omega_k})\right) = \frac{S_{yy}(\omega_k)}{S_{uu}(\omega_k)}\left[1 - \frac{|S_{yu}(\omega_k)|^2}{S_{yy}(\omega_k)S_{uu}(\omega_k)}\right] = \sigma^2(\omega_k), \quad (3-12)$$

$$\lim_{T\to\infty} \Delta_T \operatorname{cov}\left(\widehat{H}(e^{j\omega_k}), \widehat{H}^*(e^{j\omega_k})\right) = 0, \quad (3-13)$$

where

$$\Delta_T = \frac{B_T T}{2\pi \int_{-\infty}^{\infty} W^2(\alpha)\, d\alpha} = 2B_T T \quad \text{(if (3-8) is used)}, \quad (3-14)$$

and $\operatorname{cov}\{X, Y\} = E\{XY^*\} - E\{X\}E\{Y^*\}$. Thus, $\widehat{H}(e^{j\omega_k})$ on the coarse grid (3-9) is asymptotically a complex Gaussian (in the sense of [7, Sec. 4.2]) random variable, independent at distinct frequencies on the coarse grid over $(0, \pi)$, with the covariance structure (3-12).

3.2 An Equation Error Formulation

It follows from the definition of $G(e^{j\omega})$ (cf. (1-3)) that

$$-\sum_{i=1}^{n_a} G(e^{j\omega_k}) a_i e^{-j\omega_k i} + \sum_{i=0}^{n_b} b_i e^{-j\omega_k i} = G(e^{j\omega_k}) \quad (3-15)$$

for any ω_k. We rewrite (3-15) after replacing $G(e^{j\omega_k})$ with the true transfer function estimate $\widehat{H}(e^{j\omega_k})$ (see (3-10)), as

$$-\sum_{i=1}^{n_a} \widehat{H}(e^{j\omega_k}) e^{-j\omega_k i} a_i + \sum_{i=0}^{n_b} e^{-j\omega_k i} b_i = \widehat{H}(e^{j\omega_k}). \quad (3-16)$$

Using frequencies $\omega_k = 2\pi(k-1)/L_T$ for $0 \le k \le L = (L_T/2)-1$, (3-16) may be rewritten in a matrix-equation form and a least-squares solution can be found, as in [3] (but in a different context). One may also wish to split (3-16) into its real and imaginary parts and then solve it in order to preserve the real-valued nature of the parameters (see [3]).

The above least-squares formulation is equivalent to the following formulation. Choose θ to minimize the cost

$$J_{1T}(\theta) := \frac{2}{L_T} \sum_{l=0}^{\frac{L_T}{2}-1} \left| A(e^{j\omega_l};\theta)\widehat{H}(e^{j\omega_l}) - B(e^{j\omega_l};\theta)\right|^2 \quad (3-17)$$

where

$$B(e^{j\omega_l};\theta) = \sum_{i=0}^{n_b} b_i(\theta) e^{-j\omega_l}, \quad (3-18)$$

$$A(e^{j\omega_l};\theta) = 1 + \sum_{i=1}^{n_a} a_i(\theta) e^{-j\omega_l}. \quad (3-19)$$

In order to deduce some desirable theoretical properties it will be convenient to work with (3-17) in the rest of the paper.

Lemma 1. Under (AS1)-(AS4), $\lim_{T\to\infty} J_{1T}(\theta) \overset{w.p.1}{=} J_{1\infty}(\theta)$ uniformly in θ for $\theta \in \Theta_C$, any compact set, where

$$J_{1\infty}(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left| A(e^{j\omega};\theta)H(e^{j\omega}) - B(e^{j\omega};\theta)\right|^2 d\omega\,. \quad (3-20)$$

Proof: By [7, Thm. 8.11.1], convergence in (3-11) is uniform in f. In particular, for ω_l on the grid (3-9), given any $\epsilon > 0$ there exists an integer $N(\epsilon)$ such that

$$\left|\widehat{H}(e^{j\omega_l}) - H(e^{j\omega_l})\right| < \epsilon \quad w.p.1 \quad (3-21)$$

uniformly in ω_l for $T > N(\epsilon)$. Moreover, by [13, Prop. 1.2.16] and (3-21), we also have

$$\left||\widehat{H}(e^{j\omega_l})|^2 - |H(e^{j\omega_l})|^2\right| < \epsilon \quad w.p.1 \quad (3-22)$$

uniformly in ω_l for $T > N(\epsilon)$. Consider

$$\begin{aligned} D_l(\theta) &= |A(e^{j\omega_l};\theta)\widehat{H}(e^{j\omega_l}) - B(e^{j\omega_l};\theta)|^2 \\ &\quad -|A(e^{j\omega_l};\theta)H(e^{j\omega_l}) - B(e^{j\omega_l};\theta)|^2 \\ &= |A(e^{j\omega_l};\theta)|^2\left[|\widehat{H}(e^{j\omega_l})|^2 - |H(e^{j\omega_l})|^2\right] \\ &\quad + A(e^{j\omega_l};\theta)B^*(e^{j\omega_l};\theta)\left[H(e^{j\omega_l}) - \widehat{H}(e^{j\omega_l})\right] \\ &\quad + A^*(e^{-j\omega_l};\theta)B(e^{j\omega_l};\theta)\left[H(e^{j\omega_l}) - \widehat{H}(e^{j\omega_l})\right]^*. \end{aligned} \quad (3-23)$$

By compactness of Θ_C and continuity of $A(e^{-j\omega_l};\theta)$ and $B(e^{-j\omega_l};\theta)$ in θ as well as in ω_l, we have

$$\sup_{\omega_l \in [-\pi,\pi]} \sup_{\theta\in\Theta_C} \left\|A(e^{j\omega_l};\theta)\right\| \le M < \infty \quad (3-24)$$

and

$$\sup_{\omega_l \in [-\pi,\pi]} \sup_{\theta\in\Theta_C} \left\|B(e^{j\omega_l};\theta)\right\| \le M < \infty\,. \quad (3-25)$$

Therefore, by (3-21)-(3-25), given any ϵ_1 $(= 3M^2\epsilon) > 0$, there exists an integer $N(\epsilon_1)$ such that

$$|D_l(\theta)| < \epsilon_1 \quad \text{w.p.1} \quad \forall\theta\in\Theta_C,\ \forall\omega_l,\ \forall T > N(\epsilon_1). \quad (3-26)$$

Now define

$$\overline{J}_{1T}(\theta) := \frac{2}{L_T}\sum_{l=0}^{\frac{L_T}{2}-1} \left|A(e^{j\omega_l};\theta)H(e^{j\omega_l}) - B(e^{j\omega_l};\theta)\right|^2. \quad (3-27)$$

Using (3-17), (3-26) and (3-27) it follows that

$$\left|J_{1T}(\theta) - \overline{J}_{1T}(\theta)\right| \le \frac{2}{L_T}\sum_{l=0}^{\frac{L_T}{2}-1} |D_l(\theta)| < \epsilon_1 \quad \text{w.p.1} \quad (3-28)$$

$\forall \theta \in \Theta_C$ and $\forall T > N(\epsilon_1)$. Finally, for large L_T as the frequency spacing becomes finer and finer, using the integral approximation to the summation in (3-27) it follows that

$$\lim_{L_T \to \infty} \overline{J}_{1T}(\theta) = \frac{1}{\pi} \int_0^{\pi} \left| A(e^{j\omega};\theta)H(e^{j\omega}) - B(e^{j\omega};\theta) \right|^2 d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| A(e^{j\omega};\theta)H(e^{j\omega}) - B(e^{j\omega};\theta) \right|^2 d\omega = J_{1\infty}(\theta). \quad (3-29)$$

The desired result then follows from (3-28) and (3-29). □

4 Convergence Analysis

Define

$$\widehat{\theta}_T^{(1)} = \arg\{\min_\theta J_{1T}(\theta)\} \quad (4-1)$$

and

$$\overline{\theta}^{(1)} = \arg\{\min_\theta J_{1\infty}(\theta)\}. \quad (4-2)$$

Using Lemma 1 and some standard arguments we can establish Theorem 1.

Theorem 1. Under (AS1)-(AS4),

$$\lim_{T\to\infty} \widehat{\theta}_T^{(1)} \overset{\text{w.p.1}}{\in} \mathcal{D}^{(1)}$$

where

$$\mathcal{D}^{(1)} := \left\{ \theta \;\middle|\; J_{1\infty}(\theta) = J_{1\infty}(\overline{\theta}^{(1)}) \right\}. \quad (4-3)$$

Proof: Mimic the proof of Theorem 1 in [11] using Lemma 1. Note that the convergence to the set $\mathcal{D}^{(1)}$ is to be interpreted in the sense of Ljung [5, p. 215]. □

The properties of $\overline{\theta}^{(1)}$ have been studied in [9]. First we need some definitions.

Def. **Sufficient Order Model Set:** The true model $H(q^{-1})$ is of the type (1-3) such that the true model orders n_{a0} and n_{b0} satisfy $\min(n_a - n_{a0}, n_b - n_{b0}) \geq 0$. •

Def. **Reduced Order Model Set (Undermodeling):** Either the true model $H(q^{-1})$ is not of the type (1-3), or it is but the true model orders n_{a0} and n_{b0} satisfy $\min(n_a - n_{a0}, n_b - n_{b0}) < 0$. •

It has been shown in [9] that under the **sufficient order** case, $\mathcal{D}^{(1)}$ equals the set

$$\mathcal{D}^{(so)} := \left\{ \theta \;\middle|\; B(q^{-1};\theta)/A(q^{-1};\theta) = H(q^{-1}) \right\}. \quad (4-4)$$

Under **undermodeling** (reduced order case), by [9], $A(q^{-1};\overline{\theta}^{(1)})$ is minimum-phase; hence the fitted model $\widehat{G}(q^{-1}) = B(q^{-1};\overline{\theta}^{(1)})/A(q^{-1};\overline{\theta}^{(1)})$ is stable. Moreover, under undermodeling, $\overline{\theta}^{(1)}$ is unique (i.e. $\mathcal{D}^{(1)}$ is a singleton), and $J_{1\infty}(\overline{\theta}^{(1)}) > 0$.

Using the above results from [9] and Theorem 1, the following result is immediate.

Theorem 2. Under (AS1)-(AS4) and undermodeling,

$$\lim_{T\to\infty} \widehat{\theta}_T^{(1)} \overset{\text{w.p.1}}{=} \overline{\theta}^{(1)}$$

where $\overline{\theta}^{(1)}$ is unique and is given by (4-2). Under (AS1)-(AS4) and sufficient order modeling,

$$\lim_{T\to\infty} \widehat{\theta}_T^{(1)} \overset{\text{w.p.1}}{\in} \mathcal{D}^{(so)}.$$

If $\min(n_a - n_{a0}, n_b - n_{b0}) = 0$, then $\mathcal{D}^{(so)}$ is a singleton. •

5 Conclusions

A frequency-domain solution to the least-squares equation error system identification problem was considered using the power spectrum and the cross-spectrum of the IO data to estimate the IO parametric transfer function. The proposed approach was shown to yield a unimodal performance surface, consistent identification in colored noise and sufficient-order case, and stable fitted models under undermodeling for arbitrary stationary inputs so long as they are persistently exciting of sufficiently high order. This is unlike quite a few existing approaches, such as the prediction error method, the output error method and and the instrumental variable method.

6 References

[1] P.A. Regalia, "An unbiased equation error identifier and reduced-order approximations," *IEEE Trans. Signal Processing*, vol. SP-42, pp. 1397-1412, June 1994.

[2] P.A. Regalia and P. Stoica, "Stability of multivariable least-squares models," *IEEE Signal Processing Letters*, vol. 2, pp. 195-196, Oct. 1995.

[3] J.K. Tugnait and Y. Ye, "Stochastic system identification with noisy input-output measurements using polyspectra," *IEEE Trans. Automatic Control*, vol. AC-40, pp. 670-683, April 1995.

[4] T. Söderström and P. Stoica, *System Identification.* Prentice Hall Intern.: London, 1989.

[5] L. Ljung, *System Identification: Theory for the User.* Prentice-Hall: Englewood Cliffs, N.J., 1987.

[6] T. Soderstrom and P. Stoica, "On the stability of dynamic models obtained by least-squares identification," *IEEE Trans. Automatic Control*, vol. AC-26, pp. 575-577, April 1981.

[7] D.R. Brillinger, *Time Series Data Analysis and Theory.* Holt, Rhinehart & Winston,: New York, 1975.

[8] P.A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control.* Marcel Dekkar: New York, 1995.

[9] C.T. Mullis and R.A. Roberts, "The use of second-order information in the approximation of discrete-time linear systems," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-24, pp. 226-238, June 1976.

[10] B.L.S. Prakasa Rao, *Asymptotic Theory of Statistical Inference.* New York: Wiley, 1987.

[11] J.K. Tugnait, "Stochastic system identification with noisy input using cumulant statistics," *IEEE Transactions on Automatic Control*, vol. AC-37, pp. 476-485, April 1992.

A Novel Approach to Rank Determination of Multichannel Data Covariance Matrices *

Petar M. Djurić
Department of Electrical Engineering
State University of New York at Stony Brook
Stony Brook, New York 11794-2350
djuric@sbee.sunysb.edu

Abstract

The rank selection problem of a multichannel data covariance matrix is addressed by the Bayesian methodology. A maximum a posteriori solution is derived, and a bootstrap technique for its implementation proposed. Our rule is tested on simulated sensor array data that represent random signals embedded in white Gaussian noise. The tests include comparisons with the popular AIC and MDL criteria. The results show that the Bayesian rule outperforms them, particularly for low signal-to-noise ratios and small direction-of-arrival separations.

1. Introduction

In many signal processing applications, the principle of rank reduction plays an important role [4]. In sensor array processing it is often applied to determine the number of signals that arrive at an array using a finite set of observed data vectors. The rank reduction is practically a model selection problem, and as such, in recent years, has been addressed by exploiting information theoretic criteria [1], [5]–[7].

In this paper, we examine the same problem and propose a maximum a posteriori (MAP) solution that is in form similar to the well known selection rules of Akaike (AIC) and Rissanen (MDL) [5]. Our rule has a different penalty for overparameterization, and unlike the AIC and MDL, the penalty is determined from the observed data. It contains terms that include covariance matrices of the estimated model parameters. To estimate these matrices, we apply a bootstrap technique as proposed in [3]. Our rank selection procedure has been tested by computer simulations and compared to the AIC and MDL.

*This work was supported by the National Science Foundation under Award No. MIP-9506743.

2. Problem Statement

We formulate the problem using standard assumptions and the notation from [5]. Namely, a set of $p \times 1$ complex data vectors $\mathbf{x}(t)$, $t = 1, 2, ..., N$, are observed, where

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t). \quad (1)$$

Here, $\mathbf{A}$ is a $p \times m$ ($p > m$) complex matrix of full rank whose columns are associated with different signals and are parameterized by unknown signal parameters, and $\mathbf{s}(t)$ is an $m \times 1$ random complex zero mean vector whose elements are the waveforms of the m signals. The term $\mathbf{n}(t)$ denotes a $p \times 1$ complex noise vector, which is a realization of a stationary and ergodic Gaussian process with zero mean and covariance matrix $E(\mathbf{n}(t)\mathbf{n}^H(t)) = \sigma^2\mathbf{I}$. The noise and the signals are independent.

The covariance matrix of the data can be expressed as

$$\mathbf{R} = \mathbf{\Psi} + \sigma^2\mathbf{I} \quad (2)$$

where $\mathbf{\Psi}$ is the signal covariance matrix given by

$$\mathbf{\Psi} = \mathbf{A}\mathbf{S}\mathbf{A}^H \quad (3)$$

with $\mathbf{S}$ being defined by $\mathbf{S} = E(\mathbf{s}(t)\mathbf{s}^H(t))$. Since $\mathbf{A}$ is a full column rank matrix and $\mathbf{S}$ is by assumption nonsingular, the rank of $\mathbf{\Psi}$ is equal to m. On the other hand, the rank of $\mathbf{R}$ is p, and its $(p-m)$ smallest eigenvalues are equal to σ^2. Thus, if we knew $\mathbf{R}$, by observing its eigenvalues we could directly find the number of signals in the data. However, $\mathbf{R}$ is almost never available in practical applications, but instead, is estimated from the observed data vectors. To determine the number of signals based on the eigenvalues of the estimated $\mathbf{R}$,

0-8186-7576-4/96 $5.00

$\hat{\mathbf{R}}$, is not an easy task because the smallest eigenvalues of $\hat{\mathbf{R}}$ are usually not easily distinguished from the remaining eigenvalues. Our objective is to examine this problem and determine the rank of $\boldsymbol{\Psi}$ from the estimated matrix $\hat{\mathbf{R}}$.

3. Criterion for Rank Reduction

As mentioned before, the rank reduction has been addressed by several authors who used the information theoretic criteria AIC and MDL [5]–[7]. Here, we propose a different approach, which is based on the MAP criterion. We assume that the rank k can take one of the q values, $k = 1, 2, .., q$, where $q < p$. For each k we have a parametric model, which is denoted by $\mathcal{M}_k$. The model $\mathcal{M}_k$ is described by the k largest eigenvalues of $\hat{\mathbf{R}}$, $\hat{\lambda}_l$, $l = 1, 2, ..., k$, their associated eigenvectors, $\hat{\mathbf{v}}_l$, and the noise variance $\hat{\sigma}_k^2$.

Since our objective is to find the rank that has the maximum a posteriori probability, our criterion can be expressed as

$$\hat{k} = \arg\max_k \{p(\mathcal{M}_k|\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(N))\} \quad (4)$$

where $p(\mathcal{M}_k|\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(N))$ is the a posteriori probability of the model given the data records $\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(N)$. If all the rank hypotheses are a priori equally probable, the criterion (4) becomes,

$$\hat{k} = \arg\max_k \{f(\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(N)|\mathcal{M}_k)\} \quad (5)$$

where $f(\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(N)|\mathcal{M}_k)$ is the marginal density of the data given the model $\mathcal{M}_k$. The marginal density is obtained from

$$f(\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(N)|\mathcal{M}_k)$$
$$= \int_{\boldsymbol{\Theta}_k} f(\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(N)|\boldsymbol{\theta}_k, \mathcal{M}_k) f(\boldsymbol{\theta}_k|\mathcal{M}_k) d\boldsymbol{\theta}_k \quad (6)$$

where $\boldsymbol{\Theta}_k$ is the parameter space of the k-th model, and $f(\boldsymbol{\theta}_k|\mathcal{M}_k)$ is the a priori density of the model parameters.

We can show that the criterion (5) can be approximately expressed by [2]

$$\hat{k} = \arg\min_k \{-\ln f(\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(N)|\hat{\boldsymbol{\theta}}_k, \mathcal{M}_k)$$
$$- \frac{1}{2}\ln|\hat{\mathbf{C}}_k| - \frac{d_k}{2}\ln(2\pi)\} \quad (7)$$

where $\hat{\boldsymbol{\theta}}_k$ is the maximum likelihood estimate of $\boldsymbol{\theta}_k$, $\hat{\mathbf{C}}_k$ is the estimated covariance matrix of $\hat{\boldsymbol{\theta}}_k$, and d_k is the dimension of the model's parameter space. If $\mathbf{v}_{1r}$, $\mathbf{v}_{1i}$, $\mathbf{v}_{2r}$, $\mathbf{v}_{2i}$, $\cdots$, $\mathbf{v}_{ki}$ denote the real and imaginary components of the eigenvectors, then $\boldsymbol{\theta}_k^T = [\lambda_1\ \lambda_2\ ...\ \lambda_k\ \hat{\sigma}_k^2\ \mathbf{v}_{1r}^T\ \mathbf{v}_{1i}^{T(-1)}\ \mathbf{v}_{2r}^T\ \mathbf{v}_{2i}^{T(-1)} \cdots \mathbf{v}_{ki}^{T(-1)}]$, where the vectors $\mathbf{v}_{li}^{(-1)}$ are of size $p-1$ with elements identical to the first $p-1$ elements of $\mathbf{v}_{li}$.

Now, if we apply the selection rule (7), we may experience a scaling problem. Namely, for two sets of data which are only related by a scaling constant, the rule may yield two different results. This is unacceptable, and therefore we modify (7) so that the rule is based on the predictive densities $f(\mathbf{x}(L+1), \mathbf{x}(L+2), \cdots, \mathbf{x}(N)|\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(L), \mathcal{M}_k)$, for $k = 1, 2, \cdots, q$, where the data records $\mathbf{x}(1)$, $\mathbf{x}(2)$, $\cdots$, $\mathbf{x}(L)$ can be considered as training data records. With the same approximations used for obtaining (7), we can show that the modified rule becomes

$$\hat{k} = \arg\min_k \{-\ln f(\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(N)|\hat{\boldsymbol{\theta}}_k, \mathcal{M}_k)$$
$$+ \ln f(\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(L)|\tilde{\boldsymbol{\theta}}_k, \mathcal{M}_k) - \frac{1}{2}\ln\frac{|\hat{\mathbf{C}}_k|}{|\tilde{\mathbf{C}}_k|}\} \quad (8)$$

where $\hat{\boldsymbol{\theta}}_k$ and $\tilde{\boldsymbol{\theta}}_k$ are the maximum likelihood estimates of the model parameters obtained from all the data and the first L data records, respectively, and $\hat{\mathbf{C}}_k$ and $\tilde{\mathbf{C}}_k$ are the estimated covariance matrices of $\hat{\boldsymbol{\theta}}_k$ and $\tilde{\boldsymbol{\theta}}_k$, respectively.

Now, using the model assumptions, our rule simplifies to

$$\hat{k} = \arg\min_k \left\{ -\ln\frac{\prod_{l=k+1}^p \hat{\lambda}_l^N}{\hat{\sigma}_k^{2(p-k)N}} + \ln\frac{\prod_{l=k+1}^p \tilde{\lambda}_l^L}{\tilde{\sigma}_k^{2(p-k)L}} \right.$$
$$\left. -\frac{1}{2}\ln|\hat{\mathbf{C}}_k| + \frac{1}{2}\ln|\tilde{\mathbf{C}}_k| \right\} \quad (9)$$

where $\hat{\sigma}_k^2 = \frac{1}{p-k}\sum_{l=k+1}^p \hat{\lambda}_l$, and $\tilde{\sigma}_k^2 = \frac{1}{p-k}\sum_{l=k+1}^p \tilde{\lambda}_l$.

A critical step of the procedure that implements (8) is the evaluation of the covariance matrices $\hat{\mathbf{C}}_k$ and $\tilde{\mathbf{C}}_k$. These matrices can be obtained by a bootstrap technique, which is described in the next section along with some other details of the procedure.

4. Implementation by the Bootstrap Technique

To evaluate (8) for every k, first we find the correlation matrix $\hat{\mathbf{R}}$ according to

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=1}^N \mathbf{x}(t)\mathbf{x}(t)^H = \frac{1}{N}\mathbf{X}\mathbf{X}^H \quad (10)$$

where $\mathbf{X} = [\mathbf{x}(1)\mathbf{x}(2)\cdots\mathbf{x}(N)]$. Similarly, we obtain $\tilde{\mathbf{R}}$ from the first L data vectors. Next, we determine the eigenvalues and eigenvectors of $\hat{\mathbf{R}}$ and order them such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p$. From the $p-k$ smallest eigenvalues, we obtain $\hat{\sigma}_k^2$, and from the so estimated $\hat{\sigma}_k^2$ and the eigenvalues, we determine the first term in (9). We repeat these steps for the matrix $\tilde{\mathbf{R}}$, and find the second term in (9).

To estimate $\hat{\mathbf{C}}_k$ and $\tilde{\mathbf{C}}_k$, we use a bootstrap approach [3]. First we estimate the parameters of the k-th model for M different sets of bootstrap data $\mathbf{X}_l^\star$, $l = 1, 2, \cdots, M$, where $\mathbf{X}_l^\star$ is a $p \times N$ matrix whose columns are randomly chosen from the columns of the actual data matrix $\mathbf{X}$, i.e.,

$$\begin{aligned}\mathbf{X}_l^\star &= [\mathbf{x}_l^\star(1)\mathbf{x}_l^\star(2)\cdots\mathbf{x}_l^\star(N)] \\ &= [\mathbf{x}(l_1)\mathbf{x}(l_2)\cdots\mathbf{x}(l_N)]. \end{aligned} \quad (11)$$

It should be noted that some columns from the original matrix may appear more than ones in $\mathbf{X}_l^\star$, and some not at all. From each of the M bootstrap matrices, we first estimate the model parameters, and then determine the covariance matrix of the parameters. The same procedure is repeated for the $\tilde{\mathbf{X}}_l^\star$ data records to estimate $\tilde{\mathbf{C}}_k$. Once $\hat{\mathbf{C}}_k$ and $\tilde{\mathbf{C}}_k$ are found, we compute the overall criterion of the examined model.

Recall that the $k-$th model parameters are the largest k eigenvalues, the associated eigenvectors, and the noise variance. The eigenvectors have to be treated carefully for two reasons: the first is that they are normalized, and the second, that they are not unique. Since the eigenvectors satisfy

$$\mathbf{v}_l^H \mathbf{v}_l = 1, \quad l = 1, 2 \cdots k \quad (12)$$

not all the elements of $\mathbf{v}_l$ are free parameters. If $\mathbf{v}_l$ is of length p, due to (12), the number of free parameters is $2p-1$. Therefore, in defining $\boldsymbol{\theta}_k$, we have to exclude the non-free parameters. In our definition of $\boldsymbol{\theta}_k$ we exclude the last component of the imaginary part of each eigenvector. Therefore, the sizes of the $\hat{\mathbf{C}}_k$ and $\tilde{\mathbf{C}}_k$ matrices are $(2pk+1) \times (2pk+1)$.

Note also that if $\mathbf{v}_l$ is the eigenvector corresponding to the $l-$th eigenvalue of $\mathbf{R}$, i.e.

$$\mathbf{R}\mathbf{v}_l = \lambda_l \mathbf{v}_l$$

then any vector $\mathbf{v}_l(\varphi) = e^{j\varphi}\mathbf{v}_l$ is also a legitimate eigenvector of $\mathbf{R}$. Since we use the eigenvectors to compute the covariance matrix of the model parameters, this ambiguity is undesirable. So, in our implementation of the bootstrap algorithm we proceed as follows. If the maximum likelihood estimate of the $l-$th eigenvector obtained form $\mathbf{X}$ is $\hat{v}_l$, and the estimate from the bootstrap data $\mathbf{X}^\star$ is $\hat{\mathbf{v}}_l^\star$, we rotate the vector $\hat{\mathbf{v}}_l^\star$ by an angle φ so that we minimize

$$d = (\hat{\mathbf{v}}_l - e^{j\varphi}\hat{\mathbf{v}}_l^\star)^H(\hat{\mathbf{v}}_l - e^{j\varphi}\hat{\mathbf{v}}_l^\star).$$

The angle φ that minimizes d is

$$\varphi = \arctan \frac{\mathrm{Im}(\hat{\mathbf{v}}_l^H \hat{\mathbf{v}}_l^\star)}{\mathrm{Re}(\hat{\mathbf{v}}_l^H \hat{\mathbf{v}}_l^\star)}.$$

Finally, after the rotation, we have to choose the sign of $e^{j\varphi}\hat{\mathbf{v}}_l^\star$ so that the rotated vector points in the same direction as $\hat{\mathbf{v}}_l$. The same steps are implemented in evaluating every eigenvector. So is the case in determining the eigenvectors of $\tilde{\mathbf{C}}_k$.

5. Simulation Results

We have tested the MAP rule in three experiments and compared it with the AIC and MDL by using computer simulated data. The columns of the matrix $\mathbf{A}$ were given by

$$\mathbf{a}_k^T(\phi) = [1 \; e^{-j\pi\sin(\phi_k)} \; e^{-j2\pi\sin(\phi_k)} \; \cdots \; e^{-j(p-1)\pi\sin(\phi_k)}] \quad (13)$$

where $k = 1, 2, \cdots, m$, and ϕ_k is the direction of arrival of the k-th signal.

In all the experiments, there were two signals ($m = 2$) whose amplitudes were given by

$$\mathbf{s}^T(t) = [e^{-j\eta_1(t)} \; e^{-j\eta_2(t)}] \quad (14)$$

where η_1 and η_2 are independent and uniformly distributed random variables in the interval $(0, 2\pi)$. The number of sensors was $p = 7$, and the maximum possible rank $q = 4$. The evaluation of the covariance matrices was carried out by $M = 300$ bootstrap data matrices, and the number of training data records was L=10. In each experiment there were 100 independent trials. The used AIC and MDL rules were

$$\hat{k}_{\mathrm{AIC}} = \arg\min_k \left\{ -2\ln \frac{\prod_{l=k+1}^{p} \hat{\lambda}_l^N}{\hat{\sigma}^{2(p-k)N}} + 2k(2p-k) \right\} \quad (15)$$

and

$$\hat{k}_{\mathrm{MDL}} = \arg\min_k \left\{ -\ln \frac{\prod_{l=k+1}^{p} \hat{\lambda}_l^N}{\hat{\sigma}^{2(p-k)N}} + \frac{k(2p-k)\ln N}{2} \right\}. \quad (16)$$

In the first experiment the signal-to-noise ratio (SNR), defined by $\mathrm{SNR} = 10\log(1/\sigma^2)$, was equal to 0 dB, and the directions of arrival were $\phi_1 = 20°$ and $\phi_2 = 28°$. The number of data records was $N = 80$. The results are shown in Table 1. Each entry represents the number of times the MAP, AIC, and MDL

	$k=1$	$k=2$	$k=3$	$k=4$
MAP	0	99	1	0
AIC	0	95	4	1
MDL	0	100	0	0

Table 1. Performance of the MAP, AIC, and MDL rules in 100 trials for SNR=0 dB, ϕ_1=20°, ϕ_2 = 28°, and N = 80. The correct model is $\mathcal{M}_2$.

	$k=1$	$k=2$	$k=3$	$k=4$
MAP	2	89	8	1
AIC	13	74	12	1
MDL	77	23	0	0

Table 2. Performance of the MAP, AIC, and MDL rules in 100 trials for SNR=-3 dB, ϕ_1= 20°, ϕ_2 = 28°, and N = 50. The correct model is $\mathcal{M}_2$.

rules chose the models with ranks $k = 1, 2, 3$, and 4, respectively out of 100 trials. From the results we observe that all the rules showed excellent performance.

In the second experiment, we decreased the SNR to - 3 dB and the number of data records to $N = 50$, but kept all the remaining parameters identical to those in experiment 1. The results are shown in Table 2. The performance of the MDL degraded significantly. The AIC performed better, and the MAP was the best.

Finally, in the third experiment, we decreased the separation of directions of arrival by setting $\phi_1 = 22°$, and $\phi_2 = 28°$, increased the SNR to 0 dB, and kept the remaining parameters unchanged as in experiment 2. The results are shown in Table 3. Again, the MDL performed poorly, and the MAP had the best performance.

6. Conclusions

A new approach to rank determination of covariance matrices has been proposed. It is based on the MAP criterion and implemented by the bootstrap method. The method in this paper requires assumptions of a specific structure of the covariance matrix. Current research is focused on relaxing these assumptions to make the current procedure applicable in a wider set of scenarios.

	$k=1$	$k=2$	$k=3$	$k=4$
MAP	2	92	5	1
AIC	7	89	3	1
MDL	47	53	0	0

Table 3. Performance of the MAP, AIC, and MDL rules in 100 trials for SNR=0 dB, ϕ_1= 22°, ϕ_2= 28° and N = 50. The correct model is $\mathcal{M}_2$.

References

[1] Z. D. Bai, B. Q. Miao, and C. R. Rao, "Estimation of direction of arrival of signals: asymptotic results," in *Advances in Spectrum Analysis and Array Processing*, S. Haykin, ed., pp. 327-347, 1991.

[2] P. M. Djurić, "Model selection based on asymptotic Bayes theory," Proceedings of the Seventh SP Workshop on Statistical Signal & Array Processing, pp. 7-10, Quebec City, Canada, 1994.

[3] B. Efron and R. J. Tibshurani, *An Introduction to the Bootstrap*, New York: Chapman and Hall, 1993.

[4] L. L. Scharf and D. W. Tufts, "Rank reduction for modeling stationary signals," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-35, pp. 350-355, 1987.

[5] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-33, pp. 387-392, 1985.

[6] D. B. Williams, "Counting the degrees of freedom when using AIC and MDL to detect signals," IEEE Transactions on Signal Processing, vol. 42, pp. 3282-3283, 1994.

[7] K. M. Wong, Q.-T. Zhang, J. P. Reilly, and P. C. Yip, "On information theoretic criteria for determining the number of signals in high resolution array processing," vol. 38, pp. 1959-1971, 1990.

On Rate of Convergence of Some Consistent Estimates of the Number of Signals

K. W. Tam
Portland State University
Portland, Oregon 97207 USA
hmkt@psuorvm.cc.pdx.edu

Y. Wu
York University
North York, Ontario Canada M3J1P3
wuyh@mathstat.yorku.ca

Abstract

Using the information theoretic criterion the authors obtained in [4] three consistent estimates of the number of signals for an additive model with white noise. In this paper the rates of convergence for the probabilities of wrong detections as a function of the sample size are studied. It is proved that under certain conditions and for a fairly general class of penalty terms, the probabilities of wrong detection are exponentially decreasing.

1 Introduction

In signal processing, a problem of great interest is the determination of the number of signals transmitted in the presence of noise. The received signal vector $\boldsymbol{x}(t)$ is $p \times 1$ complex vector given by

$$\boldsymbol{x}(t) = A\boldsymbol{s}(t) + \boldsymbol{n}(t)$$

where A is $p \times q$ matrix $A = [A(\phi_1), A(\phi_2), \ldots, A(\phi_q)]$ and $A(\phi_i)$ is a $p \times 1$ complex vector which depends on some unknown ϕ_i associated with the direction of arrival for the ith signal, $\boldsymbol{s}(t) = (s_1(t), s_2(t), \ldots, s_q(t))'$, $s_i(t)$ is the ith complex waveform signal, and $\boldsymbol{n}(t)$ is a $p \times 1$ complex vector associated with the noise. The assumptions made here are (1) $q < p$; (2) $\boldsymbol{s}(t)$ is complex multivariate normal with mean vector $\mathbf{0}$ and nonsingular covariance matrix Ψ; (3) the noise vector $\boldsymbol{n}(t)$ is complex multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2 I_p$, where I_p is the $p \times p$ identity matrix, and $\boldsymbol{n}(t)$ is also independent of the signals. The covariance matrix of the $\boldsymbol{x}(t)$ is given by

$$\Sigma = A\Psi A^* + \sigma^2 I_p$$

where A^* denotes the transpose of the complex conjugate of$A$. The number of signals transmitted, q, is equal to the rank of $A\Psi A^*$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ be the eigenvalues of Σ,

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q > \lambda_{q+1} = \cdots = \lambda_p = \sigma^2.$$

Let $\{\boldsymbol{x}(t_1), \boldsymbol{x}(t_2), \ldots, \boldsymbol{x}(t_n)\}$ be a set of observations and $nS = \sum_{i=1}^{n} \boldsymbol{x}(t_i)\boldsymbol{x}^*(t_i)$. Then $E(S) = \Sigma$. Suppose that $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_p$ are the eigenvalues of S. Let H_k denote the hypothesis that

$$H_k : \lambda_1 \geq \cdots \geq \lambda_k > \lambda_{k+1} = \cdots = \lambda_p = \sigma^2.$$

M_k is the model that H_k is true. Let

$$I(k, C_n) = L(k) + \nu(k,p)C_n. \tag{1}$$

Here $L(k)$ is a statistic which will be specified later, $\nu(k,p)$ denotes the number of free parameters that has to be estimated under H_k and C_n is some constant chosen to depend on n. The criterion for determing the number of signals is to estimate the number of signals q by $\hat{q}_n$ which is chosen so that

$$I(\hat{q}_n, C_n) = \min\{I(0, C_n), I(1, C_n), \cdots, I(p-1, C_n)\}.$$

$\nu(k,p)C_n$ in this case is called the penalty term. The choice of C_n is crucial and its selection was discussed in [5]. In general C_n is chosen to satisfy the following conditions:

$$\lim_{n\to\infty} \frac{C_n}{n} = 0; \qquad \lim_{n\to\infty} \frac{C_n}{\log\log n} = \infty, \tag{2}$$

and $\nu(k,p)$ is either $\frac{k}{2}$ or $\frac{k}{2}(2p-k+1)$. Let

$$L_1(k) = -n\{\sum_{i=k+1}^{p} \log\delta_i - (p-k)\log(\frac{1}{p-k}\sum_{i=k+1}^{p}\delta_i)\}.$$

$$\Delta_k = -\{\sum_{i=k+1}^{p} \log\lambda_i - (p-k)\log(\frac{1}{p-k}\sum_{i=k+1}^{p}\lambda_i)\}.$$

It can be shown that $-L_1(k)$ is the likelihood ratio test statistic for testing H_k under the assumptions of normality and independence of observations.

0-8186-7576-4/96 $5.00

With $L(k) = L_1(k)$ in (1) it was proved in [5] that the estimate of q by $\hat{q}$ is consistent. Its rate of convergence was shown to be exponential in [1]. The same result was obtained in [2] where white noise is not assumed. Recently a different formulation of $L_1(k)$ was provided in [3]. However it is not known if it will give a consistent estimate of q. Based on the result of [3], the authors proposed several consistent estimates in [4]. The main purpose of this paper is to investigate the rate of convergence for three of these consistent estimates. It will be proved that for a fairly general class of C_n functions, the rate is again exponential.

2 Preliminary Lemmas

Lemma 2.1: Suppose that $0 < \alpha \leq \sigma^2/(2p)$ and $\max_{i,j} |\sigma_{ij} - s_{ij}| \leq \alpha$. The followings are true:

1. $0 \leq L_1(k) \leq np^3\alpha^2/\sigma^4$, if $k \geq q$;
2. $L_1(k) + \nu(k,p)C_n - L_1(q) - \nu(q,p)C_n \geq n(\Delta_k/2 - 4p^2\alpha/\sigma^2 - p^3\alpha^2/\sigma^4) - \nu(p-1,p)C_n$, if $k < q$

Proof: For the proof, see Theorem 3.1 of [1].

Let $\Sigma = \{\sigma_{ij}\}_{i,j=1}^p$ and $S = \{s_{ij}\}_{i,j=1}^p$. The following result can be found in [1]:

$$\max_{i,j} |\sigma_{ij} - s_{ij}| \leq \alpha \Rightarrow |\lambda_i - \delta_i| \leq p\alpha, \ i = 1, 2, \ldots, p \tag{3}$$

For $\xi > 0$, define

$$\begin{aligned}\bar{\Lambda}_{ac}(\xi) = n(&\sum_{i=1}^{k} \frac{\delta_i}{\bar{\lambda}_i} + \sum_{i=k+1}^{p} \frac{\delta_i}{\bar{\lambda}_{k+1}}) + (n-p+1)\sum_{i=1}^{k} \log \bar{\lambda}_i \\ &+ (p-k)(n-k)\log \bar{\lambda}_{k+1} + \sum_{\substack{i,j=1 \\ i<j}}^{k} \log(\bar{\lambda}_i - \bar{\lambda}_j + \xi) \\ &+ \sum_{i=1}^{k}(p-k)\log(\bar{\lambda}_i - \bar{\lambda}_{k+1} + \xi) \\ &- \sum_{\substack{i,j=k+1 \\ i<j}}^{p} \log(\delta_i - \delta_j + \xi),\end{aligned}$$

where $\bar{\lambda}_i$ and $\bar{\lambda}_{k+1}$ are solutions of

$$\begin{aligned}\bar{\lambda}_j &= \delta_j - \frac{\bar{\lambda}_j}{n} \sum_{\substack{i=1 \\ i \neq j}}^{k} \frac{\bar{\lambda}_i}{\bar{\lambda}_j - \bar{\lambda}_i + \xi} \\ &\quad - \frac{p-k}{n} \frac{\bar{\lambda}_j \bar{\lambda}_{k+1}}{\bar{\lambda}_j - \bar{\lambda}_{k+1} + \xi}, \quad j = 1, 2, \ldots, k \\ \bar{\lambda}_{k+1} &= \hat{\sigma}^2 + \frac{1}{n}\sum_{i=1}^{k} \frac{\bar{\lambda}_i \bar{\lambda}_{k+1}}{\bar{\lambda}_i - \bar{\lambda}_{k+1} + \xi}. \end{aligned} \tag{4}$$

$$|\bar{\lambda}_i - \bar{\lambda}_j| \leq \rho\xi, \ |\bar{\lambda}_i| \leq n^{\frac{2}{5}} \log n, \quad i,j = 1, 2, \ldots, k+1$$

where $0 < \rho < 1$ is a constant, $\hat{\sigma}^2 = \frac{1}{p-k}\sum_{i=k+1}^{p} \delta_i$.

Lemma 2.2: There is a $N > 0$ such that for all $n \geq N$, the system of equations (4) has at most one solution.

Proof: Define

$$\begin{aligned}g_j(x_1, x_2, \cdots, x_{k+1}) &= \delta_j - \frac{x_j}{n} \sum_{\substack{i=1 \\ i \neq j}}^{k} \frac{x_i}{x_j - x_i + \xi} \\ &\quad - \frac{p-k}{n} \frac{x_j x_{k+1}}{x_j - x_{k+1} + \xi}, \text{ for } j = 1, 2, \ldots, k\end{aligned}$$

$$g_{k+1}(x_1, x_2, \cdots, x_{k+1}) = \hat{\sigma}^2 + \frac{1}{n}\sum_{i=1}^{k} \frac{x_i x_{k+1}}{x_i - x_{k+1} + \xi}$$

and $G(\vec{x}) = (g_1, g_2, \ldots, g_{k+1})(\vec{x})$. Then $G(\vec{y}) = \vec{y}$ if and only if $\vec{y}$ is a solution of (4). Let $D_n = \{\vec{x} : |x_i - x_j| \leq \rho\xi, \ |x_j| \leq n^{2/5}\log n, \ i,j = 1, 2, \ldots, k+1\}$. Then D_n is convex. It is easy to show that on D_n,

$$\begin{aligned}M &= \sum_{j=1}^{k+1} || \nabla g_j(\vec{z})|| \\ &\leq (p-1)(k+1)^{3/2} \frac{n^{2/5}\log n(n^{2/5}\log n + \xi)}{n(1-\rho)^2\xi^2}\end{aligned}$$

where $||.||$ denotes the Euclidean norm. Suppose that there exist two distinct fixed points $\vec{x}$ and $\vec{y}$ in D_n. By the mean value theorem there is a $\vec{z} \epsilon D_n$ on the line segment joining $\vec{x}$ and $\vec{y}$ such that

$$\vec{x} - \vec{y} = G(\vec{x}) - G(\vec{y}) = G'(\vec{z})(\vec{x} - \vec{y})$$

Let N be so that for all $n \geq N$, $M < 1$. Then $||\vec{x} - \vec{y}|| \leq M||\vec{x} - \vec{y}|| < ||\vec{x} - \vec{y}||$. This is a contradiction and the uniqueness of solution is proved.

Lemma 2.3: Suppose that $\rho\xi > \lambda_1 - \sigma^2$ and $\max_{i,j} |\sigma_{ij} - s_{ij}| \leq \alpha_0 < \min\{\sigma^2/(2p), (\rho\xi - \lambda_1 + \sigma^2)/(4p)\}$. Then there is a $N > 0$ such that for $n \geq N$ the system (4) has a unique solution and

$$\begin{aligned}|\delta_i - \bar{\lambda}_i| &\leq \frac{\gamma}{n}, \quad i = 1, 2, \ldots, k \\ |\hat{\sigma}^2 - \bar{\lambda}_{k+1}| &\leq \frac{\gamma}{n}\end{aligned} \tag{5}$$

where $\gamma = p(\lambda_1 + 2p\alpha_0)^2/[(1-\rho)\xi]$.

Proof: Consider the subset of R^{k+1}, $E_k = \{|x_i - \delta_i| \leq \alpha_0, \ i = 1, 2, \ldots, k, \ |x_{k+1} - \hat{\sigma}^2| \leq \alpha_0$ and $|x_i - x_j| \leq \rho\xi\}$. Let N be large enough such that $(p-1)[\lambda_1 + (p+1)\alpha_0]^2/(N(1-\rho)\xi) \leq \alpha_0$. If $n \geq N$, then it can showed that

$$\begin{aligned}&|\delta_j - g_j(\vec{x})| \leq \alpha_0, \ j = 1, 2, \ldots, k \\ &|\hat{\sigma}^2 - g_{k+1}(\vec{x})| \leq \alpha_0, \\ &|g_i(\vec{x}) - g_j(\vec{x})| \leq |\delta_i - \delta_j| + \alpha_0 \leq \lambda_1 - \sigma^2 + 4p\alpha_0 \leq \rho\xi \\ &\qquad i,j = 1, 2, \ldots, k+1.\end{aligned}$$

Therefore $G(\vec{x}) \in E_k$. Clearly $G(\vec{x})$ is continous. By the fixed point theorem the system (4) has a solution $(\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_{k+1})$ in E_k and it follows that

$$|\delta_i - \bar{\lambda}_i| \leq \frac{1}{n}\frac{p(\lambda_1 + 2p\alpha_0)^2}{(1-\rho)\xi}, \quad i = 1, 2, \ldots, k$$

$$|\hat{\sigma}^2 - \bar{\lambda}_{k+1}(\vec{x})| \leq \frac{1}{n}\frac{p(\lambda_1 + 2p\alpha_0)^2}{(1-\rho)\xi}.$$

By Lemma 2.2, the solution is unique.

Lemma 2.4: Suppose that $\rho\xi > \lambda_1 - \sigma^2$ and $\max_{i,j} |\sigma_{ij} - s_{ij}| \leq \alpha \leq \alpha_0 < \min\{\sigma^2/(2p), (\rho\xi - \lambda_1 + \sigma^2)/(4p)\}$. There is a $N > 0$ such that for $n > N$, we have

(a) $\frac{\delta_i}{\bar{\lambda}_i} = 1 + R_1(i,k), \quad i = 1, 2, \ldots, k$

(b) $\sum_{i=k+1}^{p} \frac{\delta_i}{\bar{\lambda}_{k+1}} = (p-k)[1 + R_2(k)]$

(c) $\log(\bar{\lambda}_i) = \log \delta_i + R_3(i,k), \quad i = 1, 2, \ldots, k$

(d) $\log(\bar{\lambda}_{k+1}) = \log(\hat{\sigma}^2) + R_4(k)$

(e) $\log(\bar{\lambda}_i - \bar{\lambda}_j + \xi) = \log(\delta_i - \delta_j + \xi) + R_5(i,j,k)$
$i, j = 1, 2, \ldots, k, \ i < j$

(f) $\log(\bar{\lambda}_i - \bar{\lambda}_{k+1} + \xi) = \log(\delta_i - \hat{\sigma}^2 + \xi) + R_6(i,k)$
$i, j = 1, 2, \ldots, k$

where $|R_i| \leq \eta/n \ (1 \leq i \leq 6), \gamma$ as in Lemma 2.3 and $\eta = \max\{\frac{(\lambda_1 + p\alpha_0)\gamma}{(\sigma^2 - (p+1)\alpha_0)^2}, \frac{\gamma}{\sigma^2 - (p+1)\alpha_0}, \frac{2\gamma}{(1-\rho)\xi}\}$

Proof: The proof requires simple calculus.

3 Rates of Convergences

Suppose that the assumptions made in Lemma 2.3 are satisfied. Using the expansion of Lemma 2.4, we may rewrite

$$\bar{\Lambda}_{ac}(\xi) = L_1(k) + n\sum_{i=1}^{p} \log \delta_i + np + \beta(n,k)$$

The following simple bound for $\beta(n,k)$ can be obtained by the estimates of Lemma 2.4 and (3).

$$|\beta(n,k)| \leq 2p^2[\eta + |\log(\lambda_1 + p\alpha_0)| \vee |\log(\sigma^2 - p\alpha_0)| + |\log(\lambda_1 - \lambda_p + 2p\alpha_0 + \xi)| \vee |\log((1-\rho)\xi - 2\gamma/N)|] \stackrel{\text{def}}{=} \tilde{\Delta}_k.$$

For the first type of estimates define $L(k)$ in (1) by

$$L(k) = \bar{\Lambda}_{ac}(\xi) + p_{ac} \tag{6}$$

where $p_{ac} = \frac{1}{2}k(2p-k-1)\log n - \sum_{i=p-k+1}^{p} \log \Gamma(i)$.

With N given in Lemma 2.3 and $n \geq N$, we have the following theorem.

Theorem 3.1: Assume that $\rho\xi > \lambda_1 - \sigma^2$ and assume also that the following conditions are satisfied:

(a) $0 < \alpha \leq \alpha_0 < \min\{\sigma^2/(2p), (\rho\xi - \lambda_1 + \sigma^2)/(4p)\}$;

(b) $\min_k\{\nu(k+1,p) - \nu(k,p))C_n\} > 2np^3\alpha^2/\sigma^4 + 2\max_k\{\tilde{\Delta}_k\} + \sum_{i=1}^{p} \log \Gamma(i)$;

(c) $\min_{k \leq q}\{\Delta_k\} > 4p^2\alpha/\sigma^2 + p^3\alpha^2/\sigma^4 + \nu(p-1,p)C_n/n + p^2 \log n/n + 2\max_k\{\tilde{\Delta}_k\}/n$.

Then

$$P(\hat{q}_n \neq q | H_q) \leq \sum_i \sum_j P(|s_{ij} - \sigma_{ij}| \geq \alpha).$$

Proof: Suppose that $\max_{i,j} |\sigma_{ij} - s_{ij}| \leq \alpha$. By the assumptions, the results of Lemma 2.3 and 2.4 hold. By the definition of $I(k, C_n)$, we have

$$I(k,C_n) - I(q,C_n) = \frac{1}{2}(k-q)[2p - (q+k) - 1]\log n + [L_1(k) + \nu(k,p)C_n - L_1(q) - \nu(q,p)C_n] + \beta(n,k) - \beta(n,q) + \sum_{i=p-q+1}^{p} \log \Gamma(i) - \sum_{i=p-k+1}^{p} \log \Gamma(i).$$

Hence, for $k > q$ it follows from lemma 2.1 and the assumption (b) that

$$I(k,C_n) - I(q,C_n) \geq (\nu(q+1,p) - \nu(q,p))C_n - 2np^3\alpha^2/\sigma^4 - 2\max_k\{\tilde{\Delta}_k\} - \sum_{i=1}^{p} \log \Gamma(i) > 0. \tag{7}$$

For $k < q$, it follows from lemma 2.1 and assumption (c) that

$$I(k,C_n) - I(q,C_n) \geq n(\Delta_k - 4p^2\alpha/\sigma^2 - p^3\alpha^2/\sigma^4 - \nu(p-1,p))C_n - 2\max_k\{\tilde{\Delta}_k\} - p^2 \log n > 0. \tag{8}$$

In view of (7) and (8), we have $\hat{q} = q$. Therefore,

$$P(\hat{q}_n \neq q | H_q) \leq \sum_i \sum_j P(|s_{ij} - \sigma_{ij}| \geq \alpha).$$

If $\nu(k,p)$ is a strictly increasing function of k and

$$\alpha = \alpha(n) \downarrow 0, \ \frac{C_n}{n} \to 0, \ \frac{C_n}{n\alpha^2} \to \infty, \tag{9}$$

then for $\rho\xi > \lambda_1 - \sigma^2$ we have the following theorem.

Theorem 3.2: If $\nu(k,p)$, $\alpha(n)$ and C_n satisfy (9), then the probability of wrong detection using $\hat{q}_n$ satisfies the following inequality:

$$P(\hat{q}_n \neq q|H_q) \leq \sum_i \sum_j P(|s_{ij} - \sigma_{ij}| \geq \alpha)$$

Proof: It is obvious that the assumptions made in Theorem 3.1 hold. Then the result follows from Theorem 3.1.

For $\xi > 0$ the second type of estimates is defined by letting $L(k)$ in (1) to be

$$L_2(k) = (n-p+1)\sum_{i=1}^{k} \log \delta_i + (p-k)(n-k)\log(\hat{\sigma}^2)$$

$$+ \sum_{\substack{i,j=1 \\ i<j}}^{k} \log(\delta_i - \delta_j + \xi) + \sum_{i=1}^{k} (p-k)\log(\delta_i - \hat{\sigma}^2 + \xi)$$

$$- \sum_{\substack{i,j=k+1 \\ i<j}}^{p} \log(\delta_i - \delta_j + \xi) = L_1(k) + n\sum_{i=1}^{p} \log \delta_i$$

$$+ \beta(n,k) - \sum_{\substack{i,j=k+1 \\ i<j}}^{p} \log(\delta_i - \delta_j + \xi) \qquad (10)$$

$$\beta(n,k) = \sum_{i=1}^{k} (p-k)\log(\delta_i - \hat{\sigma}^2 + \xi) + (1-p)\sum_{i=1}^{k} \log \delta_i$$

$$+ 2 \sum_{\substack{i,j=k+1 \\ i<j}}^{p} \log(\delta_i - \delta_j + \xi) - k(p-k)\log(\hat{\sigma}^2)$$

For the third type of estimates, use

$$I(k, C_n) = L_2(k) + p_{ac} + \nu(k,p)C_n \qquad (11)$$

with $L_2(k)$ as in (10). Let $\hat{q}_n$ be the estimate obtained by either the second and third type, then the following theorem with proof similar to that of Theorem 3.2 is true.

Theorem 3.3: If $\nu(k,p)$, $\alpha(n)$ and C_n satisfy (9), then the probability of wrong detection using $\hat{q}_n$ satisfies the following inequality:

$$P(\hat{q}_n \neq q|H_q) \leq \sum_i \sum_j P(|s_{ij} - \sigma_{ij}| \geq \alpha)$$

In the following, $\hat{q}_n$ will denote an estimate given by any of the three types mentioned.

Theorem 3.4: Suppose that $\mathbf{x}_1, \mathbf{x}_2, \cdots$ are i.i.d. vectors of order $p \times 1$ such that $E(\mathbf{x_1}) = 0$, $E(\mathbf{x}_1\mathbf{x}_1^*) = \Sigma$ and $E|\mathbf{x_1}|^{2\kappa} < \infty$ for some $\kappa > 1$. Also let C_n in (1) be chosen so that C_n, $\alpha(n)$ and $\nu(k,p)$ satisfy (9). Then for any $s > \kappa$, we have

$$P(\hat{q}_n \neq q|H_q) = O(n/(n\alpha)^{\kappa}) + O((n\alpha^2)^{-s})$$

as $n \to \infty$.

Proof: Indentical to that of Theorem 3.2 [1].

Similarly the following results as in [1] are true:

Corolary 3.1: In Theorem 3.4, if we take $\alpha = \alpha(n) \downarrow 0$ as a slowly varying function and $C_n = n\alpha$, then

$$P(\hat{q}_n \neq q|H_q) = O(n^{1-\kappa}(\alpha)^{-\kappa})$$

as $n \to \infty$.

Theorem 3.5: Suppose that $\mathbf{x}_1, \mathbf{x}_2, \cdots$ are i.i.d. with $E(\mathbf{x}_1) = \mathbf{0}$, $E(\mathbf{x}_1\mathbf{x}_1^*) = \Sigma$ and $E\{\exp(\kappa|\mathbf{x}_1|^2\} < \infty$ for some $\kappa > 0$. Then

$$P(\hat{q}_n \neq q|H_q) \leq c\exp(-bn\alpha^2)$$

as $n \to \infty$ for some constant $b > 0$ and $c > 0$.

Corollary 3.2: If $\alpha(n) \downarrow 0$ is a slowly varying function, $C_n = \alpha(n)n$ and the conditions of Theorem 3.5 are satisfied, for any $\epsilon > 0$

$$P(\hat{q}_n \neq q|H_q) \leq c\exp(-bn^{1-\epsilon}).$$

References

[1] Z. D. Bai, P. R. Krishnaiah, P. R. and L. C. Zhao, "On rates of convergence of efficient detection criteria in signal processing with white noise", *IEEE Trans. Inform. Theory, vol. 35*, pp.380-388, 1989

[2] Tam, K. W. and Wu, Yuehua, "On the convergence rate of general information theorectic criteria in signal processing when the covariance matrix is arbitrary", *IEEE Trans. Inform. Theory, vol 37*, pp.1667-1671, 1991

[3] K. M. Wong, Q. T. Zhang, J. P. Reily and P. C. Yip, "On Information Theorectic Criteria for Determining the Number of Signals in High Resolution Array Processing", *IEEE Trans Acoust., Speech and Signal Processing, V38"*, pp.1959-1970, 1990

[4] Y. Wu and K. W. Tam, "On Estimation of Number of Signals", Submitted

[5] L. C. Zhao, P. R. Krishnaiah and Z. D. Bai, "On Determination of Number of Signals in Presence of White Noise", *J. Multivariate Anal. V20*, pp.1-25, 1986

Asymptotic Statitics of AR Spectral Estimators for Processes Containing Mixed Spectrum

Peter Sherman
Iowa State University
Ames, Iowa U.S.A

Lang White
DSTO
Australia

Joanna Spanjaard
CRASys
Australia

Robert Bitmead
Austr. Natl. University
Australia

Abstract

We address the influence of point spectrum on the large sample statistics of the AR(n) spectral estimator for fixed n as well as for the case where n approaches infinity. For fixed n we obtain the distribution of this estimator. We also obtain approximate expressions for its mean and variance. These expressions involve the nth order Capon spectrum. Using recently discovered convergence properties of this spectrum as n approaches infinity, we show that these expressions depend on the ratio of the AR(n) to the nth order Capon spectrum. This ratio gives insight into the statistical influence of point spectrum on the AR(n) spectral estimator, based on the well known difference in the resolving properties of these two spectra. Simulations are included to support the theoretical results. Finally, it is hoped that our attempt to bring to bear a number of recently published results in this area will also contribute to a better understanding of it, and possibly stimulate further investigations.

Introduction

This work addresses the statistics of spectral estimators associated with a zero mean wide sense stationary (wss) random process having mixed spectrum. From the Wold decomposition, any wss random process, Y_t , has a decomposition of the form

$$Y_t = X_t + U_t \qquad (1a)$$

where X_t has an absolutely continuous spectral density, and where U_t is is independent of Y_t and is perfectly predictable given $\{ U_s \ ; \ s < t \}$. In this work we restrict U_t to be a harmonic process; that is,

$$U_t = \sum_k A_k \sin(\omega_k t + \theta_k) \qquad (1b)$$

where the $\{\theta_k\}$ are independent, and identically distributed (iid) uniformly over the interval $[-\pi, \pi]$, and where $\{A_k, \omega_k\}$ are unknown parameters. Consequently, Y_t has an autocorrelation of the form

$$r_y(\tau) = r_x(\tau) + r_u(\tau)$$

$$= r_x(\tau) + \sum_k \frac{A_k^2}{2} \cos(\omega_k \tau) \qquad (2)$$

where $\{r_x(\tau)\}_{-\infty}^{\infty}$ is absolutely summable. The statistics of spectral estimators related to the regular process, X_t, have been studied extensively over the years. Only relatively recently, however, have those of Y_t received much attention. There are a number of reasons for this recent interest. One is no doubt due to the increasing importance of such processes in the engineering and physical sciences. An example is the spectral analysis of signals associated with periodic systems such as rotating machinery in order to identify cyclostationary behavior [1]. Complications introduced by the presence of a harmonic process in these areas are illustrated in [1]. In fact, given the adverse influence of such a process on practically all methodologies related to not only spectral estimation, but also to system identification and feedback control, one might wonder why processes such as (1) have been of such limited interest. At least a partial answer to this question relates to the mathematical difficulties imposed by (1b). Afterall, a key assumption found in all these areas is that the autocorrelation function decay sufficiently fast; whereas (2) does not decay at all.

The goal of this paper is to characterize the large sample statistical properties of a particular class of spectral density estimators, namely autoregressive (AR) estimators. We consider both a fixed order, n, model, as well as when n approaches infinity. Much of this characterization will be obtained by piecing together recent results of other researchers, and in particular, those published in statistical journals. A significant portion will also follow directly from the convergence results of [2] related to the family of Capon spectral estimators (see [1] for more recent related references). Consequently, while we believe that this work contains valuable original contributions, it is also our intent to contribute by combining a collection of recent developments along these lines into a self-contained work. To begin, we define the spectral density and power spectrum for the process (1):

0-8186-7576-4/96 $5.00

$$S_y(\omega) \triangleq \sum_{-\infty}^{\infty} r_y(\tau) e^{-i\omega\tau}$$
$$= S_x(\omega) + \frac{A_k^2}{4}\,\delta(\omega \pm \omega_k)\ , \qquad (3a)$$

$$P_y(\omega) \triangleq \lim_{n\to\infty}\ (2n+1)^{-1} \sum_{-n}^{n} r_y(\tau) e^{-i\omega\tau}$$
$$= \frac{A_k^2}{4}\delta_{kr}(\omega \pm \omega_k)\ . \qquad (3b)$$

We remark in the mixed spectrum setting it follows from (3) that the spectral density is not well-defined, in the sense that it becomes unbounded as the number of available autocorrelation lags approaches infinity. Moreover, the power spectrum contains no information about the spectral density. Hence, it is natural to expect that the AR(n) spectral estimator will be poorly behaved near the point spectrum frequencies, and furthermore, that in the limit (as $n\to\infty$) it will not exist at these frequencies. To arrive rigorously at these conjectures, consider the AR(n) prediction model

$$\hat{y}_t = -\sum_{k=1}^{n} a_k y_{t-k} \;=\; Y_t^{-} a \qquad (4)$$

where $a \triangleq [a_1,...,a_n]^{tr}$, and define the prediction error $\sigma^2 \triangleq E(y_t - \hat{y}_t)^2$. For clarity of understanding and notation, we present the minimum variance and least squares equations.

Minimum Variance Approximation of a and σ^2

$$a_{mv} = -R^{-1} r \quad ; \quad \sigma_{mv}^2 = r_y(0) + a^t r \qquad (5)$$

where $\quad R \triangleq \{r_y(i-j) = r_{i-j}\}_{i,j=1}^{n}$

and $r \triangleq [r_y(1),...,r_y(n)]^{tr}$. In the statistical sense used throughout this paper a_{mv} is not an estimator, since it is not random. For this reason, a_{mv} is termed the minimum variance approximant of a.

Least Squares Estimation of a and σ^2

$$\hat{a} = -\hat{R}^{-1}\hat{r} \quad ; \quad \hat{\sigma}^2 = \hat{r}_y(0) + \hat{a}^t\hat{r} \qquad (6)$$

$$\hat{r}_j = (1/N)\sum_{1}^{N-j} y_t y_{t+j}\ . \qquad (7)$$

Main Results

We now summarize the key results in this work.

THEOREM 1. (Li et al [3]) . Let $\hat{r} \triangleq [\hat{r}_0,...,\hat{r}_n]^{tr}$ whose elements are given by (7). Then

$$N^{-1/2}\ (\hat{r} - r)\ \underset{N\to\infty}{\overset{d}{\to}}\ \mathcal{N}(0,\Sigma)\ .$$

The elements of Σ are complex expressions, and so are omitted here for brevity. This recent (1994) result is an extension of the 1990 result of [4] for white X_t to the colored case. It leads immediately to our first result.

THEOREM 2. If for any sufficiently large N we have $\hat{R} > c > 0$ almost surely, then

$$\hat{a} \;\dot{\sim}\; \mathcal{N}(a_{mv}, \sigma_{mv}^2 R^{-1}/N) \quad and$$
$$\hat{\sigma}^2 \;\dot{\sim}\; \mathcal{N}(\sigma_{mv}^2\ , \sigma_{mv}^2(r_o - \sigma_{mv}^2)/N)\ .$$

The proof of this theorem is similar to that in [5] (p.352) for the case where U_t is absent. Our proof uses similar types of convergence results for random variables as those used in [5], but combined with a continuity property of means and variances, along with Theorem 1 to obtain normality. In [5] normality was obtained using a martingale difference argument which does not apply here. This theorem also extends the result of [6], which is essentially the same as that of [5], to the mixed spectrum setting.

To continue on to the statistics of the AR(n) spectral estimator, define $z = z(\omega) \triangleq [e^{-in\omega},,\ldots,1]^{tr}$,

$$\hat{\rho}(e^{i\omega}) = z^*[1\ ,\hat{a}^{tr}\ ]^{tr} \text{and } \rho_{mv}(e^{i\omega}) \triangleq z^*\ [1,a_{mv}^{tr}\ ]^{tr}.$$

Here, $*$ denotes the conjugate transpose. Then

$$\hat{S}(\omega) = \frac{\hat{\sigma}^2}{|\,\hat{\rho}(e^{i\omega})\,|^2} \text{ and } S_{mv}(\omega) = \frac{\sigma_{mv}^2}{|\,\rho_{mv}(e^{i\omega})\,|^2}.$$

We will also require the Capon nth order spectrum

$$S_{cap}(\omega) \triangleq n/\ z^* R^{-1} z. \qquad (8)$$

Our second result is then the following theorem.

THEOREM 3. For large N we have

$$\hat{\rho}(e^{i\omega}) \dot{\sim} N(\ \rho_{mv}(e^{i\omega})\ ,\ \frac{n\ \sigma_{mv}^2}{S_{cap}(\omega)N}\ ).$$

The correlation coefficient, γ, between $\hat{\rho}(e^{i\omega})$ and $\hat{\sigma}^2$ is given by

$$\gamma \;\cong\; \frac{\rho_{mv}(e^{i\omega}) S_{mv}(\omega)}{n(r_o - \sigma_{mv}^2)^{1/2} N}\ .$$

It follows from theorems 2 and 3 that for large N, the AR(n) spectral estimator has a distribution which is the ratio of the normal random variable and a non-central chi-squared random variable with two degrees of freedom. Moreover, for large N these random variables are approximately independent. To gain further insight into this estimator we use a first order Taylor expansion to obtain approximate expressions for its mean and variance. This results in

$$E[\hat{S}(\omega)] \cong S_{mv}(\omega) \left\{ \frac{1}{1+\frac{n\, S_{mv}(\omega)}{N\, S_{cap}(\omega)}} \right\} \quad ; \qquad (9a)$$

and

$$Var[\hat{S}(\omega)] \cong \frac{4n\sigma^2_{mv}(r_o - \sigma^2_{mv})S_{mv}(\omega)}{N^2 S_{cap}(\omega)} . \qquad (9b)$$

It must be emphasized that all of the above results are for a specified model order, n. Explicit dependence on n of any quantity has been omitted for notational convenience only. Expressions (9) offer some interesting insight. In particular, the nth order Capon spectrum plays a role in both expressions. It is well known that this spectrum can also be expressed as

$$S_{cap}(\omega) = \left\{ \sum_{k=0}^{n} [S^{(k)}_{mv}(\omega)]^{-1} / n \right\}^{-1} \qquad (10)$$

where $S^{(k)}_{mv}(\omega)$ is the kth order minimum variance spectral approximant. Since (10) involves a sort of averaging of higher resolution AR spectra, it is also well known that its resolution is notably less than the AR(n) spectral approximant [7]. Hence, the mean (9a) is liable to experience significant bias in the region of strong narrowband spectral components, and in particular, near point spectrum frequencies. The variance (9b) will also be influenced in these regions, possibly in an oscillatory manner, due to the spectral oscillations in the AR(n) spectrum induced by the point spectrum [7].

Next, we consider the properties of $\hat{S}^{(n)}(\omega)$ as $n \to \infty$. From the fact that the AR(n) spectral approximant converges to $S_x(\omega)$ at all point of continuity, it follows that the nth order Capon spectral approximant also converges to the same. Thus, at frequencies sufficiently removed from the point spectrum the above theorem, along with (9) and (10), give the large n statistical description of the AR(n) spectral estimator. They also show that the condition $n/N \to 0$ is sufficient for reasonable behavior for large order and data lengths. The difficulty in identifying this large n behavior in the mixed spectrum setting is greatest near the point spectrum frequencies. In fact, it required two theorems in [4] to rigorously prove that the AR(n) spectral approximant becomes unbounded at these frequencies as $n \to \infty$. This behavior, however, is a trivial consequence of the convergence results in [2].

Example. For the process (1a) let X_t be AR(2) with $a = [1.13\ ,\ -0.64]'$, and let U_t be a single sinusoid with $A = 1$ and $\omega = \pi/4$. It follows that the AR(2) mv approximant values are a_{mv}=[-.94 , .57]' and σ^2_{mv}=1.61. For N=100 samples per record, theorem 2 yields the predicted approximate distributions

$$\hat{a} \sim \mathcal{N}\left(\begin{bmatrix} -.94 \\ .57 \end{bmatrix}, \begin{bmatrix} .007 & -.004 \\ -.004 & .007 \end{bmatrix} \right) ; \hat{\sigma}^2 \sim \mathcal{N}(1.61, .034).$$

Using 1000 simulations of (1a) we obtained estimates

$$\hat{a} \sim \mathcal{N}\left(\begin{bmatrix} .93 \\ -.57 \end{bmatrix}, \begin{bmatrix} .007 & -.003 \\ .003 & .007 \end{bmatrix} \right) ; \hat{\sigma}^2 \sim \mathcal{N}(1.64, .050).$$

The approximate normality of $\hat{\sigma}^2$ and $\hat{a}$ are reflected in Figure 1 and Figure 2, respectively.

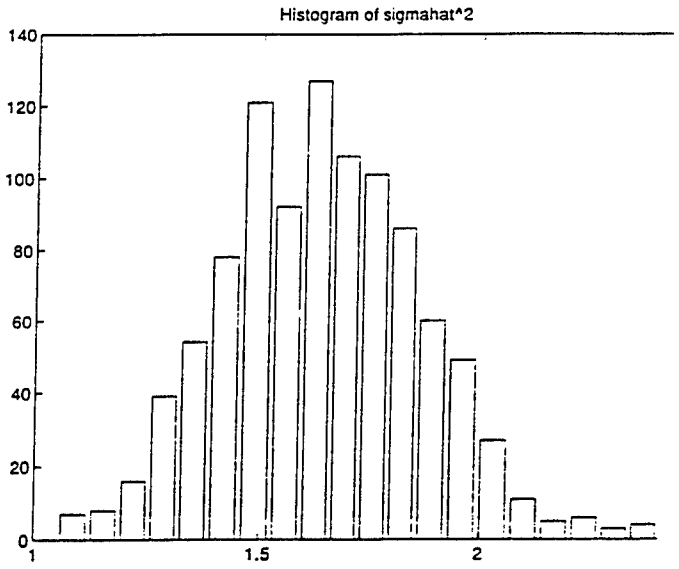


Figure 1. Histogram for $\hat{\sigma}^2$.

The accuracy of the mean approximation (9a) is shown against the sample mean in Figure 3. A comparison of (9b) and the sample variance, however, revealed major difference at all but very high frequencies far removed from the sinusoid. The sample variance was two orders of magnitude higher than (9b) in the region of the tone. At this stage it is not known whether this difference is due to the sample size, N, to the Taylor series approximation, or to some combination of the two.

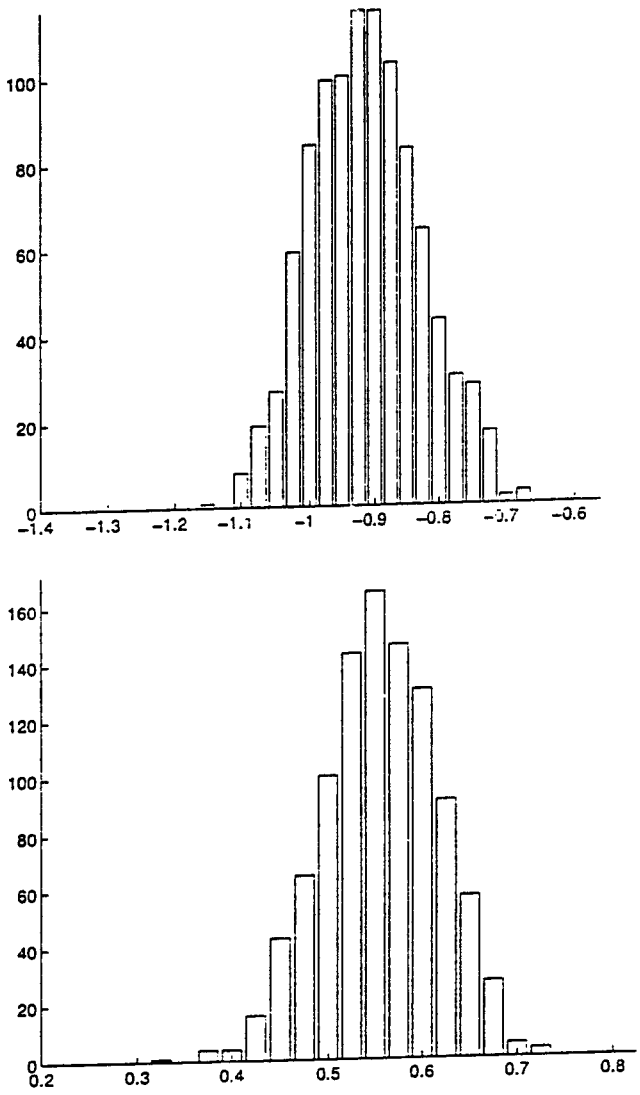


Figure 2. Histograms for $\widehat{a}_1$(top) and $\widehat{a}_2$(bottom).

Conclusions

The above results provide a partial description of the statistical behavior of AR spectral estimators for random processes having mixed spectrum. The example illustrated the claims of mv parameter estimator normality for record sizes as small as N=100 samples. Furthermore, the approximate expression for the mean of the AR spectrum compared well against the sample mean. While not shown here, investigation of the normal distribution claim in Theorem 3 also proved reasonable for the above example. But The variance expressions in both (9b) and in Theorem 3 were nowhere near to the sample variances at any but the very highest frequencies. A further investigation for this and other ARMA types of noise processes supported the rate dependence of the variance on N, but actual variances were dramatically different from predicted ones, except in the most simple casees where the noise was essentially white.

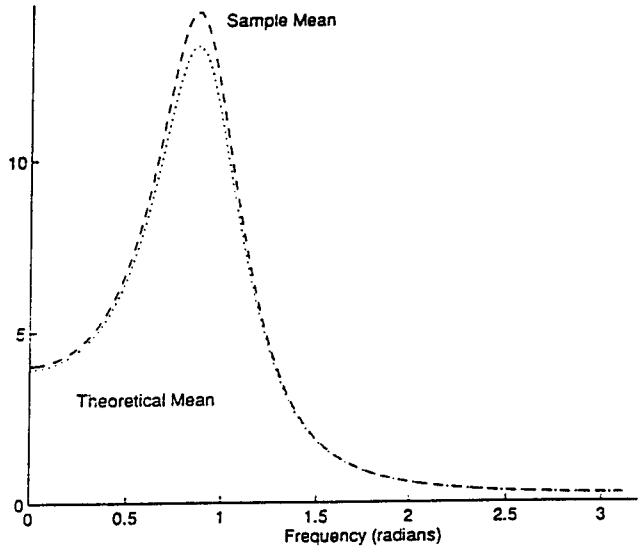


Figure 3. Comparison of (9a) and simulation estimate.

References

[1] Sherman,P & White, L. "Periodic spectral analysis of diesel vibration data" J. Amer. Acous. Soc., November 1994.

[2] Foias,C., Frazho,F. & Sherman, P. "A geometric approach to the ML spectral estimator for sinusoids in noise", IEEE IT-34(5), 1066-1070, 1988.

[3] Li,T., Kedem,B. & Yakowitz,S. "Asymptotic normality of sample autocovariances with an application in frequency estimation", Stoch. Poc. & Appl.,52,329-349, 1994.

[4] Mackisack,M. & Poskitt,D. "Some properties of autoregressive estimates for processes with mixed spectra" J. Time Series Anal., 11(4), 325-337, 1990.

[5] Caines,P. Linear Stochastic Systems, John Wiley & Sons, 1988.

[6] Stoica,P., Nehorai,A. & Kay,S. "Statistical analysis of LS AR estimator in the presence of noise", IEEE ASSP-35(9), 1273-1281, 1987.

[7] Sherman,P. & Lou,K. "On the family of ML spectral estimates for mixed spectrum identification", IEEE ASSP-39(3), 644-655, 1991.

[8] Szego, G. Orthogonal Polynomials, Providence, RI, Amer. Math. Soc., 1975.

AR Spectral Estimation with Randomly Missed Observations

Sina Mirsaidi and Jacques Oksman
École Supérieure d'Électricité - Service des Mesures
Plateau de Moulon
91192 Gif-sur-Yvette France
Mirsaidi@soleil.supelec.fr, Oksman@supelec.fr

Abstract

This paper represents a new spectral estimation method for the time series with missed observations. An Auto-Regressive modeling approach is adopted. The AR parameters are estimated by optimizing a weighted mean-square error criterion. The method can be used in real-time, adaptive contexts where the AR parameters are time varying. In general, both regularly and randomly missed observations can be handled by this method. The spectral estimates are compared to those obtained by well known AR parameter estimators used in the cases where none of the signal samples is missed. The performance of the method is illustrated by some numerical examples.

1. Introduction

In many practical situations, periodically sampled signals with missed observations may be encountered. This is caused by a variety of reasons such as accidentally loss of some portions of data, failure of the measurement equipment, etc. In some applications where data compression is needed, one may wish to reduce the whole number of data samples. This may result in a periodically sampled signal with "missed" observations.

Some important recent works in this field are [1][2][3]. Jones [1] has developed a maximum likelihood algorithm for ARMA time series with missed observations. He uses state-space representation and Kalman filtering to compute the likelihood function of the ARMA parameters and this function is then maximized using some non linear optimization procedure. Rozen and Porat [2] have developed an algorithm for the problem of spectral estimation through the ARMA modeling of stationary processes with missing observations. This algorithm is asymptotically optimal in the sense of achieving the smallest error-variance when the number of data approaches infinity. All of the mentioned methods handle *only* stationary time series and *cannot* be used in an adaptive context where the AR parameters are time varying.

In this paper, we present a new method of AR spectral estimation when the data are not consecutive, but some of the observations are missed. In general, both regularly and randomly missed observations can be handled by this method [4]. The method is based on non-linear optimization of a weighted squared error criterion. All the formulae obtained are recursive, and real-time spectral estimation of *non stationary* signals can also be handled [5].

2. Description of the method

The basic idea of this method is very similar to that of the methods used in adaptive identification contexts (RLS, LMS, ...). For the convenience and without loss of generality, in what follows, we suppose that the period of sampling is equal to 1.

We suppose that $\{y_n\}$ is a discrete time zero-mean AR process defined as follows :

$$y_n = \boldsymbol{\theta}^T \boldsymbol{y}_n + v_n \qquad (1)$$

where v_n is a zero-mean white process with variance σ_v^2, $\boldsymbol{\theta}^T = [\theta_1, \ldots, \theta_M]$ is the vector of the AR parameters and $\boldsymbol{y}_n^T = [y_{n-1}, \ldots, y_{n-M}]$ is the vector of the last M signal samples, M being the order of the AR model. We suppose that the signal $\{y_n\}$ is subjected to random skipping or deletion of some samples. Let $\{t_1, \ldots, t_n\}$ be the set of instants where the signal samples are *not* missed. Our aim is to compute the vector $\boldsymbol{\theta}$ that minimizes the following cost function :

$$J_{t_n} = \frac{1}{n}\sum_{i=1}^{n} \omega_{t_i} e_{t_i}^2 = \frac{1}{n}\sum_{i=1}^{n} \omega_{t_i}(y_{t_i} - \hat{y}_{t_i})^2 \qquad (2)$$

where e_{t_i} is the prediction error at instant t_i and $\hat{y}_{t_i}$ is the estimate of y_{t_i}. In order to compute the value of $\hat{y}_{t_i}$, we use a well known result of the prediction theory that is recalled below.

0-8186-7576-4/96 $5.00

The optimal k-step-ahead linear predictor let $\{y_n\}$ be defined as in (1) and $y_n, \ldots, y_{n-M}$ be known. The best linear mean square estimation of y_{n+k} is obtained by the following recursion :

$$\hat{y}_{n+k} = \sum_{i=1}^{M} \theta_i \hat{y}_{n+k-i} \qquad (3)$$

This means that at each instant t_i, in order to obtain the value of $\hat{y}_{t_i}$, one has to use the recursion (3) for $n = t_{i-1}$ and for $k = 1, 2, \ldots, t_i - t_{i-1}$. In addition, each missed sample y_j where $j < t_{i-1}$, has to be replaced by its estimated value. The algorithm can be summarized as below :

1. Computation of $\hat{y}_{t_n}$ at each instant t_n, using the optimal linear predictor described above,
2. Non linear optimization of the cost function J_{t_n},
3. Prediction of missed samples between instants t_n and t_{n+1} using the last estimated AR parameters.

At step 2, one has to compute the gradient of the cost function $\frac{\partial J_{t_n}}{\partial \theta}$. This is subsequently used in some non linear optimization procedure to minimize J_{t_n}. Formal description of the algorithm is given in [6]. Details of the gradient computation can be found in [4][7] and is not given here but what is important is that the gradient can be updated recursively at each instant t_n. In addition, the use of an iterative optimisation procedure (descent algorithms such as : gradient or variable-metric methods) together with an exponential weighting factor such as $\omega_{t_i} = \lambda^{t_n - t_i}$, afford the possibility of operating in non- stationary environments.

3. Some discussions about the cost function

Eq. 3 shows that y_{t_n} is a polynomial function of θ and hence, J_{t_n} is not a quadratic cost function as it may be superficially expected, it is rather a polynomial whose degree at instant t_n depends on the number of missed observations until t_n. This may cause the problem of convergence to a local minimum and not necessarily to the global one. One solution may be to repeat the algorithm with several initial values to increase the chance of finding the global minimum. However, the cost function J_{t_n} has some interesting properties, at least in some special cases. For example, the following proposition has been proved [4].

Proposition *Suppose that $\{y_n\}$ is an arbitrary AR(1) process with parameter θ^*. Assume that the random pattern of misses is a Bernoulli-type one in which each measurement has a fixed probability $q = 1 - p$ of being missed and that the misses are independent. If we define the cost function as below :*

$$J_{t_n} = E\left(\omega_{t_n} e_{t_n}^2\right) = E\left(\omega_{t_n} (y_{t_n} - \hat{y}_{t_n})^2\right) \qquad (4)$$

where $\hat{y}_{t_n}$ is obtained by the method described above and if we set $\hat{y}_1 = y_1$, then J_{t_n} is a convex polynomial of degree $n-1$ with the minimum at $\theta = \theta^*$.

The proposition describes the statistical behavior of the present method in the case of AR(1) processes. In several examples tested in the case of AR(2) processes, only one minimum has been observed for the cost function. The extension of the proposition to AR(p) processes has not yet been done. However, the simulations, partly discussed in the following section, give satisfying results in the cases where the AR process has a larger order.

4. Simulations

In all the examples, we consider the random Bernoulli pattern of misses where it is supposed that each sample has the probability $q = 1 - p$ of being missed and the misses are independent.

Example 1 In this example we illustrate the performance of the proposed algoritm in spectral reconstruction. In each case, and as a reference for comparison, the spectral estimate obtained by a classical AR estimator in the case where none of the samples is missed, is also given. The approach used in this case is the forward-backward approach where the sum of least-squares criterion for a forwad model and the analogous criterion for a time-reversed model is minimised [8].

The first test spectrum is a two peak one. It is supposed that the two peaks represent the sum of two zero-mean independent signals. Each signal is obtained by filtering white noise by a first order Butterworth filter. The experiment is repeated 100 times, each time using an independent realisation of the test signal. The average spectral estimate is obtained by computing the average of the estimates over these 100 independent trials of the experiment. The normalised frequencies and bandwidths of the peaks are :

$$f_1 = 0.3, \quad \Delta f_1 = 0.005 \quad f_2 = 0.35, \quad \Delta f_2 = 0.005 \qquad (5)$$

The probability of missing each sample is $q = 0.4$. There is approximately 15 dB of difference between the amplitudes of the sharp peaks. Fig.1 shows plots of the average estimated spectra for both cases : *with* and *without* missed samples. We note the correct reconstruction of more informative portions of the spectrum.

The second test spectrum is that of the vowel "i" (in French!) spoken by a male speaker. The probability of

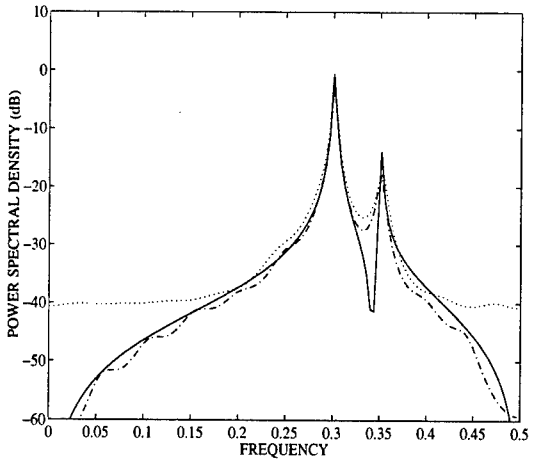


Figure 1. Spectral reconstruction. Original PSD: _____, estimated PSD: $q = 0\% : -.-.-, q = 40\% : ...., (M = 20)$.

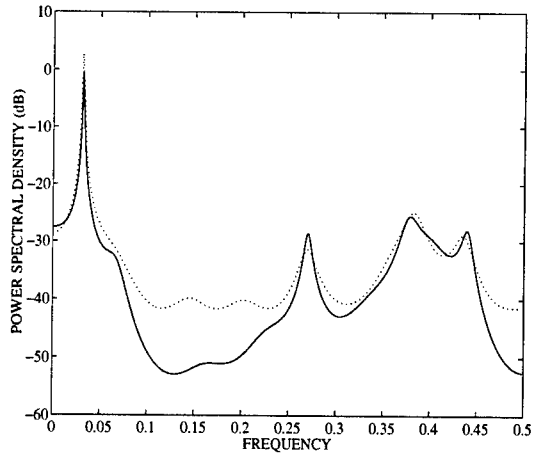


Figure 2. Spectral reconstruction. Estimated PSD : $q = 0\% :$ ____, $q = 40\% : ...., (M = 20)$.

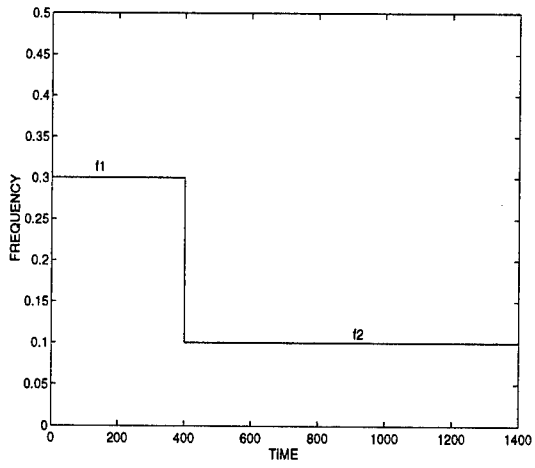


Figure 3. The frequency variation of the test signal in example 2.

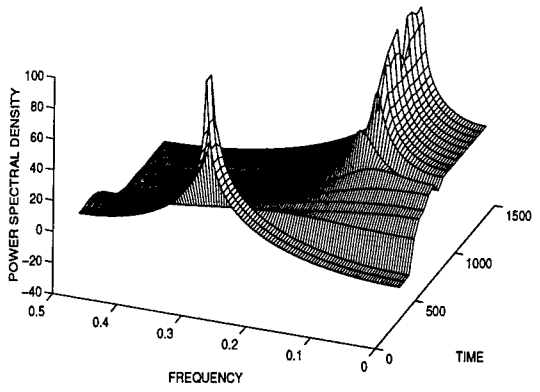


Figure 4. Time-frequency evolution of the spectral estimate for the test signal in example 2. $\lambda = 0.99, q = 40\%$.

missing each sample is $q = 0.4$. The estimated spectra are shown in Fig.2.

Example 2 In this example the performance of the proposed algorithm in the non-stationary environments is studied. In order to test the parameter tracking capacity of the method, we have considered a sinusoid that is subjected to an abrupt change in frequency as demonstrated in Fig.3. The period of sampling is $T = 1$. The AR model order is $M = 2$. Fig.4 shows the time-frequency evolution of the spectral estimates. We note the correct estimation of the frequencies $f_1 = 0.3$ and $f_2 = 0.1$. The evolution of the AR parameters as a function of time is shown in Fig.5. We note that the choice of a forgetting factor $\lambda = 0.99$ increases considerably the tracking capacity of the algorithm.

Eample 3 The convergence behavior of the mean squared prediction error for different values of q is illustrated here. The test signal is an AR(2) pocess with the parameters $\theta^T = [1 \ -0.3 \ 0.5]$. Fig.6 shows the average results obtained from 100 independent realisations of the AR(2) process. It is important to note that the speed of convergence is the same for different values of q. Clearly, the residulal error is greater for larger values of q. This is obviously because of the accumulation of the errors due to missed sample estimations.

From the previous and numerous other examples, the

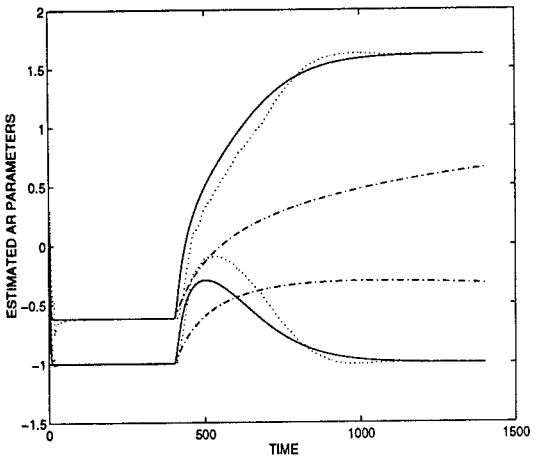


Figure 5. AR estimation for the test signal in example 2. $\lambda = 0.99, q = 0\% :$ ____, $\lambda = 0.99, q = 40\% :$ ....., $\lambda = 1, q = 0\% : -.-.$

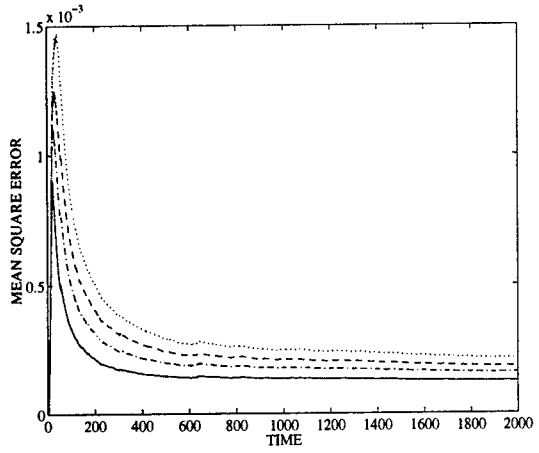


Figure 6. Convergence behavior of the mean squared error. $(M = 2), q = 0\% :$ ____, $q = 20\% : -.-., q = 30\% :$ - - - - $q = 40\% :$ .....

following points are noted :

- The performance of the AR estimators in both of the cases (*with* and *without* missed samples) are similar, particularly in the more informative zones of the spectrum.
- In the case where some of the samples are lost, a residual noise level is observed in the spectral estimates. This becomes more pronounced as the number of lost samples increases. The level is situated at -40 dB for the signals in the example 1. The reason is obviously the lack of information from the signal.
- A higher order AR estimator is needed to resolve neighbouring spectral peaks with the same fidelity as in single-peak cases. For peaks with greater amplitude ratios, higher model orders should be used.
- In the case of spectra with larger bandwidths, one must choose larger model orders in order to have spectral estimates with the same fidelity as in the case of sharp peak spectra. This is because the AR models are less adapted to these kinds of spectra than those with sharp peaks.

5. Conclusion

We presented a parametric spectral estimation technique for signals with incomplete data based on AR modeling. The method is adaptive and can be applied to non stationary cases. Both regularly and randomly missed data can be handled. Simulation results show the high performance of this method even in the cases where a large number of samples is lost.

References

[1] R. H. Jones. Fitting a continuous time auto regression to discrete data. *Applied times series analysis III*, pages 651–682, 1981.

[2] Y. Rozen and B. Porat. Optimal ARMA parameter estimation based on the sample covariances for data with missing observations. *IEEE Trans. Inform. Theory*, 35(2):342–349, 1989.

[3] E. A. Parzen. Times series analysis of irregularly observed data. *Lect. Notes Stat*, 25, 1984.

[4] S. Mirsaidi. AR identification in the case of missing observations. Technical report, École supérieure d'Électricité, Gif-sur-Yvette, FRANCE, 1995.

[5] S. Mirsaidi and J. Oksman. Adaptive signal reconstruction from incomplete data. *Proc. 30th CISS, Princeton, NJ*, March 1996.

[6] S. Mirsaidi and J. Oksman. Extension d'algorithmes de type gradient au cas d'échantillonnage non périodique. *Proc. Quinzièm Colloque Gretsi, Juans les Pins, FRANCE*, 1:415–418, September 1995.

[7] S. Mirsaidi, G. Fleury, and J. Oksman. LMS-like AR modeling in the case of missing observations. Submitted to . *IEEE trans. Acoust., Speech, Signal Processing.*

[8] S. Haykin. *Adaptive filter theory.* Prentice Hall, 1986.

An Adaptive Modified Covariance Algorithm For Spectral Analysis

Kyriakos Kitsios*, Andreas Spanias*, Bruno Welfert**, Philipos Loizou*

*Department of Electrical Engineering
**Department of Mathematics
Arizona State University, Tempe, AZ 85287, USA

Abstract

An optimum Block Modified Covariance Algorithm is developed for computing time-varying autoregressive (AR) parameters. The method presented here differs from those presented previously [3] in that it uses optimally selected time-varying convergence factors such that the block mean square error is minimized from one iteration to the next. In particular, the algorithm developed here, called Block Modified Covariance Algorithm with individual adaptation of parameters (BMCAI), uses individual time-varying convergence factors computed using modified covariance matrix approximations along with the Gauss-Seidel method. Even though the BMCAI is gradient based it retains the attractive spectral matching properties of fixed-window least squares modified covariance algorithms while at the same time providing capabilities for time-varying spectral estimation.

1. Introduction

This paper is concerned with the development of an efficient algorithm for least-squares forward-backward prediction (FBP). Unconstrained FBP requires matrix inversion and most of the originally proposed algorithms compute AR parameters based on a fixed-window approach. Marple developed a fast Cholesky algorithm (FCA) which requires $O(p^2)$ operations and more recently a fast QR algorithm (FQRA) [1] which was shown to have improved numerical behavior relative to the FCA. The fast inversion algorithms [1] are order recursive and operate on a fixed N-point record, i.e., they are non-adaptive. A family of fixed-order sliding-window block gradient algorithms for FBP, namely the block modified covariance algorithms (BMCAs), were proposed recently by Spanias [3]. In particular, the BMCA worked reasonably well in a series of "benchmark" simulations, however its performance deteriorated considerably in scenario requiring estimation of the spectral content of multiple closely-spaced sinusoids. This is mainly because the BMCA uses a single convergence factor (or step size μ_B) which does not allow for fast adaptation in cases where the modified covariance matrix has large eigenvalue disparity. In this paper, we concentrate on the development of multiple convergence factors for adapting the AR parameters. The use of multiple convergence factors in adaptive FBP was motivated by work done in adaptive FIR system identification by Mikhael *et al* [2]. The difference between the algorithms presented in this paper and those presented by Mikhael are: a) the algorithms presented here are intended for modified covariance linear prediction in which the structure of the equations to be solved is distinctly different than that encountered in FIR system identification, b) the algorithms presented are studied in the context of spectral estimation applications and deal with the idiosyncrasies of some of complex spectral estimation examples such as multiple closely spaced sinusoids, and c) the proposed methods go a step beyond Mikhael's work in the sense that the computation of the individual μ_B is done efficiently using fast and stable Gauss-Seidel numerical methods tailored specifically to deal with the structure of the modified covariance equations. The latter is the most important contribution of the paper in that it provides opportunities for reducing the complexity of the algorithms by using approximates of the modified covariance matrix while maintaining the attractive performance characteristics of least squares MC spectral estimators.

The rest of the paper is organized as follows. Section 2 presents the BMCA and Section 3 describes an algorithm that uses individual step sizes for adapting the AR parameters (BMCAI). An efficient Gauss-Seidel it-

0-8186-7576-4/96 $5.00

erative procedure for computing the optimum convergence factors is also presented in this section. Section 4, presents simulations using the BMCAI and Section 5 gives the conclusions.

2. The Block MC Algorithm

In this section, a general technique for formulating the BMCAI is presented. We begin by defining the following parameters: let i be the block index, p the order of the AR model, N the number of samples for prediction, $2L$ the length of the processed block, n the time index, $a_k(i)$ the k-th adjustable parameter in the i-th block ($k = 1, 2, 3, ..., p$), $x(n)$ the input signal for linear prediction (adaptive filter), $e_\ell(i)$ the ℓ-th error signal in the i-th block ($\ell = 1, 2, ..., 2L$), and S the number of samples per block shift.

At the i-th iteration, the objective is to minimize the cost function $J(i+1) = \frac{1}{2}\boldsymbol{e}_{fb}^T(i+1)\boldsymbol{e}_{fb}(i+1)$ where the $2L \times 1$ error vector $\boldsymbol{e}_{fb}(i)$ is given by

$$\begin{aligned}\boldsymbol{e}_{fb}(i) &= [e_f(iS+p+1)...e_f(iS+N) \\ &\quad e_b(iS+p+1)...e_b(iS+N)]^T \end{aligned} \quad (1)$$

and $e_f(n)$ and $e_b(n)$ are the forward and backward prediction errors

$$e_f(n) = x(n) - \sum_{k=1}^{p} a_k(i)x(n-k), \quad (2)$$

$$e_b(n) = x(n-p) - \sum_{k=1}^{p} a_k(i)x(n-p+k). \quad (3)$$

Equations (1), (2) and (3) can be written block-wise as

$$\boldsymbol{e}_{fb}(i) = \boldsymbol{x}(i) - \boldsymbol{X}_{fb}(i)\boldsymbol{a}(i) \quad (4)$$

where the $2L \times 1$ vector $\boldsymbol{x}(i)$ is given by

$$\begin{aligned}\boldsymbol{x}(i) &= [x_f(iS+p+1).....x_f(iS+N) \\ &\quad x_b(iS+1).....x_b(iS+N-p)]^T \end{aligned} \quad (5)$$

and the $2L \times p$ matrix $\boldsymbol{X}_{fb}(i)$ and $p \times 1$ vector $\boldsymbol{a}(i)$ are defined by

$$\boldsymbol{X}_{fb}(i) = \begin{bmatrix} x(iS+p) & ... & x(iS+1) \\ x(iS+p+1) & ... & x(iS+2) \\ . & . & . \\ x(iS+N-1) & ... & x(iS+N-p) \\ x(iS+2) & ... & x(iS+p+1) \\ x(iS+3) & ... & x(iS+p+2) \\ . & . & . \\ x(iS+N-p+1) & ... & x(iS+N) \end{bmatrix}$$

$$\boldsymbol{a}(i) = [a_1(i), a_2(i), ..., a_p(i)]^T. \quad (6)$$

The BMCA uses the following update formula $\boldsymbol{a}(i+1) = \boldsymbol{a}(i) - \mu\nabla_{fb}(i)$, with $\nabla_{fb}(i) = -\frac{2}{2L}\boldsymbol{X}_{fb}^T(i)\boldsymbol{e}_{fb}(i)$. The condition for convergence of the algorithm is $0 < \mu < 2L/\lambda_{max}$, where λ_{max} is the largest eigenvalue of $E(\boldsymbol{X}_{fb}^T(i)\boldsymbol{X}_{fb}(i))$.

3. The BMCAI

In this section, we propose the use of individual convergence factors that are optimally chosen to adapt *individual filter parameters*. The step sizes are updated at each block iteration.

3.1. Problem formulation

We now consider the relation

$$\boldsymbol{a}(i+1) = \boldsymbol{a}(i) - \boldsymbol{M}(i)\nabla_{fb}(i) \quad (7)$$

to update the parameters, where $\boldsymbol{M}(i)$ is a $p \times p$ diagonal matrix containing the p convergence factors, i.e.,

$$\boldsymbol{M}(i) = \begin{bmatrix} \mu_1(i) & & \\ & \ddots & \\ & & \mu_p(i) \end{bmatrix}. \quad (8)$$

As in all block gradient algorithms, the block gradient vector $\nabla_{fb}(i)$ is replaced by an estimated block gradient vector which is given by

$$\widehat{\nabla}_{fb}(i) = \frac{1}{L}\frac{\partial J(i)}{\partial \boldsymbol{a}(i)} = -\frac{1}{L}\boldsymbol{X}_{fb}^T(i)\boldsymbol{e}_{fb}(i). \quad (9)$$

From (7), (8), (9) one obtains the general form of the parameter updating formula in matrix vector form as:

$$\boldsymbol{a}(i+1) = \boldsymbol{a}(i) + \frac{1}{L}\boldsymbol{M}(i)\boldsymbol{X}_{fb}^T(i)\boldsymbol{e}_{fb}(i). \quad (10)$$

In the parameter update (10), there are p individual time-varying convergence factors, $\mu_k(i)$ ($k = 1, 2, ..., p$). These factors are chosen at each iteration i so as to minimize the functional $J(i+1)$. To this end, the forward and backward errors are expanded using the truncated Taylor series

$$\begin{aligned}\boldsymbol{e}_{fb}(i+1) &= \boldsymbol{e}_{fb}(i) + \frac{\partial \boldsymbol{e}_{fb}(i)}{\partial \boldsymbol{a}(i)}\Delta\boldsymbol{a}(i) \\ &= \boldsymbol{e}_{fb}(i) - \boldsymbol{X}_{fb}(i)(\boldsymbol{a}(i+1) - \boldsymbol{a}(i)) \\ &= \boldsymbol{e}_{fb}(i) - \frac{1}{L}\boldsymbol{X}_{fb}(i)\boldsymbol{M}(i)\boldsymbol{X}_{fb}^T(i)\boldsymbol{e}_{fb}(i) \\ &= \boldsymbol{e}_{fb}(i) - \boldsymbol{X}_{fb}(i)\boldsymbol{M}(i)\boldsymbol{q}(i) \end{aligned} \quad (11)$$

with $\boldsymbol{q}(i) = \frac{1}{L}\boldsymbol{X}_{fb}^T(i)\boldsymbol{e}_{fb}(i) = -\nabla_{fb}(i)$. Here the partial derivative $\frac{\partial \boldsymbol{e}_{fb}(i)}{\partial \boldsymbol{a}(i)}$ is obtained from (4) and reduces to $-\boldsymbol{X}_{fb}(i)$.

The next step is to choose $\boldsymbol{M}(i)$ such that $J(i+1)$ is minimized. This is done by setting

$$\frac{\partial J(i+1)}{\partial \mu_k(i)} = 0 \tag{12}$$

for $k = 1, ..., p$. This leads to the system of equations

$$\boldsymbol{q}^T(i)\boldsymbol{M}(i)\boldsymbol{R}(i)\frac{\partial \boldsymbol{M}(i)}{\partial \mu_k(i)}\boldsymbol{q}(i) = \boldsymbol{q}^T(i)\frac{\partial \boldsymbol{M}(i)}{\partial \mu_k(i)}\boldsymbol{q}(i) \tag{13}$$

for $k = 1, ..., p$, or

$$\boldsymbol{R}(i)\boldsymbol{M}(i)\boldsymbol{q}(i) = \boldsymbol{q}(i). \tag{14}$$

Equivalently,

$$\boldsymbol{M}(i)\boldsymbol{q}(i) = \boldsymbol{R}^{-1}(i)\boldsymbol{q}(i), \tag{15}$$

Therefore the updating formula (7) becomes

$$\boldsymbol{a}(i+1) = \boldsymbol{a}(i) + \boldsymbol{M}(i)\boldsymbol{q}(i) = \boldsymbol{a}(i) + \boldsymbol{R}^{-1}(i)\boldsymbol{q}(i). \tag{16}$$

The last equation is the weight update equation for the BMCAI with individual adaptation of parameters. Its main drawback is the requirement of computing the solution of a system of equations of order p. The associated cost can become intolerable especially for high-order prediction. The following section gives an approach which can be used to approximate $\boldsymbol{R}^{-1}(i)\boldsymbol{q}(i)$ in an efficient manner.

3.2. Implementation via a Gauss-Seidel Iteration

The matrix inversion for computing the vector $\boldsymbol{R}^{-1}(i)\boldsymbol{q}(i)$ in (16) can be avoided altogether by solving the system

$$\boldsymbol{R}(i)\boldsymbol{z}(i) = \boldsymbol{q}(i) \tag{17}$$

for $\boldsymbol{z}(i)$ via an iterative method (which only requires matrix-vector products), then updating

$$\boldsymbol{a}(i+1) = \boldsymbol{a}(i) + \boldsymbol{z}(i). \tag{18}$$

More precisely, $\boldsymbol{z}(i)$ is replaced by $\boldsymbol{z}^{(k)}(i)$ obtained by applying k iterations

$$\boldsymbol{z}^{(k)}(i) = \boldsymbol{z}^{(k-1)}(i) + \boldsymbol{Q}^{-1}(i)\left(\boldsymbol{q}(i) - \boldsymbol{R}(i)\boldsymbol{z}^{(k-1)}(i)\right) \tag{19}$$

starting with a given vector $\boldsymbol{z}^{(0)}(i)$. Here $\boldsymbol{Q}(i)$ is a matrix approximating $\boldsymbol{R}(i)$. Since the system (17) is symmetric and generally positive definite, for efficiency we will only consider Gauss-Seidel iterations, i.e.,

$$\boldsymbol{Q}(i) = \boldsymbol{D}(i) + \boldsymbol{L}(i), \tag{20}$$

Algorithm	Multiplies	Additions
BMCA	$4Lp + p$	$4Lp$
BMCAI	$p(4N - 3p/2 + 3/2)$	$p(4N - 3p/2 - 1/2) + 1$

Table 1. Computational Complexity of BMCA algorithms $L = N - p$.

where $\boldsymbol{D}(i)$ and $\boldsymbol{L}(i)$ are the diagonal and (strictly) lower triangular parts of $\boldsymbol{R}(i)$, respectively. Note that the matrix $\boldsymbol{R}(i)$ is not always diagonally dominant (at least for the input data used), which explains why the Jacobi method (corresponding to $\boldsymbol{Q}(i) = \boldsymbol{D}(i)$) did not converge when applied to (17). In our experiments only 2 or 3 iterations were sufficient to obtain a good approximation of $\boldsymbol{z}(i)$ when starting with $\boldsymbol{z}^{(0)}(i) = 0$. For two iterations, this is equivalent to approximating $\boldsymbol{z}(i)$ by

$$\begin{aligned}\boldsymbol{z}^{(2)}(i) &= (\boldsymbol{D}(i) + \boldsymbol{L}(i))^{-1}. \\ &\left(\boldsymbol{q}(i) - \boldsymbol{L}^T(i)(\boldsymbol{D}(i) + \boldsymbol{L}(i))^{-1}\boldsymbol{q}(i)\right).\end{aligned} \tag{21}$$

In order to reduce the computational complexity of the algorithm the sum $\boldsymbol{D}(i) + \boldsymbol{L}(i)$ can be directly updated without forming $\boldsymbol{R}(i+1)$, by considering the lower triangular part (including the diagonal) of the recursion

$$\boldsymbol{R}(i+1) = \boldsymbol{R}(i) + \boldsymbol{V}^T(i+1)\boldsymbol{W}(i+1) \tag{22}$$

i.e.,

$$\boldsymbol{D}(i+1) + \boldsymbol{L}(i+1) = \boldsymbol{D}(i) + \boldsymbol{L}(i) + \boldsymbol{Y}(i+1) \tag{23}$$

where $\boldsymbol{Y}(i+1)$ is the lower triangular part of $\boldsymbol{V}^T(i+1)\boldsymbol{W}(i+1)$. Note that with $\boldsymbol{V}(i)$ and $\boldsymbol{W}(i)$ defined as

$$\boldsymbol{V}(i) = \begin{bmatrix} -\boldsymbol{x}_1(i) \\ \boldsymbol{x}_L(i) \\ -\boldsymbol{x}_{L+1}(i) \\ \boldsymbol{x}_{2L}(i) \end{bmatrix} \quad \text{and} \quad \boldsymbol{W}(i) = \begin{bmatrix} \boldsymbol{x}_1(i) \\ \boldsymbol{x}_L(i) \\ \boldsymbol{x}_{L+1}(i) \\ \boldsymbol{x}_{2L}(i) \end{bmatrix}. \tag{24}$$

where $\boldsymbol{x}_\ell(i)$ denotes the ℓ-th row of $\boldsymbol{X}_{fb}(i)$ ($\ell = 1, ..., 2L = 2(N-p)$), then $\boldsymbol{Y}(i+1)$ can be computed efficiently.

The computational complexity of the BMCAI relative to that of the BMCA is given in Table 1

4. Simulation Results

The performance of the BMCAI is examined in terms of its ability to resolve closely-spaced sinusoids

embedded in noise. The PSD obtained using the BMCAI compared favorably against that obtained with the BMCA. In Fig. 1, we show a simulation with 10 closely-spaced spectral peaks of a process given by

$$x(n) = \sum_{i=1}^{10} A_i \cos(\omega_i n) + W(n) \qquad (25)$$

for $n = 1, ..., 32$, with $A_i = 0.1i$, $\omega_i = \frac{2\pi(10+1.5(i-1))}{f_s}$ and $Q = 10^{-4}$ (noise variance). Here $f_s = 64$ is the sampling frequency (in Hertz) and $W(n)$ a pseudo-random white-noise sequence. The prediction order was taken to be equal to 32. The plot in Figure 1 are formed by overlapping the spectra obtained using the BMCAI with individual adaptation of parameters based on Gauss-Seidel iterations, for 10 independent realizations. Each realization is a 100-sample record of the above input time series. The relative phases change randomly from realization to realization. Note that although the sinusoids are very closely spaced in frequency and the available data records are quite short, the BMCAI tracks accurately the frequencies (Fig. 1a and b) without missing any spectral peak. The BMCA on the other hand (Fig. 1c) fails to resolve one of the peaks.

5. Conclusions

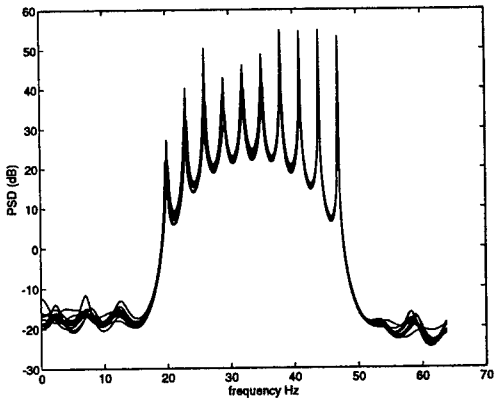
In this paper, the formulation of a block modified covariance algorithm with individual convergence factors (BMCAI) has been presented. The convergence factors are optimally selected to minimize the combined forward-backward squared error in each block. The BMCAI computes the individual convergence factors using Gauss Seidel iterations. The BMCAI has been applied in AR spectral estimation and outperformed the existing BMCA.

References

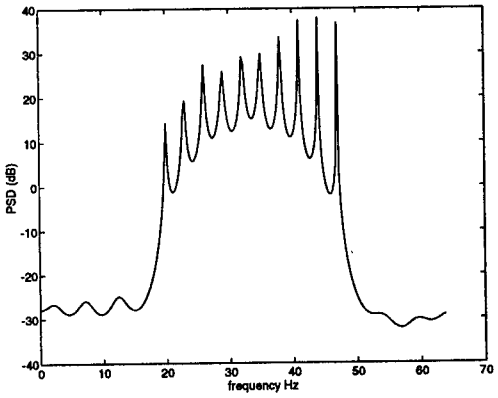
[1] S. L. Marple. A fast computational algorithm for the modified covariance method of linear prediction. *Digital Signal Processing 1*, New York: Academic(6):124–133, 1991.

[2] W. Mikhael and F. Wu. A fast block fir adaptive digital filtering algorithm with individual adaptation of parameters. *IEEE Transactions on Circuits and Systems*, cas-36(1):1–10, Jan. 1989.

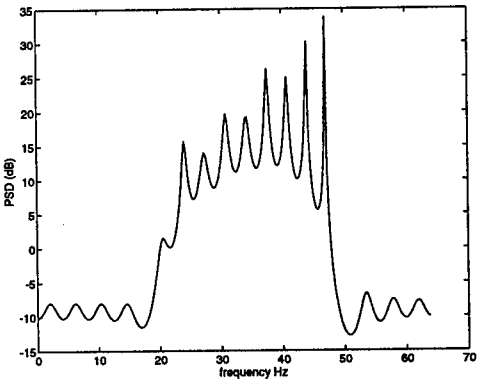
[3] A. S. Spanias. Block time and frequency domain modified covariance algorithms for spectral analysis. *IEEE Transactions on Signal Processing*, ASSP-41(6):2131–2140, June 1993.



(a)



(b)



(c)

Figure 1. (a) PSD estimation using the BMCAI based on Gauss-Seidel iterations with 10 realizations of 100-sample records, p=32 and SNR=42dB, (b) average of the ten simulations shown in (a), and (c) PSD using the BMCA with the same record and prediction order but with a fixed $\mu = 0.001$.

Simultaneous Registration and Tracking for Multiple Radars with Cluttered Measurements

Nickens N. Okello
Cooperative Research Centre for
Sensor Signal and Information Processing
SPRI Building, Technology Park,
The Levels, SA 5095, Australia
nickens.okello@cssip.edu.au

Graham W. Pulford
Dept. of Electrical and Electronic Engineering
Univ. of Melbourne, Parkville 3052, Australia
and Cooperative Research Centre for
Sensor Signal and Information Processing
g.pulford@ee.mu.oz

Abstract

Simultaneous registration and tracking has advantages over other track registration techniques because it is capable of responding to changes in registration errors. The track registration problem is presented for a network of two geographically distributed radars with unknown measurement biases that are fixed or slowly varying. The extended Kalman filter that receives unregistered and cluttered plots from the radars and outputs registered tracks, is used to carry out centralized simultaneous registration and tracking. A multisensor probablistic data association filter (PDAF) that combines locally gated plots from the radars is developed to enable the system operate under clutter. The algorithm satisfies a number of important registration design criteria.

1 Introduction

In multisensor tracking, registration is vital if errors due to site uncertainties, antenna orientation and improper caliberation of range and time are to be minimized. Errors that are fixed but unknown can be handled as part of a multisensor initialization procedure and a suitable off-line approach is the *generalized linear least-squares estimation* (GLSE) technique [1, pp. 180], [2, pp. 68].

Unfortunately sensor measurement biases can vary over time due to technical maintenance or the effect of a changing wind direction on the mechanics of a radar antenna [3, pp. 38]. This requires on-line estimation of biases and tracks under clutter using an algorithm that satisfies some basic registration design criteria [1, pp. 173].

In this paper we consider a system of two 2-D radar detectors A and B located at (η_1, ζ_1) and (η_2, ζ_2) respectively and responsible for a common cluttered surveillance region that is being traversed by a single non-maneuvering target T. We assume that the target position at time index k with respect to a common Cartesian coordinate system is $(x_1(k), x_2(k))$.

Furthermore each radar measures target position in polar coordinates with the origin of the measurement system being located at the radar antenna. We therefore assume that the target as reported by sensors A and B are at $T_A(\rho_1(k), \theta_1(k))$ and $T_B(\rho_2(k), \theta_2(k))$ respectively and that these measurements (or plots) include fixed but unknown biases $\delta\rho_i$ and $\delta\theta_i$ and measurement noise $v_i(k)$ for $i = 1, 2$. The measurement equations for the two radars therefore take on the form

$$\begin{bmatrix} \rho_i(k) \\ \theta_i(k) \end{bmatrix} = h_i(x_1(k), x_2(k)) + \begin{bmatrix} \delta\rho_i \\ \delta\theta_i \end{bmatrix} + v_i(k) \qquad (1)$$

where $v_i(k)$, $i = 1, 2$ are respectively zero-mean, mutually uncorrelated, white Gaussian noise processes of known covariance $R_i(k)$.

With the bias terms unknown it is not possible to determine the true target position. We must therefore attempt joint estimation of the target state and biases using target measurements from the two sensors. To do this, we append the radar biases to the target state to obtain the augmented vector

$$x(k) = [\, x_t^T(k) \quad b^T(k) \,]^T \qquad (2)$$

where $x_t(k) = [\, x_1(k) \;\; \dot{x}_1(k) \;\; x_2(k) \;\; \dot{x}_2(k) \,]^T$ is the target state and $b(k) = [\, \delta\rho_1(k) \;\; \delta\theta_1(k) \;\; \delta\rho_2(k) \;\; \delta\theta_2(k) \,]^T$ is the bias vector. The resulting process equation therefore takes on the form

0-8186-7576-4/96 $5.00

$$x(k+1) = \begin{bmatrix} F & 0 \\ 0 & I \end{bmatrix} x(k) + \begin{bmatrix} G_t w_t(k) \\ u(k) \end{bmatrix} \quad (3)$$

where I denotes an identity matrix of dimension 4 and $u(k)$ is a small process noise term. The equation can be more compactly written as $x(k+1) = F(k)x(k) + G(k)w(k)$, where $F(k)$ and $G(k)$ are known, and $w(k)$ is a zero-mean, white Gaussian noise process with covariance $Q(k)$. For centralized tracking, the combined measurement equation takes on the form

$$\begin{bmatrix} \rho_1(k) \\ \theta_1(k) \\ \rho_2(k) \\ \theta_2(k) \end{bmatrix} = \begin{bmatrix} h_1(x(k)) \\ h_2(x(k)) \end{bmatrix} + [0 \quad I]\, x(k) + \begin{bmatrix} v_1(k) \\ v_2(k) \end{bmatrix} \quad (4)$$

which may be rewritten as $y(k) = h(x(k)) + v(k)$ where $h(.)$ is known but in general nonlinear. The measurement noise covariance is $R(k) = \text{block} - \text{diag}(R_1(k), R_2(k))$. The process and measurement equations in (3) and (4) are in the form required for approximate conditional mean estimation by (first order) extended Kalman filtering.

2 Cluttered Environment

We now extend the method to the case of cluttered (false) measurements arising from two separate sensors that are tracking a single target through a common surveillance region. Denote the set of measurements obtained by sensor i at time k by $Y_i(k) = \{y_1^i(k), y_2^i(k), ..., y_{m_k^i}^i(k)\}$. In heavily cluttered scenarios, validation gates can be applied to reduce the number of measurements for processing. The number of validated measurements per sensor per scan m_k^i is a random variable. In addition to the unknown sensor biases, there is uncertainty as to which measurement (if any) in $Y_i(k)$ corresponds to the target of interest.

Our approach to the problem is similar to that presented in [5] where a method was developed for the fusion of multiple measurements arising from a common target. The basis of the approach is probablistic data association (PDA) [4, pp.164]. This is a suboptimal state estimation scheme which approximates the Gaussian mixture density of the target state by a single Gaussian PDF at each processing stage.

The set of mutually exclusive association hypotheses for the procedure follows:

- $\theta_{00}(k)$ - no measurement in $Y_1(k)$ or $Y_2(k)$ is a target measurement;
- $\theta_{i0}(k)$ - measurement $y_i^1(k)$ in $Y_1(k)$ is a target measurement, all other measurements in $Y_1(k)$ and $Y_2(k)$ are clutter, $i = 1, ..., m_k^1$;
- $\theta_{0j}(k)$ - measurement $y_j^2(k)$ in $Y_2(k)$ is a target measurement, all other measurements in $Y_1(k)$ and $Y_2(k)$ are clutter, $j = 1, ..., m_k^2$;
- $\theta_{ij}(k)$ - measurement $y_i^1(k)$ in $Y_1(k)$ and $y_j^2(k)$ in $Y_2(k)$ are target measurements, all other measurements in $Y_1(k)$ and $Y_2(k)$ are clutter, $i = 1, ..., m_k^1$, $j = 1, ..., m_k^2$.

We therefore have a total of $m_k^1 m_k^2 + m_k^1 + m_k^2 + 1$ possible association hypotheses each of which has an *association probability* defined for $i = 0, 1, ..., m_k^1$ and $j = 0, 1, ..., m_k^2$ by

$$\beta_{ij}(k) = \Pr\{\theta_{ij}(k) | Y_1^k, Y_2^k\} \quad (5)$$

where Y_i^k denotes the cumulative data set for sensor i.

The joint registration and tracking process consists of propagating the approximate conditional mean of the combined target and sensor bias state $\hat{x}(k-1|k-1)$ and its covariance $P(k-1|k-1)$ to obtain $\hat{x}(k|k-1)$ and $P(k|k-1)$ from which

$$\begin{aligned} \hat{y}_i(k|k-1) &= H_i(k)\hat{x}(k|k-1) \\ S_i(k) &= H_i(k)P(k|k-1)H_i(k)^T + R_i \end{aligned}$$

where $H_i(k)$ is the Jacobian of h(.) evaluated at the state prediction $\hat{x}(k|k-1)$ and $S_i(k)$ is the covariance of the innovation process of sensor $i = 1, 2$. The validation gate for each sensor is then defined by an ellipsoid centred on the predicted measurement according to

$$G_i(k) \stackrel{\text{def}}{=} \{y_j^i(k) \in \mathbf{R}^N : \xi_j^i \leq \gamma_i\}, \quad (6)$$

where

$$\xi_j^i = (y_j^i(k) - \hat{y}^i(k|k-1))^T S_i^{-1}(k)(y_j^i(k) - \hat{y}^i(k|k-1)),$$

$j = 1, ..., m_k^i$, $i = 1, 2$, and N is the dimension of $y_j^i(k)$. The measurements are assumed Gaussian and so $\xi_j^i \sim \chi^2(N)$ with N degrees of freedom. The threshold γ_i is therefore choosen from a $\chi^2(N)$ probability distribution according to $\Pr\{\xi_j^i \leq \gamma_i\} \geq P_{Gi}$ where P_{Gi} is sufficiently high. Using the argument in [4, pp.157], and the condition $\sum_{i=0}^{m_k^1} \sum_{j=0}^{m_k^2} \beta_{ij}(k) = 1$, equation (5) takes on the form

$$\beta_{ij}(k) = \begin{cases} Ce_{ij}, & \text{for } i = 1, .., m_k^1, \text{ and } j = 1, .., m_k^2 \\ Cb_i, & \text{for } i = 1, .., m_k^1, \text{ and } j = 0 \\ Cc_j, & \text{for } i = 0, \text{ and } j = 1, .., m_k^2, \\ Ca, & \text{for } i = 0, \text{ and } j = 0 \end{cases} \quad (7)$$

where C is a normalizing constant and

$$
\begin{aligned}
a &= (1-P_{D1}P_{G1})\lambda_1(1-P_{D2}P_{G2})\lambda_2 \\
b_i &= \frac{P_{D1}(1-P_{D2}P_{G2})\lambda_2}{(2\pi)^{\frac{N}{2}}|S_1(k)|^{\frac{1}{2}}} e^{-\frac{1}{2}\nu_i^1(k)^T S_1^{-1}(k)\nu_i^1(k)} \\
c_j &= \frac{(1-P_{D1}P_{G1})\lambda_1 P_{D2}}{(2\pi)^{\frac{N}{2}}|S_2(k)|^{\frac{1}{2}}} e^{-\frac{1}{2}\nu_j^2(k)^T S_2^{-1}(k)\nu_j^2(k)} \\
e_{ij} &= \frac{P_{D1}P_{D2}}{(2\pi)^N|S(k)|^{\frac{1}{2}}} e^{-\frac{1}{2}\nu_{ij}(k)^T S^{-1}(k)\nu_{ij}(k)}
\end{aligned}
$$

The vector $\nu_i^1(k)$ is the innovation at sensor 1 based on measurement i, $\nu_j^2(k)$ is the innovation at sensor 2 based on measurement j and $\nu_{ij}(k) = [\nu_i^1(k) \quad \nu_j^2(k)]^T$ is the innovation from the two sensors for $i = 1, ..., m_k^1$, $j = 1, ..., m_k^2$. P_{Di} (P_{Gi}) is the detection (gate) probability of sensor i, λ_i is the spatial density of clutter measurements for sensor $i = 1, 2$, and the probability mass function of the number of clutter measurements is Poisson for each sensor. Having obtained $\beta_{ij}(k)$, it is now possible to evaluate the conditional state estimate and error covariance.

2.1 Conditional Mean and Error Covariance

The state update equation of the PDAF therefore takes on the form $\hat{x}(k|k) = \hat{x}(k|k-1) + W(k)\nu(k)$ where $W(k)$ is the filter gain and $\nu(k)$ is the combined innovation given by $\nu(k) = \sum_{i=0}^{m_k^1}\sum_{j=0}^{m_k^2}\beta_{ij}(k)\nu_{ij}(k)$ or

$$
\nu(k) = \begin{bmatrix} \sum_{i=1}^{m_k^1}\nu_i^1(k)\left[\beta_{i0}(k) + \sum_{j=1}^{m_k^2}\beta_{ij}(k)\right] \\ \sum_{j=1}^{m_k^2}\nu_j^2(k)\left[\beta_{0j}(k) + \sum_{i=1}^{m_k^1}\beta_{ij}(k)\right] \end{bmatrix}.
$$

The error covariance associated with the updated state estimate has the form

$$
\begin{aligned}
P(k|k) &= \beta_{00}(k)P(k|k-1) + \alpha_1(k)P_1(k|k) \\
&+ \alpha_2(k)P_2(k|k) + \alpha_{12}(k)P_{12}(k|k) + \tilde{P}(k) \\
P_i(k|k) &= P(k|k-1) - W_i(k)S_i(k)W_i^T(k), \\
P_{12}(k|k) &= P(k|k-1) - W(k)S(k)W^T(k)
\end{aligned}
$$

where $W(k) = [W_1(k) \quad W_2(k)]$ and $W_i(k)$ is the gain corresponding to measurements from sensor $i = 1, 2$. With probability $\beta_{00}(k)$, none of the measurements is correct and so the covariance $P(k|k-1)$ indicating no update, appears with this weighting. Similarly, with probability $\alpha_1(k) = \sum_{i=1}^{m_k^1}\beta_{i0}(k)$, target measurement is available only to sensor 1 and so the updated covariance $P_1(k|k)$ has this weighting. Furthermore, with probability $\alpha_2(k) = \sum_{j=1}^{m_k^2}\beta_{0j}(k)$ target measurement is available only to sensor 2 and so the updated covariance $P_2(k|k)$ has this weighting. Lastly with probability $\alpha_{12}(k) = (1 - \beta_{00}(k) - \alpha_1(k) - \alpha_2(k))$ target measurement is available to both sensor 1 and 2 and so the updated covariance $P_{12}(k|k)$ appears with this weighting. The last term $\tilde{P}$ is positive semidefinite [4, pp. 324], and represents the effect of the measurement origin uncertainty since we do not know which of the $m_k^1 m_2^2 + m_k^1 + m_k^2$ validated combinations actually represents the target measurement combination. The factors $\beta_{ij}(k)$ and $\tilde{P}(k)$ are measurement dependent and so $P(k|k)$ is a stochastic Riccati equation.

3 Numerical Results and Conclusions

Figure 1 shows a network of radars A and B each with fixed but unknown range and azimuth biases. For a target originating from location (250, 95), two distinct plots each displaced from the true trajectory are obtained. Tracking was done under Gaussian measurement noise and a clutter density that generates between 0 and 6 validated clutter samples per sensor during each stage. Figure 2 shows the composition of validated measurements. Figures 3 and 4 show bias estimates and their variances.

The registration and tracking algorithm is robust under clutter condition once track initiation has been properly done. It can track fixed or slowly varying registration errors under clutter conditions and is based on a sound mathematical foundation. Furthermore, it provides quality estimates for the solution set and can be adapted to cater for a wide range of system configurations.

References

[1] Dana, M.P., Registration: A prerequisite for multiple sensor tracking. In Bar-Shalom, Y., (Ed), *Multitarget-Multisensor Tracking: Advanced Applications*, Artech House, 1990.

[2] Fischer, W.L., Muehe, C.E., and Cameron, A.G., "Registration errors in a netted air surveillance system, "MIT Lincoln Laboratory Technical Note 1980-40, September 1980.

[3] Bar-Shalom, Y.(ed), *Multitarget-Multisensor Tracking: Applications and Advances. Vol. II*, Artech House, 1992.

[4] Bar-Shalom, Y. and Fortmann, T.E., *Tracking and Data Association*, Academic Press, 1988.

[5] Pulford, G.W., and Evans R.J., "State estimation in systems with multiple simultaneous measurements", *Proc. 33rd Conf. Decision and Control*, Orlando, FL, Vol. 4, pp.3299-3300, Dec. 1994.

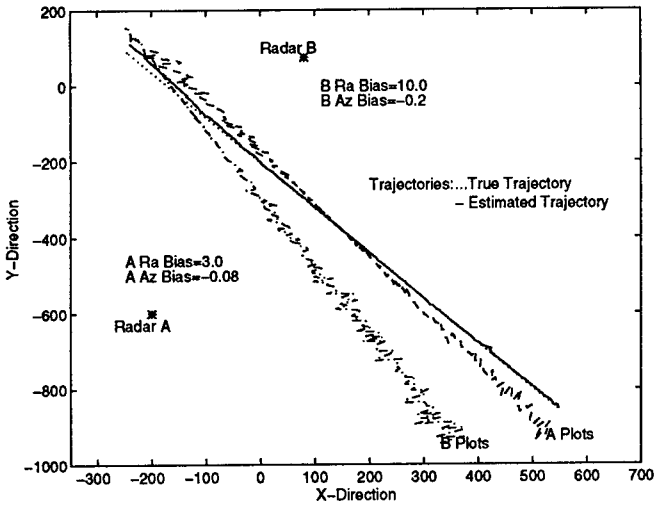


Figure 1. Trajectory estimate under clutter with $P_{Di} = 0.90$, $R_i = \text{diag}(4 \times 10^{-2}, 4 \times 10^{-4})$, $P_{Gi} = 0.989$ and $\lambda_i = 0.0002$, $i = 1, 2$.

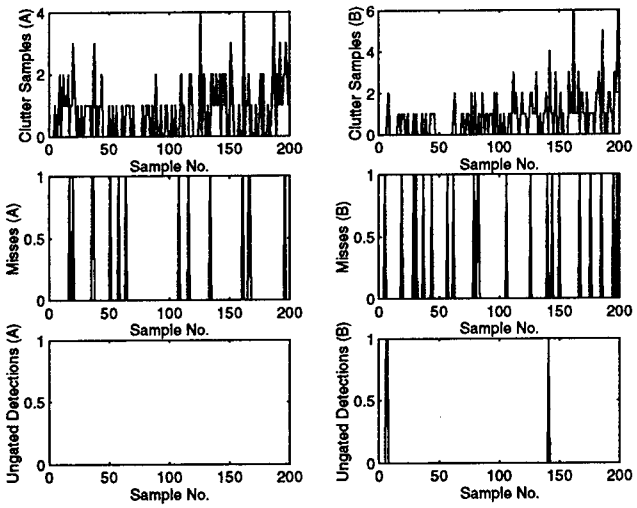


Figure 2. Variation of validated clutter, missed detections and ungated detections during the tracking process.

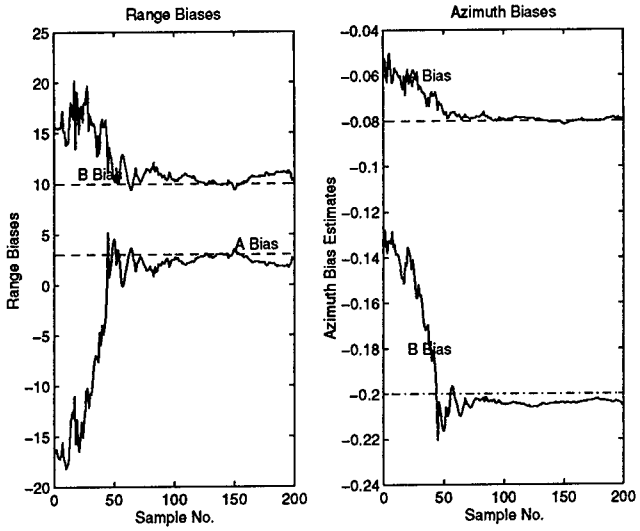


Figure 3. Estimates of radar biases during the tracking process.

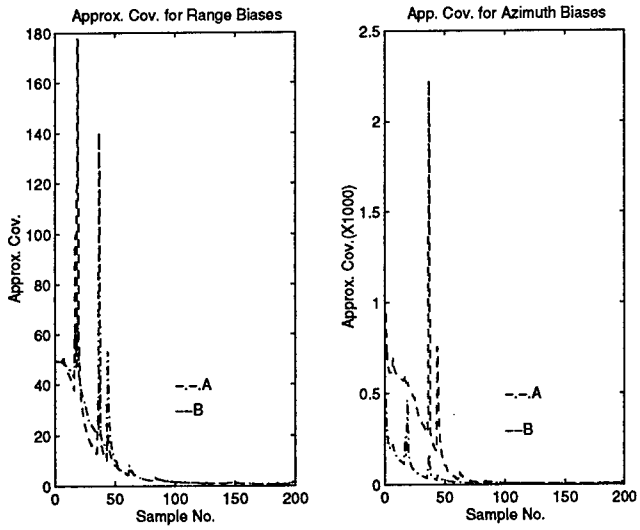


Figure 4. Variances of range and azimuth bias estimates during the tracking process.

MA-2: Array Processing I

Direction Finding Via Joint Diagonalization

A.J. Weiss

Dept. Elec. Eng. Systems,
Tel Aviv University,
Tel Aviv 69978, Israel.

B. Friedlander *

Dept. Elec.& Comp. Eng.,
University of California,
Davis, CA 95616.

Abstract

An approach to array processing (i.e., direction-finding, signal separation and reconstruction, and calibration) based on the Analytical Constant Modulus Algorithm is considered. The main advantage of this approach is that the multidimensional search associated with Maximum Likelihood based estimators or the single dimensional search associated with MUSIC type methods are eliminated.

The sensor array elements are assumed to have the same, up to a multiplicative constant, angle dependent, unknown gain pattern. We show that under this assumption it is possible to estimate the array response matrix and then use the result for direction finding, if the nominal array manifold is known, at least approximately. It is also possible to use the estimated array response matrix in order to separate and reconstruct the signals, or calibrate the array shape or response.

1 Introduction

In recent years many approaches to direction finding were proposed. All of these approaches are associated with some form of search. Maximum Likelihood based techniques like the EM algorithm [1], IQML [2], APM [3], MODE [4] and others, require multidimensional search in the parameter space. The main difficulty in using these approaches is that the algorithms tend to converge to a local stationary point and convergence to the global maximum (or minimum) is not guaranteed. Even MUSIC [5] type algorithms require a one-dimensional search which is indeed free from convergence problems but is associated with a lengthy search procedure (the search must be performed on a fine grid in order to avoid missing the narrow peaks of the MUSIC spectrum). Search-free techniques like Root-MUSIC and ESPRIT [9] require special array configurations that limit their applicability.

We consider an approach to steering vector estimation that does not rely on a search procedure. The steering vectors of the array are estimated via a short non-iterative algorithm. The estimates are close to the true steering vectors, if enough data samples are collected. If desired, the estimates can be further improved using a few iterations that are guaranteed to converge. These estimates of the steering vectors can be used for direction finding, signal separation and reconstruction, or array shape/phase calibration. The fact that the algorithm is essentially search free is its most appealing feature.

The method is based on a version of the Analytical Constant Modulus Algorithm that was recently proposed by van der Veen and Paulraj [6] for estimating constant modulus signals. However, the new approach is not limited to constant modulus signals or any other specific signals.

2 Problem Formulation

We begin by describing the data model for the observation of narrowband signals by an array of sensors.

We consider an M-element array of sensors and N narrowband signal sources, and define the $M \times 1$ vector $\mathbf{a}_n$ to be the complex array response for the nth source.

The outputs of the M array elements at the $k-$th sample are arranged in an $M \times 1$ vector,

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{u}(k) \qquad k = 1, 2 \cdots N_s; \qquad (1)$$

where $\mathbf{u}(k)$ is the noise vector, $\mathbf{s}(k)$ is the signal vector, and

$$\mathbf{A} \triangleq [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_N] \qquad (2)$$

Assuming that the signal vectors $\mathbf{s}(k)$ and the noise vectors $\mathbf{u}(k)$ are realizations of stationary, zero mean random processes, and there is no correlation between the noise and the signals, the data covariance matrix is

$$\mathbf{R} \triangleq \mathrm{E}\{\mathbf{x}(k)\mathbf{x}^H(k)\} = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \eta\mathbf{I} \qquad (3)$$

where $\mathbf{R}_s$ is the signal covariance matrix (a non-singular matrix) and $\eta\mathbf{I}$ is the noise covariance matrix.

Given N_s snapshots, the sample covariance is given by

$$\hat{\mathbf{R}} = \frac{1}{N_s}\sum_{k=1}^{N_s}\mathbf{x}(k)\mathbf{x}^H(k) \qquad (4)$$

We assume that the sensors have identical gain pattern, up to a multiplicative scalar factor. We therefore use the following model for $\mathbf{A}$.

$$[\mathbf{A}]_{m,n} = g_m e^{j\phi_{m,n}} \qquad (5)$$

The scalar g_m is the multiplicative factor of the mth sensor gain pattern, and the constants $\phi_{m,n}$ are the

*This work was supported by the Office of Naval Research under Contract No. N00014-95-1-0912

unknown sensor phase responses. We are interested in estimating the steering vector matrix, $\mathbf{A}$. As a by product, we also estimate the noise variance, η. Note that the observations, namely $\mathbf{x}(k)$, do not change if $\mathbf{A}$ is right multiplied by a diagonal matrix while $\mathbf{s}(k)$ is left multiplied by the inverse of the same diagonal matrix. This means that the steering vectors and the signals can be observed (and estimated) only up to a multiplicative complex scalar. We therefore assume, without loss of generality, that the first element of each steering vector is one.

Note that if the signals are uncorrelated then $\mathbf{R}_s$ is diagonal. In this case the scalars g_p can be estimated by observing that the elements on the diagonal of the data covariance are given by

$$\mathbf{R}_{m,m} = g_m^2 \sum_{n=1}^{N} [\mathbf{R}_s]_{n,n} + \eta \tag{6}$$

Under the assumption that $g_1 = 1$ we get

$$\sum_{n=1}^{N} [\mathbf{R}_s]_{n,n} \simeq \hat{\mathbf{R}}_{1,1} - \hat{\eta} \tag{7}$$

and

$$\hat{g}_m \simeq \sqrt{\frac{\hat{R}_{m,m} - \hat{\eta}}{\hat{R}_{1,1} - \hat{\eta}}} \tag{8}$$

where $\hat{}$ indicates estimated values.

This estimation procedure for g_m does not hold in the general case of correlated signals. We therefore resort to the assumption that g_m is given. This assumption is not restrictive in most direction finding applications. However, it might be somewhat restrictive in the signal estimation problem for which the sensor response is not of prime concern.

3 Steering Vector Estimation

The eigenvalue decomposition of the data covariance is given by

$$\mathbf{R} = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \eta\mathbf{I} = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H + \eta\mathbf{U}_n\mathbf{U}_n^H \tag{9}$$

where $\mathbf{\Lambda}_s = \mathrm{diag}\{\lambda_1, \cdots, \lambda_N\}$ is a diagonal matrix containing the N biggest eigenvalues in decreasing order, and the associated eigenvectors are the columns of the matrix $\mathbf{U}_s$. The columns of $\mathbf{U}_n$ are the remaining $M - N$ eigenvectors, associated with $\lambda_{N+1} = \cdots = \lambda_M = \eta$. Subtracting $\eta\mathbf{I}$ from the above equation we get

$$\mathbf{A}\mathbf{R}_s\mathbf{A}^H = \mathbf{U}_s\mathbf{\Gamma}_s\mathbf{U}_s^H \tag{10}$$

where $\mathbf{\Gamma}_s = \mathrm{diag}\{\lambda_1 - \eta, \cdots, \lambda_N - \eta\}$. Hence,

$$\mathbf{A} = \mathbf{U}_s\mathbf{W} \tag{11}$$

where $\mathbf{W}$ is a weighting matrix.

Based on an estimate of $\mathbf{U}_s$ we are interested in estimating the matrix $\mathbf{A}$, under the constraint that the modulus of each column is given by $[1, g_2 \cdots g_M]^T$. This raises the question of how many vectors of this form are contained in the subspace spanned by the columns of $\mathbf{A}$? We show in [8] that in some cases there are more than N such vectors in the range of A, and therefore the solution of (11) is not unique. However, in most cases, there are exactly N vectors with the given modulus in the column space of $\mathbf{A}$. The later is assumed in sequel.

The eigen decomposition of $\hat{\mathbf{R}}$ provides the estimate, $\hat{\mathbf{U}}_s$, of $\mathbf{U}_s$. Equation (11) indicates that minimizing the distance between its left hand side and the estimate of its right hand side corresponds to estimates of the steering vectors.

In order to find $\mathbf{W}$ we follow the steps of [6]. Any column vector, $\mathbf{w}$, in $\mathbf{W}$ must satisfy the equations,

$$\mathbf{w}^H\tilde{\mathbf{u}}_m\tilde{\mathbf{u}}_m^H\mathbf{w} = g_m^2, \qquad m = 1, 2, \cdots, M; \tag{12}$$

where $\tilde{\mathbf{u}}_m^H$ is the mth row vector of $\mathbf{U}_s$. These equations can be written in a different form as,

$$\mathbf{P}(\mathbf{w} \otimes \mathbf{w}^*) = \tilde{\mathbf{g}} \tag{13}$$

where the mth row of $\mathbf{P}$ is given by $\mathrm{vec}^T\{\tilde{\mathbf{u}}_m\tilde{\mathbf{u}}_m^H\}$ and

$$\tilde{\mathbf{g}} \triangleq [g_1^2, \cdots, g_M^2]^T \tag{14}$$

Define the Householder matrix

$$\mathbf{Q} \triangleq \mathbf{I} - 2\frac{\mathbf{q}\mathbf{q}^T}{\mathbf{q}^T\mathbf{q}}, \qquad \mathbf{q} \triangleq \tilde{\mathbf{g}} + \|\tilde{\mathbf{g}}\|\mathbf{e}_1 \tag{15}$$

where $\mathbf{e}_1$ is a vector of zeros except for the first element which is one. By left multiplying (13) by $\mathbf{Q}$ we get

$$\mathbf{Q}\mathbf{P}(\mathbf{w} \otimes \mathbf{w}^*) = -\|\tilde{\mathbf{g}}\|\mathbf{e}_1 \tag{16}$$

Define the $(M-1) \times N^2$ matrix $\hat{\mathbf{P}}$ to be $\mathbf{QP}$, with the first row deleted, so we get

$$\hat{\mathbf{P}}(\mathbf{w} \otimes \mathbf{w}^*) = 0 \tag{17}$$

This equation indicates that $\mathbf{w} \otimes \mathbf{w}^*$ belongs to the null space of $\hat{\mathbf{P}}$. Note that

$$\mathrm{rank}\{\hat{\mathbf{P}}\} \leq \min\{M - 1, N^2\} \leq M - 1 \tag{18}$$

Substituting in

$$\dim\{\mathrm{null}\{\hat{\mathbf{P}}\}\} = N^2 - \mathrm{rank}\{\hat{\mathbf{P}}\} \tag{19}$$

We get,

$$\dim\{\mathrm{null}\{\hat{\mathbf{P}}\}\} \geq N^2 - (M - 1) \tag{20}$$

In order that the null space $\mathrm{null}\{\hat{\mathbf{P}}\}$ will span the space of all the N vectors $\mathbf{w} \otimes \mathbf{w}^*$ we must have $\dim\{\mathrm{null}\{\hat{\mathbf{P}}\}\} = N$. Hence, we get the condition

$$N \geq N^2 - M + 1 \tag{21}$$

or

$$M \geq N^2 - N + 1 \tag{22}$$

If the condition (22) is met, then the solutions to (17) span the null space of $\hat{\mathbf{P}}$. Assume now that the null$\{\hat{\mathbf{P}}\}$ is spanned by the vectors $\mathbf{y}_1, \cdots, \mathbf{y}_N$. Each of these vectors can be obtained by a linear combination of the solutions of (17). Hence,

$$\mathbf{y}_n = \sum_{j=1}^{N} \alpha_{jn} (\mathbf{w}_j \otimes \mathbf{w}_j^*), \qquad n = 1, \cdots, N; \quad (23)$$

where α_{jn} are complex scalar coefficients. Performing the inverse vec operation on (23) we get

$$\text{vec}^{-1}\{\mathbf{y}_n\} = \sum_{j=1}^{N} \alpha_{jn} (\mathbf{w}_j^* \mathbf{w}_j^T) = \mathbf{W}^* \Lambda_n \mathbf{W}^T, \quad (24)$$

Hence, to obtain $\mathbf{W}$ we have to simultaneously diagonalize the matrices $\text{vec}^{-1}\{\mathbf{y}_n\}$, $n = 1, 2, \cdots, N$. An algorithms for performing this task can be found in [7]. Our approach is to simultaneously diagonalize only two matrices. This can be done by solving a generalized eigenvalue problem. Therefore we define,

$$\mathbf{Y}_1 \triangleq \sum_{n=1}^{N'} \text{vec}^{-1}\{\mathbf{y}_n\} = \mathbf{W}^* \tilde{\Lambda}_1 \mathbf{W}^T \quad (25)$$

and

$$\mathbf{Y}_2 \triangleq \sum_{n=N'+1}^{N} \text{vec}^{-1}\{\mathbf{y}_n\} = \mathbf{W}^* \tilde{\Lambda}_2 \mathbf{W}^T \quad (26)$$

where N' is the integer part of $N/2$. The eigen vectors, $\mathbf{v}_j$ that satisfy

$$\mathbf{Y}_1 \mathbf{v}_j = \mathbf{Y}_2 \mathbf{v}_j \beta_j \quad (27)$$

can be arranged in a matrix whose inverse is $\mathbf{W}^T$.

In order to obtain the final estimate of $\mathbf{A}$ we minimize the cost function

$$f(\hat{\mathbf{U}}_s, \mathbf{A}) \triangleq \|\hat{\mathbf{U}}_s \mathbf{W} - \mathbf{A}\|^2. \quad (28)$$

This can be done in two steps that can be repeated several times. Convergence is guaranteed since the cost function value decrease (or stay the same) in each step.

1) In this step we use the last estimate of $\mathbf{W}$ and find $\mathbf{A}$. Since the modulus of the elements of $\mathbf{A}$ are known we only have to find the phases of $\mathbf{A}$ which minimize the cost function. Obviously, the minimizing phase estimates are given by

$$\text{phase}\{\mathbf{A}_{ij}\} = \text{phase}\{[\hat{\mathbf{U}}_s \mathbf{W}]_{ij}\} \quad (29)$$

2) In this step we use the last estimate of $\mathbf{A}$ to estimate $\mathbf{W}$. The $\mathbf{W}$ that minimizes the cost function is given by

$$\mathbf{W} = \hat{\mathbf{U}}_s^H \mathbf{A} \quad (30)$$

Usually, between 3 to 10 iterations are needed.

4 Application to Direction Finding

Once the steering vector phases have been estimated, the signal directions of arrival (DOAs) can be easily extracted. If the array response is close to the free space model of propagation, then the phase of the mth element of the nth steering vector is given by

$$\phi_{m,n} = 2\pi(d_{x,m} \sin\theta_n \cos\psi_n + d_{y,m} \sin\theta_n \sin\psi_n + d_{z,m} \cos\theta_n) \quad (31)$$

where $d_{x,m}$, $d_{y,m}$, $d_{z,m}$ are the Cartesian coordinates (in wavelength units) of the mth sensor, while θ_n and ψ_n are the elevation angle (with respect to the z axis) and the azimuth (with respect to the x axis), respectively, of the nth source. Note that $\phi_{m,n}$ is known only modulo 2π. Hence, phase unwrapping must be used before applying the following method.

Define

$$\tilde{\mathbf{H}} \triangleq 2\pi \begin{bmatrix} d_{x,1} & d_{y,1} & d_{z,1} \\ d_{x,2} & d_{y,2} & d_{z,2} \\ \vdots & & \\ d_{x,M} & d_{y,M} & d_{z,M} \end{bmatrix} \quad (32)$$

$$\boldsymbol{\mu}_n = [\sin\theta_n \cos\psi_n, \sin\theta_n \sin\psi_n, \cos\theta_n]^T \quad (33)$$

$$\boldsymbol{\phi}_n = [\phi_{1,n}, \phi_{2,n} \cdots \phi_{M,n}]^T \quad (34)$$

which yield the following matrix equation

$$\tilde{\mathbf{H}} \boldsymbol{\mu}_n = \boldsymbol{\phi}_n \quad (35)$$

Hence we have

$$\hat{\boldsymbol{\mu}}_n = \tilde{\mathbf{H}}^\dagger \boldsymbol{\phi}_n \quad (36)$$

where $\tilde{\mathbf{H}}^\dagger$ is the left inverse of $\tilde{\mathbf{H}}$. The estimates of θ_n, ψ_n follow immediately.

If the array response is given by a calibration table rather than an analytic expression then the DOA is estimated by finding the calibration table entries that are close in some sense to the estimated steering vector. Interpolation is usually required to improve the system accuracy.

5 Numerical Examples

Consider a linear array of 3 sensors with element spacing of half a wavelength. The sensor gains are chosen arbitrarily to be 1.0, 1.02, 1.34. The array intercepts two equal power uncorrelated signals with Signal to Noise Ratio of 10dB. The direction of arrival (DOA) of one signal is 0° relative to broadside while the DOA of the other signal varies from 5° to 30°. The algorithm is applied to simulated data matrix with 500 snapshots. Figure 1 shows the experimental standard deviation, o, and experimental bias, *, of the steering vector phases, vs. the DOA separation. Each circle/asterisk is based on 200 experiments. The solid lines represents the Cramer-Rao bound which coincides with the theoretical performance analysis, in this case. It is apparent that the bias is negligible and the standard deviation agrees with the theoretical performance analysis and the bound.

Next, consider a linear array of 6 sensors with element spacing of half a wavelength. The sensor gains are chosen to be 1.0, 1.01, 1.35, 1.47, 1.12, 1.09. The array intercepts two correlated equal power signals with Signal to Noise Ratio of 10dB. The magnitude of the correlation coefficient is 0.95 and its phase varies from 0° to 180°. The direction of arrival (DOA) of one signal is 0° relative to broadside while the DOA of the other signal is 10°. The algorithm is applied to simulated data matrix with 500 snapshots. Figure 2 shows the experimental standard deviation, and experimental bias, of the steering vector phases vs. the correlation coefficient phase. Each circle/asterisk is based on 200 experiments. The solid line represents the theoretical performance analysis, and the dashed line represents the Cramer-Rao bound. We note that the experiments verify the theoretical performance analysis and that the bias is negligible. Observe that due to the correlation between the signals the statistical efficiency is lost.

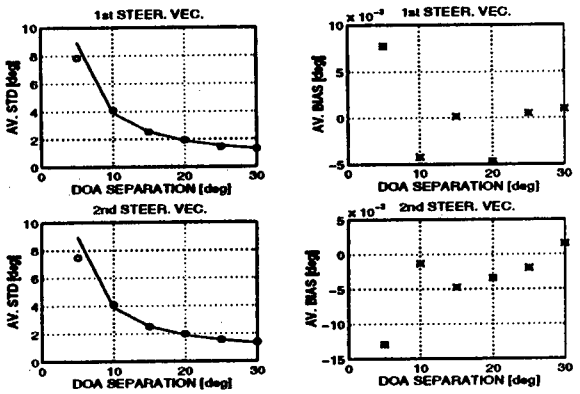


Figure 1: Standard Deviation and bias of steering vector phases vs. DOA separation.

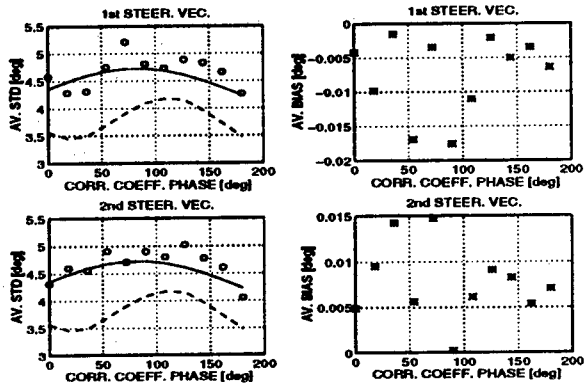


Figure 2: Standard Deviation and bias of steering vector phases vs. correlation phase.

6 Conclusions

We examined a new approach to direction-finding and signal estimation based on steering vector estimation. We showed that it is possible to estimate the array response matrix and then use the result for Direction Finding, if the nominal array manifold is known, at least approximately. It is also possible to use the array response matrix estimate in order to separate and reconstruct the signals or calibrate the array. The main advantage of the method is that the multidimensional search associated with Maximum Likelihood based estimators or the single dimensional search associated with MUSIC type methods is eliminated. The method can be applied in the presence of specular multipath (using spatial smoothing) but it is not suitable for signal separation in the presence of diffuse multipath.

References

[1] M. Feder, and E. Weinstein, "Parameter Estimation of Superimposed Signals Using the EM Algorithm," *IEEE Trans. on ASSP,* Vol. 36, pp. 477-489, Apr. 1988.

[2] Y. Bresler, and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise," *IEEE Trans. on ASSP,* Vol. 34, pp. 1081-1089, Oct. 1986.

[3] I. Ziskind, and M. Wax, "Maximum Likelihood Estimation via the Alternating Projections Maximization Algorithm," *IEEE Trans. on ASSP,* Vol. 36, pp. 1553-1560, Oct. 1988.

[4] P. Stoica, and K. Sharman, "Maximum Likelihood Methods for Direction-of-Arrival Estimation," *IEEE Trans. on ASSP,* Vol. 38, pp. 1132-1143, July 1990.

[5] R.O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *Proc. RADC Spectral Estimation Workshop,* pp. 243-258, Rome, NY, 1979.

[6] A.J. van der Veen, and A. Paulraj, "Analytical Solution to the Constant Modulus Factorization Problem," *Proceedings of ICASSP,* Detroit, MI, pp. 1433-1437, May 1995.

[7] A. Bunse-Gerstner, R. Byers, and V. Mehrmann, "Numerical Methods for Simultaneous Diagonalization," *SIAM J. Matrix Anal. Appl.,* Vol. 4, pp. 927-949, 1993.

[8] A.J. Weiss, and B. Friedlander, "Array Processing Using Joint Diagonalization," *European J. Signal Processing,* to appear.

[9] R. Roy, *et. al.*, "ESPRIT - A Subspace Rotation approach to Estimation of Parameters of Cisoids," *IEEE Trans. on ASSP,* Vol. 34, pp. 1340-2, 1986.

JOINT TARGET ANGLE AND DOPPLER ESTIMATION WITH FRACTIONAL LOWER-ORDER STATISTICS FOR AIRBORNE RADAR

Panagiotis Tsakalides and Chrysostomos L. Nikias

Signal & Image Processing Institute
Department of Electrical Engineering – Systems
University of Southern California
Los Angeles, CA 90089–2564
e-mail: tsakalid@sipi.usc.edu

ABSTRACT

We introduce a new joint spatial- and doppler-frequency high-resolution estimation technique based on the fractional lower-order statistics of the measurements of a radar array. We define the covariation matrix of the space-time radar observation vector process and employ subspace-based estimation techniques to the sample covariation matrix resulting in improved target angle and Doppler estimates in the presence of impulsive interference. We name the introduced technique "2-D Robust Covariation-Based MUSIC" or "2-D ROC-MUSIC". We show that 2-D ROC-MUSIC provides better angle/Doppler estimates than 2-D MUSIC in a wide range of impulsive interference environments and for very low signal-to-noise ratios.

1. INTRODUCTION

Most of the theoretical work in detection and estimation for radar applications has focused on the case where clutter is assumed to follow the Gaussian model. The Gaussian assumption is frequently motivated by the physics of the problem and it often leads to mathematically tractable solutions. However, in many practical instances, experimental results have been reported where clutter returns are impulsive in nature and cannot be appropriately modeled by means of the Gaussian distribution [1]. A number of distributions, based on empirical as well as theoretical grounds, have been proposed for the modeling of non-Gaussian clutter and interference environments [2, 3].

Recently, a statistical model for impulsive clutter has been proposed, which is based on the theory of symmetric alpha-stable ($S\alpha S$) random processes [4]. The model is of a statistical-physical nature and has been shown to arise under very general assumptions and to describe a broad class of impulsive interference. In particular, it has been shown in [4] that the first order distribution of the amplitude of the radar return follows a SαS law, while the first-order joint distribution of the quadrature components of the envelope of the radar return follows an isotropic stable law. In addition, the theory of *multivariate sub-Gaussian* random processes provides an elegant and mathematically tractable framework for the solution of the detection and parameter estimation problems in the presence of impulsive correlated radar clutter.

As mentioned in [5], much of the work reported for radar systems has concentrated on target detection in Gaussian or Non-Gaussian backgrounds [6, 7, 8, 9]. In this paper, we are addressing the parameter estimation problem with a space-time adaptive processing (STAP) radar operating in impulsive clutter and interference environments. We present a new subspace-based method for joint spatial- and doppler-frequency high-resolution estimation in the presence of impulsive noise which can be modeled as a complex symmetric alpha-stable ($S\alpha S$) process. In Section 2, we present some necessary preliminaries on α-stable processes. In Section 3, we formulate the STAP problem for airborne radar. In Section 4, we define the covariation matrix of the space-time radar sensor output snapshot and we show that eigendecomposition-based methods, such as the MUSIC algorithm, can be applied to the sample covariation matrix to extract the angle/Doppler information from the measurements. Finally, in Section 5, the improved performance of the proposed source localization method in the presence of a wide range of impulsive noise environments is demonstrated via Monte Carlo experiments.

2. MATHEMATICAL PRELIMINARIES

In this section, we introduce the statistical model that will be used to describe the additive noise. The model is based on the class of *isotropic* $S\alpha S$ distributions, and is well-suited for describing impulsive noise processes [4].

Stable processes satisfy the stability property which states that linear combinations of jointly stable variables are indeed stable. They arise as limiting processes of sums of independent, identically-distributed random variables via the generalized central limit theorem. They are described by their characteristic exponent α, taking values $0 < \alpha \leq 2$. Gaussian processes are stable processes with $\alpha = 2$. Stable distributions have heavier tails than the normal distribution, possess finite pth order moments only for $p < \alpha$, and

The work in this paper was supported by Rome Laboratory under Contract F30602-95-1-0001.

0-8186-7576-4/96 $5.00

are appropriate for modeling noise with outliers.

A complex random variable (r.v.) $X = X_1 + jX_2$ is isotropic $S\alpha S$ if X_1 and X_2 are jointly $S\alpha S$ and have a symmetric distribution. The characteristic function of X is given by

$$\varphi(\omega) = \mathcal{E}\{\exp(j\Re[\omega X^*])\} = \exp(-\gamma|\omega|^\alpha), \qquad (1)$$

where $\omega = \omega_1 + j\omega_2$. The *characteristic exponent* α is restricted to the values $0 < \alpha \leq 2$ and it determines the shape of the distribution. The smaller the characteristic exponent α, the heavier the tails of the density. The *dispersion* γ ($\gamma > 0$) plays a role analogous to the role that the variance plays for second-order processes. Namely, it determines the spread of the probability density function around the origin.

Several complex r.v.'s are jointly $S\alpha S$ if their real and imaginary parts are jointly $S\alpha S$. When X and Y are jointly $S\alpha S$ with $1 < \alpha \leq 2$, the *covariation* of X and Y is defined by

$$[X,Y]_\alpha = \frac{\mathcal{E}\{XY^{<p-1>}\}}{\mathcal{E}\{|Y|^p\}}\gamma_Y, \quad 1 \leq p < \alpha, \qquad (2)$$

where $\gamma_Y = [Y,Y]_\alpha$ is the dispersion of the r.v. Y, and we use throughout the convention $Y^{<p>} = |Y|^{p-1}Y^*$. Also, the *covariation coefficient* of X and Y is defined by

$$\lambda_{X,Y} = \frac{[X,Y]_\alpha}{[Y,Y]_\alpha}, \qquad (3)$$

and by using (2), it can be expressed as

$$\lambda_{X,Y} = \frac{E\{XY^{<p-1>}\}}{E\{|Y|^p\}}, \quad \text{for } 1 \leq p < \alpha. \qquad (4)$$

The covariation of complex jointly $S\alpha S$ r.v.'s is not generally symmetric and has the following properties:

P1 If X_1, X_2 and Y are jointly $S\alpha S$, then for any complex constants a and b,

$$[aX_1 + bX_2, Y]_\alpha = a[X_1,Y]_\alpha + b[X_2,Y]_\alpha;$$

P2 If Y_1 and Y_2 are independent and X_1, X_2 and Y are jointly $S\alpha S$, then for any complex constants a, b and c,

$$[aX_1, bY_1 + cY_2]_\alpha = ab^{<\alpha-1>}[X_1,Y_1]_\alpha + ac^{<\alpha-1>}[X_1,Y_2]_\alpha;$$

P3 If X and Y are independent $S\alpha S$, then $[X,Y]_\alpha = 0$.

3. STAP PROBLEM FORMULATION

Space-time adaptive processing (STAP) refers to multidimensional adaptive algorithms that simultaneously combine the signals from the elements of an array antenna and the multiple pulses of a coherent radar waveform, to suppress interference and provide target detection [10, 5, 11].

Consider a uniformly spaced linear array radar antenna consisting of N elements, which transmits a coherent burst of M pulses at a constant pulse repetition frequency (PRF) f_r and over a certain range of directions of interest. The array receives signals generated by q narrow-band moving targets which are located at azimuth angles $\{\theta_k;\ k = 1,\ldots,q\}$ and have relative velocities with respect to the radar $\{\nu_k;\ k = 1,\ldots,q\}$ corresponding to Doppler frequencies $\{f_k;\ k = 1,\ldots,q\}$. Since the signals are narrow-band, the propagation delay across the array is much smaller than the reciprocal of the signal bandwidth, and it follows that, by using a complex envelop representation, the array output can be expressed as [10]:

$$\mathbf{x}(t) = \mathbf{V}(\Theta,\varpi)\mathbf{s}(t) + \mathbf{n}(t), \qquad (5)$$

where

- $\mathbf{x}(t) = [x_1(t),\ldots,x_{MN}(t)]^T$ is the array output vector (N: number of array elements, M: number of pulses, t may refer to the number of the coherent processing intervals (CPI's) available at the receiver);
- $\mathbf{s}(t) = [s_1(t),\ldots,s_q(t)]^T$ is the signal vector emitted by the sources as received at the reference sensor 1 of the array;
- $\mathbf{V}(\Theta,\varpi) = [\mathbf{v}(\vartheta_1,\varpi_1),\ldots,\mathbf{v}(\vartheta_q,\varpi_q)]$ is the *space-time steering matrix* ($\varpi_k = \frac{f_k}{f_r}$);
- Space-Time steering vector: $\mathbf{v}(\vartheta_k,\varpi_k) = \mathbf{b}(\varpi_k) \otimes \mathbf{a}(\vartheta_k)$;
  - $\mathbf{a}(\vartheta_k) = [1, e^{j2\pi\vartheta_k},\ldots,e^{j(N-1)2\pi\vartheta_k}]^T$ is the spatial steering vector ($\vartheta_k = \frac{d}{\lambda_0}\cos(\theta_k)$);
  - $\mathbf{b}(\varpi_k) = [1, e^{j2\pi\varpi_k},\ldots,e^{j(M-1)2\pi\varpi_k}]^T$ is the temporal steering vector.
- $\mathbf{n}(t) = [n_1(t),\ldots,n_{MN}(t)]^T$ is the noise vector.

Assuming the availability of P coherent processing intervals (CPI's) $t_1,\ldots,t_P$, the data can be expressed as

$$\mathbf{X} = \mathbf{V}(\Theta,\varpi)\mathbf{S} + \mathbf{N}, \qquad (6)$$

where $\mathbf{X}$ and $\mathbf{N}$ are the $MN \times P$ matrices

$$\mathbf{X} = [\mathbf{x}(t_1),\ldots,\mathbf{x}(t_P)], \qquad (7)$$

$$\mathbf{N} = [\mathbf{n}(t_1),\ldots,\mathbf{n}(t_P)], \qquad (8)$$

and $\mathbf{S}$ is the $q \times P$ matrix

$$\mathbf{S} = [\mathbf{s}(t_1),\ldots,\mathbf{s}(t_P)]. \qquad (9)$$

Our objective is to jointly estimate the directions-of-arrival $\{\theta_k;\ k = 1,\ldots,q\}$ and the Doppler frequencies $\{f_k;\ k = 1,\ldots,q\}$ of the source targets.

4. THE ARRAY COVARIATION MATRIX

We will assume that the q signal waveforms are non-coherent, statistically independent, complex isotropic $S\alpha S$ ($1 < \alpha \leq 2$) random processes with zero location parameter and covariation matrix $\mathbf{\Gamma}_S = \mathrm{diag}(\gamma_{s_1},\ldots,\gamma_{s_q})$. Also, the noise vector $\mathbf{n}(t)$ is a complex isotropic $S\alpha S$ random process with the same characteristic exponent α as the signals. The noise is assumed to be independent of the signals with covariation matrix $\mathbf{\Gamma}_N = \gamma_n\mathbf{I}$.

Now, we define the *covariation matrix*, $\mathbf{\Gamma}_X$, of the observation vector process $\mathbf{x}(t)$ as the matrix whose elements are the covariations $[x_i(t), x_j(t)]_\alpha$ of the components of $\mathbf{x}(t)$. By using properties P1-P3, we obtain the following expression for the covariation of the sensor measurements:

$$[x_i(t), x_j(t)]_\alpha = \sum_{k=1}^{q} v_i(\vartheta_k, \varpi_k) v_j^{<\alpha-1>}(\vartheta_k, \varpi_k)\gamma_{s_k} + \gamma_n \delta_{i,j} \quad i,j = 1, \ldots, MN. \tag{10}$$

In matrix form, (10) gives the following expression for the covariation matrix of the observation vector:

$$\mathbf{\Gamma}_X \triangleq [\mathbf{x}(t), \mathbf{x}(t)]_\alpha = \mathbf{V}(\Theta, \varpi)\mathbf{\Gamma}_S \mathbf{V}^{<\alpha-1>}(\Theta, \varpi) + \gamma_n \mathbf{I}, \tag{11}$$

where the (i,j)th element of matrix $\mathbf{V}^{<\alpha-1>}(\Theta, \varpi)$ results from the (j,i)th element of $\mathbf{V}(\Theta, \varpi)$ according to the operation

$$[\mathbf{V}^{<\alpha-1>}(\Theta, \varpi)]_{i,j} = [\mathbf{V}(\Theta, \varpi)]_{j,i}^{<\alpha-1>} \tag{12}$$

Clearly, when $\alpha = 2$, i.e., for Gaussian distributed signals and noise, the expression for the covariation matrix is identical to the well-known expression for the covariance matrix:

$$\mathbf{R}_X = \mathbf{V}(\Theta, \varpi)\mathbf{\Sigma}\mathbf{V}^H(\Theta, \varpi) + \sigma^2 \mathbf{I}, \tag{13}$$

where $\mathbf{\Sigma}$ is the signal covariance matrix.

When the amplitude response of the sensors equals unity, it follows that

$$[\mathbf{V}^{<\alpha-1>}(\Theta, \varpi)]_{i,j} = [\mathbf{V}(\Theta, \varpi)]_{j,i}^{*}, \tag{14}$$

and thus the covariation matrix can be written as

$$\mathbf{\Gamma}_X = \mathbf{V}(\Theta, \varpi)\mathbf{\Gamma}_S \mathbf{V}^H(\Theta, \varpi) + \gamma_n \mathbf{I}. \tag{15}$$

Observing (15), we conclude that standard subspace techniques can be applied to the covariation or the covariation coefficient matrices of the observation vector to extract the angle/Doppler information. In practice, we have to estimate the covariation matrix from a finite number of array sensor measurements. A proposed estimator for the covariation coefficient $\lambda_{x_i(t),x_j(t)}$ is called the *fractional lower order (FLOM) estimator* and is given by [12, 13]

$$\hat{\lambda}_{x_i(t),x_j(t)} = \frac{\sum_{t=1}^{n} x_i(t)\hat{x}_j^{<p-1>}(t)}{\sum_{t=1}^{n} |x_j(t)|^p} \tag{16}$$

for some $0 \leq p < \alpha/2$. We will refer to the new algorithm resulting from the eigendecomposition of the array covariation coefficient matrix as the **2-D Robust Covariation-Based MUSIC** or **2-D ROC-MUSIC**.

5. EXPERIMENTAL RESULTS

In this section, we show results on the resolution capability and estimation accuracy of the 2-D ROC-MUSIC and 2-D MUSIC methods. The array is linear with five sensors spaced a half-wavelength apart ($N = 5$). The number of transmitted pulses is $M = 10$. Three moving targets impinge on the array from directions $\mathbf{\Theta} = [-20°, -40°, 40°]$ and they have Doppler values $\mathbf{D} = [-0.3, -0.2, 0.3]$). The

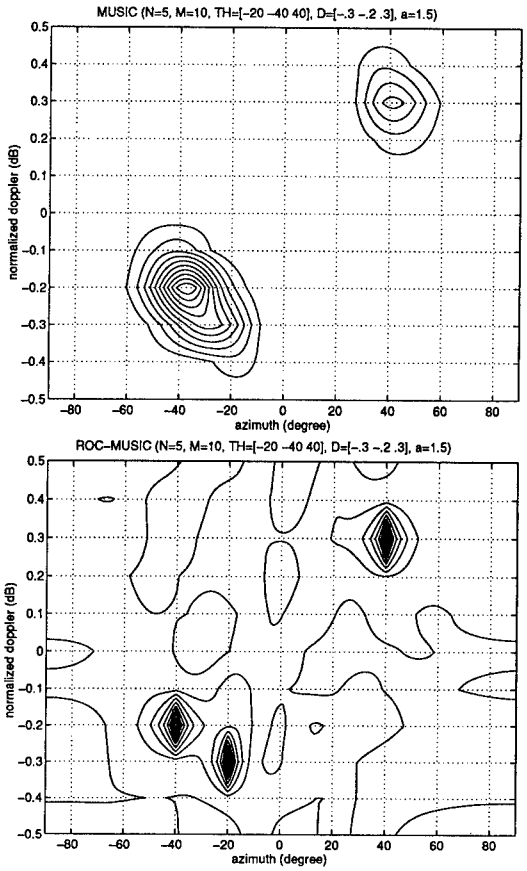


Figure 1: 2-D MUSIC and 2-D ROC-MUSIC angle-Doppler spectra ($N = 5$, $M = 10$, $\mathbf{\Theta} = [-20°, -40°, 40°]$ $\mathbf{D} = [-0.3, -0.2, 0.3]$). Additive stable noise ($\alpha = 1.5$, $\gamma_n = 4$).

number of snapshots available to the algorithms is $P = 1000$. The noise follows the bivariate isotropic stable distribution with $\alpha = 1.5$.

Since the alpha-stable family for $\alpha < 2$ determines processes with infinite variance, we define an alternative signal-to-noise ratio. Namely, we define the *Generalized SNR (GSNR)* to be the ratio of the signal power over the noise dispersion γ_n:

$$GSNR = 10\log\left(\frac{1}{\gamma_n M}\sum_{t=1}^{M}|s(t)|^2\right). \tag{17}$$

The GSNR is 22.3 dB ($\gamma_n = 1$). The characteristic exponent α of the additive noise is unknown to the ROC-MUSIC algorithm. The parameter p in the estimation of the covariation matrix (cf. (16)): was set equal to $p = 0.8$. Clearly, MUSIC can be thought as a special case of ROC-MUSIC with $p = 2$.

In Figure 1, isosurfaces of space-time spectral estimates are shown for the 2-D ROC-MUSIC and the 2-D MUSIC algorithms. We can see that the 2-D MUSIC method exhibits poor resolution performance and it does not resolve the two closely-spaced moving targets. On the other hand, the 2-D

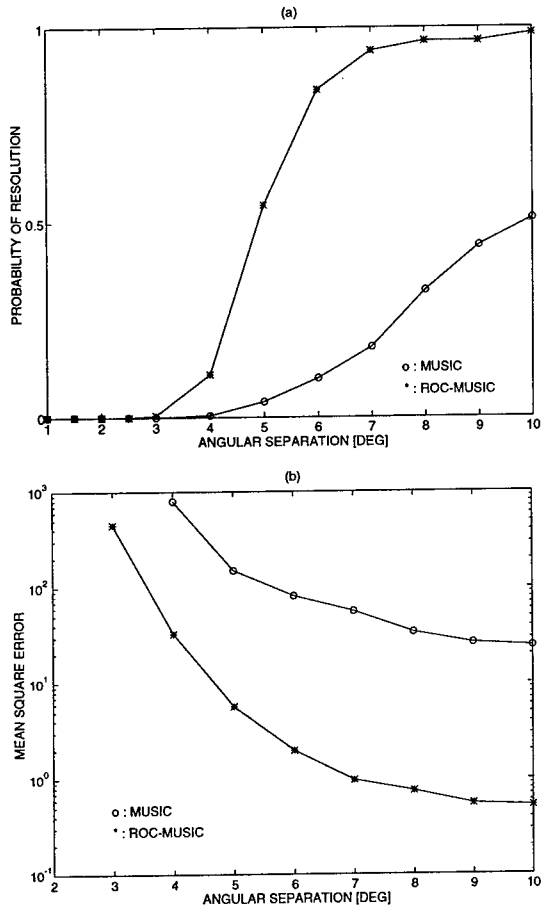


Figure 2: Probability of resolution (a) and mean square error (b) as functions of the source angular separation, $\alpha = 1.5$.

ROC-MUSIC method exhibits high-resolution capabilities for non-Gaussian additive noise environments.

Figure 2 illustrates the variation of the algorithmic performance with respect to the spatial angle separation of the two closely spaced incoming targets for GSNR= 22.3 dB, ($\alpha = 1.5$). As expected, the resolution capability of both algorithms improves with increased angle separation between the two sources. But for a given probability of resolution, the 2-D ROC-MUSIC algorithm requires a lower angle separation threshold than the 2-D MUSIC algorithm.

6. CONCLUSIONS

We considered the problem of target angle and Doppler estimation with an airborne radar employing space-time adaptive processing. We introduced a new joint spatial- and doppler-frequency high-resolution estimation technique based on the fractional lower-order statistics of the measurements of a radar array. We showed that the proposed 2-D ROC-MUSIC algorithm provides better angle/Doppler estimates than the 2-D MUSIC method, and it can result to improved STAP radar systems operating in impulsive interference environments.

7. REFERENCES

[1] I. S. Reed, C. L. Nikias, and V. Prasanna, "Multidisciplinary research on advanced high-speed, adaptive signal processing for radar sensors," Annual Tech. Rep., University of Southern California, Jan. 1996.

[2] S. A. Kassam, *Signal Detection in Non-Gaussian Noise.* Berlin: Springer-Verlang, 1988.

[3] D. Middleton, "Threshold detection in non-Gaussian interference environments: Exposition and interpretation of new results for EMC applications," *IEEE Trans. Electrom. Compat.*, vol. 26, pp. 19–28, Feb. 1984.

[4] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications.* New York: John Wiley and Sons, 1995.

[5] J. Ward, "Cramér-Rao bounds for target angle and Doppler estimation with space-time adaptive processing radar," in *Twenty-Ninth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA), October 30-November 1 1995.

[6] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, pp. 115–127, March 1986.

[7] M. Rangaswamy, D. Weiner, and A. Ozturk, "Non-Gaussian random vector identification using spherically invariant random processes," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, pp. 111–123, Jan. 1993.

[8] K. J. Sangrton and K. R. Gerlach, "Coherent detection of radar targets in a non-Gaussian background," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, pp. 330–340, April 1994.

[9] G. A. Tsihrintzis and C. L. Nikias, "Data adaptive algorithms for signal detection in sub-gaussian impulsive interference," *IEEE Trans. Signal Processing.* submitted on January, 1996, pp. 25.

[10] J. Ward, "Space-time adaptive processing for airborne radar," Tech. Rep. 1015, Lincoln Laboratory, Dec. 1994.

[11] S. U. Pillai, W. C. Lee, and J. Guerci, "Multichannel space-time adaptive processing," in *Twenty-Ninth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA), Oct. 30-Nov. 1 1995.

[12] P. Tsakalides, *Array Signal Processing with Alpha-Stable Distributions.* PhD thesis, University of Southern California, Los Angeles, California, December 1995.

[13] P. Tsakalides and C. L. Nikias, "The robust covariation-based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive noise environments," *IEEE Trans. Signal Processing*, July 1996. Also available as technical report USC-SIPI-278.

Convex Optimization for Antenna Array Processing

Hervé Lebret
ENSTA
32, bd Victor
75739 PARIS CEDEX 15, FRANCE
lebret@ensta.fr

Abstract

Antenna array pattern synthesis deals with choosing the complex weights of an antenna array in order to satisfy a set of specifications or to say if such a set is feasible. It appears that these problems can often be expressed as convex optimization problems which can be solved numerically with algorithms such as interior point methods. Two examples are given dealing with constrained adaptive array processing and robustness issues.

1 Introduction

It is well known that the antenna pattern of a linear array in direction θ is given by the amplitude of

$$G(\theta) = \sum_{i=1}^{N} w_i e^{j\frac{2\pi}{\lambda}x_i \cos\theta} \tag{1}$$

where the complex weights w_i are the parameters. The position of the elements are given by x_i whereas λ is the wavelength. The pattern is easily generalized to any array geometry with two angular variables (azimuth and elevation).

What may be not as well known is that the array pattern is a convex function of the real and imaginary parts of the weights. This important property makes possible the solution of many antenna array synthesis problems using convex optimization and more particularly recently developed algorithms (interior point methods).

This is all the more interesting as many other problems arising in array processing are convex. For instance, the noise or signal power with general form $w^T Rw$ where R is some covariance matrix and w^T denotes the conjugate of w are convex functions. The weight level $||w||$, defined as a given norm of the weights vector is another convex function. More generally, convex quadratic functions appear often in antenna array design and we see in the next section how they can be solved.

2 Convex optimization

A convex optimization problem can be defined as the minimization of a convex function over a convex set. The important property is that for a convex problem, any local optimum is in fact global. Furthermore, by using optimality conditions or more generally the theory of duality, it is possible to obtain lower bounds on the optimal value and an absolute required precision on the desired results.

It is impossible here to describe more precisely the properties of convex optimization. Let us simply mention the recent book by Hiriart-Urruty and Lemaréchal [3]. Even more interesting is the development of very efficient algorithms called interior point methods. The book by Nesterov and Nemirovsky [7] is the most complete account on the subject.

Finally the article of S. Boyd and L. Vandenberghe [9] shows that convex optimization is of much interest in many engineering fields. Let us show now how to express an antenna array pattern synthesis problem as a convex optimization one.

3 Pattern synthesis as a convex program

In general, it is possible to design optimal antenna array patterns by solving particular convex optimization problems of the general form

$$\begin{array}{ll} \text{minimize} & e^T x \\ \text{subject to} & ||A_i x + b_i||^2 \leq c_i^T x + d_i, i = 1, \ldots, L, \end{array} \tag{2}$$

where $A_i \in \mathbf{R}^{m\times n}$, $b_i \in \mathbf{R}^n$, $c_i, e, x \in \mathbf{R}^n$ and $d_i \in \mathbf{R}$. These are called quadratically constrained convex quadratic programs (QCQP). They can be solved with an algorithm described in appendix A. Let us notice that if a given

0-8186-7576-4/96 $5.00

objective $f(x)$ is given, it is easy to replace its minimization with the following problem

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & f(x) \leq t \end{array} \quad (3)$$

so that the choice of a linear objective is general.

Let us now express the array patern as a quadratic function in order to recover the general form of QCQP problem (2). Expression (1) is a linear complex function of the weights so that its amplitude squared is a quadratic function of the real and imaginary parts of the complex weights. Generally a normalization constraint $G(\theta_0) = 1$ is used so that it is possible to eliminate one of the weights as

$$w_N = e^{-j\frac{2\pi}{\lambda}x_N \cos\theta_0}\left(1 - \sum_{i=1}^{N-1} w_i e^{j\frac{2\pi}{\lambda}x_i \cos\theta_0}\right)$$

Combining the quadratic expression of the beam pattern and the elimination leads to the general expression $||A_i x + b_i||^2$ for a given $|G(\theta_i)|^2$ where x includes the real and imaginary parts of the first (N-1) weights. In this case, $m = 2$ and $n = 2(N-1)$. Therefore if we want to constrain the beam pattern in L different directions we have to choose $c_i = 0$ and the constrained level d_i for $i = 1$ to L.

A similar expression can be derived for the positive power $w^T Rw$ so that any constrained power can be included in problem (2) with correspondent $c = 0$ and d giving the power constraint. A difference is that here $m = N$. The objective to be minimized is also in general one of the previous quadratic expressions $||A_0 x + b_0||^2$, so that we can replace it with the minimization of $t = e^T x$ with the constraint $||A_0 x + b_0||^2 \leq e^T x$. In this case $c_0 = e$ and $d_0 = 0$ using the formulation (3). This also implies that x is modified in order to add the new variable t, so that $n = 2N - 1$. Finally we could also add constraints on the weights norm which can be very interesting for the signal to noise ratio.

We want to show now through some simulation examples, the interesting applications of interior point methods to antenna array processing.

4 Simulation Examples

Applications of convex optimization to antenna array processing are numerous [6, 5]. We want to show here two kinds of applications: creating broad gaps in the pattern for intefe-rence rejection with applications to adaptive beamforming and designing antenna patterns with robustness properties.

4.1 Adaptive arrays

In adaptive arrays problems, it is generally desired to put zeros in directions corresponding to interferences. Neverthess, it is sometimes more efficient to create a broad angular zone where the pattern is minimum even if not zero.

4.1.1 Constrained pattern

As an example, we deal with the minimization of the pattern level around 70° while keeping a 0dB level at 90° for a 32-element linear regular array. The element distance is half a wavelength. The minimized area is 15° wide around 70° and we also want the pattern level to remain less than -12dB in the sidelobe area (except of course the minimized region).

The problem is discretized in the angular directions in order to be expressed with the general form (2). The figure 1 presents the optimal result. Because the problem is **convex**, it is possible to state that within the required precision (which is here 10^{-6}), it is impossible to find weights giving a better rejection level with the given constraints. It is also possible to compare the results with a given adaptive technique.

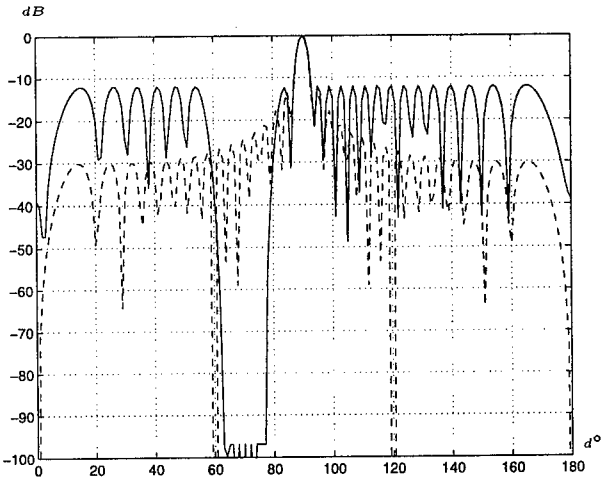


Figure 1. Optimized pattern for interference rejection around 70 degrees. The straight line is the solution of the convex problem, whereas the dashed line gives the standard beamformer.

4.1.2 Constrained adaptive beamforming

We can now have a slightly different approach. Constrained adaptive beamforming is an important issue as recent articles show it [2, 8]. The previous simulation could be criticized as the interference position needs to be known. Let us assume therefore that the region previously mentionned corresponds to clutter where the beam pattern level has to be less than -40dB. We can use the signal covariance matrix R as in standard array processing. For the simulation, we assume that this matrix is built with

- a signal of interest in direction 90° with level 0dB,
- four interferences at position $20, 45, 50, 70°$, with identical level (60dB)

- a white noise density (-60dB).

Figure 2 gives the result of standard adaptive beamforming with the same array as above, that is the minimization of $w^T Rw$ subject to $G(90°) = 1$. The four interferences are eliminated as expected. The new problem becomes the minimization of the signal power $w^T Rw$ with constraints on the clutter zone (less than -40dB), the mainlobe zone (less than 0.08 dB) and the sidelobe zone (less than -12dB) and with a normalization constraint of 0dB at 90°. Figure 3 gives the beam pattern for the constrained adaptive beamforming problem. The interferences are once again cancelled, furthermore the constraints on the clutter zone are achieved.

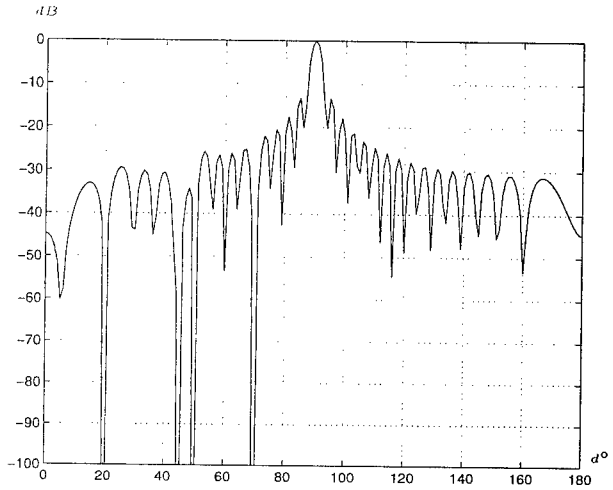


Figure 2. Interference rejection through standard adaptive beamforming.

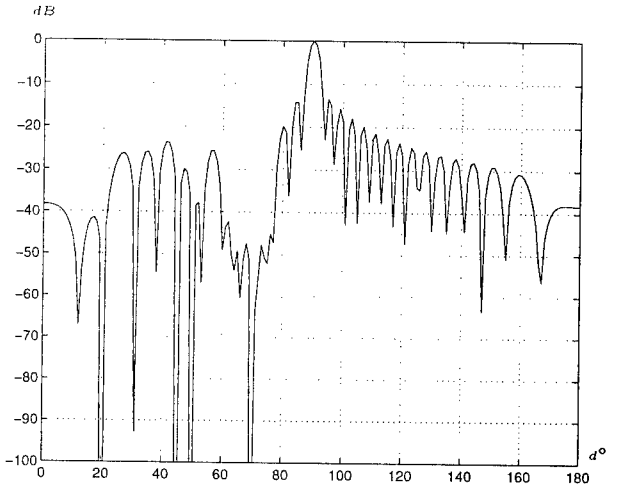


Figure 3. Interference rejection through constrained adaptive beamfroming.

4.2 Robustness issues

The problem of robustness is particularly important for antenna array design. We will very quickly show here some results. More details can be found in [4]. The main ideas come here from a series of papers by Evans [10] and Cantoni [1] . Here we are just interested in the robustness of the weights themselves. More precisely, it is known that the optimal weights have to be discretized for implementation. What is the influence of the quantization on the optimal results?

These problems can also be expressed as convex optimization problems and figure 4 shows such an example. The figure shows the optimal sidelobe level obtained with quantization steps Δw smaller than 6.10^{-3}. This means that the difference between any quantized weight and the corresponding optimal weight is less than Δw in modulus. The straight line corresponds to a 10-element array with a mainlobe width of 25° whereas the dashed line is for a 30-element array with a mainlobe width of 25°. For both cases, the mainlobe direction is 45°. The optimal sidelobe level is obviously increasing with the quantization step.

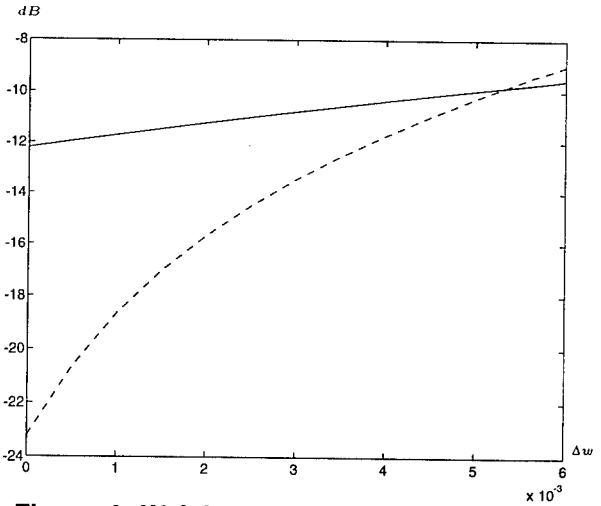


Figure 4. Weight robustness: optimal sidelobe level vs. quantization step

5 Conclusion

Through two examples, we tried to show the advantages of convex optimization. Of course all optimization problems are not convex, but it is of much interest to recognize convex ones and use their properties. From our point of view, the main one is to be able to state if a problem can be solved or not, and if it can be, to say with an absolute precision what the optimal result is. Another important question is the

real-time capabilities of such algorithms which are already very efficient. The advance of digital computers will give answers although the problem remains opened.

A An interior point algorithm

The algorithm used to solve problem (2) minimizes the function

$$\begin{aligned}\phi(x,l) &= q\log(e^T x - l) \\ &+ \sum_{i=1}^{L} \log\left((c_i^T x + d_i) - \|A_i x + b_i\|^2\right) \quad (4)\end{aligned}$$

where l is a lower bound on the objective, through

$x_1 \leftarrow$ initial feasible point ;
$l_1 \leftarrow$ initial lower bound;
$k \leftarrow 0$;
repeat {
  $k+1 \leftarrow k$;
  minimize $\phi(x_k)$ through:
    $y_1 \leftarrow x_k$; $i \leftarrow 0$;
    repeat {
      $i+1 \leftarrow i$;
      compute v_i, Newton direction of $\phi(y_i)$
      then $y_{i+1} \leftarrow y_i + \alpha_i v_i$ with:
      $\alpha_i = \text{argmin}_\alpha \phi(y_i + \alpha v_i)$;
    } **until convergence**;
  $x_{k+1} \leftarrow y_{i+1}$;
  compute a new lower bound l_{k+1};
} **until** $e^T x_{k+1} - l_{k+1} <$ tol;

Let us add that α_i is computed through a line search and the updated lower bound l_k is obtained thanks to optimality conditions of the minimized function ϕ.

References

[1] A. Cantoni, X. G. Lin, and K. L. Teo. A new approach to the optimization of robust antenna array processors. *IEEE Trans. Antennas Propag.*, 41(4):403–411, Apr. 1993.

[2] L. B. Fertig and J. H. McClellan. Dual forms for constrained adaptive filtering. *IEEE Trans. Signal Processing*, 42(1):11–23, Jan. 1994.

[3] J. B. Hiriart-Uruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms*. Springer-Verlag, 1993.

[4] H. Lebret. *Synthèse de diagrammes de réseaux d'antennes par optimisation convexe*. PhD thesis, Université de Rennes I, Nov. 1994.

[5] H. Lebret. Antenna pattern synthesis through convex optimization. In F. T. Luk, editor, *Advanced Signal Processing Algorithms*, Proc. SPIE 2563, pages 182–192, 1995.

[6] H. Lebret and S. P. Boyd. Antenna array pattern synthesis via convex optimization. Submitted to *IEEE Trans. on Signal Processing*, September 1995.

[7] Y. E. Nesterov and A. S. Nemirovsky. *Interior point polynomial methods in convex programming*. SIAM, 1993.

[8] F. Qian and B. D. V. Veen. Quadratically constrained adaptive beamforming for coherent signals and interference. *IEEE Trans. Signal Processing*, 43(8):1890–1900, Aug. 1995.

[9] L. Vandenberghe and S. Boyd. Semidefinite programming. To be published in *SIAM Review*, July 1994.

[10] I. Webster, R. Evans, and A. Cantoni. Robust perturbation algorithms for adaptive antenna arrays. *IEEE Trans. Antennas Propag.*, 38(2):195–201, Feb. 1990.

Instrumental Variable Subspace Tracking with Applications to Sensor Array Processing and Frequency Estimation*

TONY GUSTAFSSON AND MATS VIBERG

Chalmers University of Technology
Department of Applied Electronics
S-412 96 Gothenburg, Sweden
email: tonyg@ae.chalmers.se

Abstract

Recursive methods for subspace tracking with applications to 'on-line' direction of arrival estimation, have lately drawn considerable interest. In this paper, Instrumental Variable (IV) generalizations of the Projection Approximation Subspace Tracking (PAST) algorithm are proposed. The IV-approach is motivated by the fact that PAST delivers biased estimates when the noise vectors are not spatially white. The resulting basic IV-algorithm has a computational complexity of $3mn + \mathcal{O}(n^2)$ *complex multiplications, where* m *is the dimension of the measurement vector and* n *is the subspace dimension. The performance of the proposed algorithms in tracking sinusoids in colored noise is illustrated by computer simulations.*

1 Introduction

One aspect of the sensor array signal processing field that has drawn much attention is the application of high-resolution frequency and direction of arrival (DOA) estimation techniques to non-stationary environments, see for example [1]. A drawback of traditional subspace methods, in this scenario, is that the singular value decomposition (SVD) is time consuming to update. A specific example of a successful subspace tracking algorithm is the Projection Approximation Subspace Tracking (PAST) algorithm [5]. The basic idea of PAST is that a projection like unconstrained criterion is approximated, which leads to a RLS-like algorithm for tracking the signal subspace. The DOA (or frequency) estimates can then be taken as the angles of the eigenvalues of a matrix obtained using the shift-invariant structure of the subspace (Uniform Linear Array, ULA). However, PAST assumes that the noise is spatially white, and tends to deliver biased estimates whenever this requirement is not fulfilled. This fact is the motivation of the algorithms proposed herein. The aim of this paper is to present Instrumental Variable (IV) generalizations of PAST. For a treatment of IV methods in the context of identifying linear systems, see [3]. Like all other IV-methods we require that an IV-vector, that is uncorrelated with the noise vector, can be found. As long as this requirement is fulfilled, the noise vectors can be allowed to have arbitrary (temporal and spatial) color. A certain rank condition must also be fulfilled. One possible approach to find the instruments is to consider an array that is divided into sub-arrays. Then the outputs of one of the sub-arrays can be taken as instruments. Then, if the sub-arrays are sufficiently far apart, the noise in the main sub-array is uncorrelated with the IV vector. For a discussion on temporal IV's, see [4]. In the following, $\mathcal{R}(\mathbf{A})$ denotes the subspace spanned by the columns of $\mathbf{A}$ and $\rho(\mathbf{A})$ denotes the rank of $\mathbf{A}$.

2 Problem Formulation

Let $\mathbf{z}(t) \in \mathcal{C}^{m\times 1}$ be the observed data vector. In the array case, $\mathbf{z}(t)$ consists of the samples of an array with m sensors. In time series (sum of complex sinusoids) problems, $\mathbf{z}(t) = [z(t), \ldots, z(t+m-1)]^T$ consists of m consecutive samples of an observed scalar signal. It is assumed that $\mathbf{z}(t)$ consists of n narrow-band plane waves impinging on an antenna array or n complex sinusoids corrupted by additive noise. Here the subspace dimension n, $n < m$, is assumed to be known. Hence, the following data model will be studied, see for exam-

*This work was supported in part by the Swedish Research Council for Engineering Sciences (TFR).

ple [5]:

$$\mathbf{z}(t) = \mathbf{\Gamma}\mathbf{x}(t) + \mathbf{e}(t) \tag{1}$$

where $\mathbf{e}(t)$ is additive noise with arbitrary covariance matrix $\mathbf{C}_e = E[\mathbf{e}(t)\mathbf{e}^H(t)]$. The structure of $\mathbf{\Gamma}$ may generally be arbitrary, but in this paper we focus on the special case of a ULA. The matrix $\mathbf{\Gamma}$ is deterministic and is constructed as $\mathbf{\Gamma} = [\gamma(\omega_1)\dots\gamma(\omega_n)]$ where $\gamma(\omega_k) = [1\ e^{j\omega_k}\dots e^{j\omega_k(m-1)}]^T$ is a so-called steering-vector. Implicit in the definitions above is that the subspace $\mathcal{R}(\mathbf{\Gamma})$ might be slowly time-varying, *i.e.* $\mathcal{R}(\mathbf{\Gamma}) = \mathcal{R}(\mathbf{\Gamma}(t))$.

With samples $\mathbf{z}(t), t = 1,\dots,$ we are interested in deriving an efficient algorithm which estimates $\mathcal{R}(\mathbf{\Gamma})$ at time instant t, given the subspace estimate at time instant $t-1$ and the sample $\mathbf{z}(t)$. Typical for IV approaches are the following. Assume that there exists an IV vector $\boldsymbol{\xi}(t) \in \mathcal{C}^{l\times 1}$, $l \geq n$ such that

A1: $E[\mathbf{e}(t)\boldsymbol{\xi}^H(t)] = \mathbf{0}$

A2: $\rho(E[\mathbf{x}(t)\boldsymbol{\xi}^H(t)]) = \rho(\mathbf{C}_{x\xi}) = n$

Assumption **A2** is made in order to ensure that $\rho(\mathbf{\Gamma}\mathbf{C}_{x\xi}) = n$, which implies that $\mathcal{R}(\mathbf{\Gamma}\mathbf{C}_{x\xi}) = \mathcal{R}(\mathbf{\Gamma})$. For the time series case, this assumption is discussed in [2]. Assumption **A2** is not necessary for guaranteeing DOA identifiability, see [4]. However, for reasons of simplicity, it is assumed to hold throughout the paper.

3 Basic IV algorithm (IV-PAST)

Consider the solutions to ($\mathbf{W} \in \mathcal{C}^{m\times n}$):

$$\begin{aligned} &V(\mathbf{W}) = \\ &E[\mathbf{z}(t)\boldsymbol{\xi}^H(t)] - \mathbf{W}\mathbf{W}^H E[\mathbf{z}(t)\boldsymbol{\xi}^H(t)] = \mathbf{0} \\ &\Leftrightarrow V(\mathbf{W}) = \mathbf{\Gamma}\mathbf{C}_{x\xi} - \mathbf{W}\mathbf{W}^H\mathbf{\Gamma}\mathbf{C}_{x\xi} = \mathbf{0}. \end{aligned} \tag{2}$$

Provided **A2** is fulfilled, by definition of the orthogonal projector, all solutions to (2) will be of the form $\mathbf{W} = \mathbf{U}_\Gamma\mathbf{T}$ where $\mathcal{R}(\mathbf{U}_\Gamma) = \mathcal{R}(\mathbf{\Gamma})$, $\mathbf{U}_\Gamma \in \mathcal{C}^{m\times n}$ is orthogonal, and $\mathbf{T} \in \mathcal{C}^{n\times n}$ is an arbitrary unitary matrix. Thus, for all solutions to (2) we have that

$$\mathbf{W}\mathbf{W}^H = \mathbf{\Pi}_\Gamma = \mathbf{\Gamma}\left(\mathbf{\Gamma}^H\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}^H = \mathbf{U}_\Gamma\mathbf{U}_\Gamma^H \tag{3}$$

is the orthogonal projector onto the space spanned by the columns of $\mathbf{\Gamma}$. To derive a practical algorithm, consider the solutions to (compare with (2))

$$\begin{aligned} \bar{V}(\mathbf{W}(t)) = \sum_{k=1}^{t}\gamma^{t-k}\Big(&\mathbf{z}(k)\boldsymbol{\xi}^H(k) - \\ &-\mathbf{W}(t)\mathbf{W}^H(t)\mathbf{z}(k)\boldsymbol{\xi}^H(k)\Big) = \mathbf{0}, \end{aligned} \tag{4}$$

where γ is the forgetting factor ($0 \ll \gamma < 1$). Using the projection approximation idea of [5], $\mathbf{h}(k) \triangleq \mathbf{W}^H(k-1)\mathbf{z}(k)$, then gives

$$\mathbf{W}(t) = \hat{\mathbf{C}}_{z\xi}(t)\hat{\mathbf{C}}_{h\xi}^{-1}(t) \tag{5}$$

with obvious definitions of the estimates of the covariance matrices. Using the matrix inversion lemma, the following algorithm is obtained:

$$\mathbf{h}(t) = \mathbf{W}^H(t-1)\mathbf{z}(t) \tag{6a}$$

$$\boldsymbol{\varepsilon}(t) = \mathbf{z}(t) - \mathbf{W}(t-1)\mathbf{h}(t) \tag{6b}$$

$$\mathbf{W}(t) = \mathbf{W}(t-1) + \boldsymbol{\varepsilon}(t)\mathbf{K}(t) \tag{6c}$$

$$\mathbf{P}(t) = \frac{1}{\gamma}\left(\mathbf{P}(t-1) - \mathbf{P}(t-1)\mathbf{h}(t)\mathbf{K}(t)\right) \tag{6d}$$

$$\mathbf{K}(t) = \frac{\boldsymbol{\xi}^H(t)\mathbf{P}(t-1)}{\gamma + \boldsymbol{\xi}^H(t)\mathbf{P}(t-1)\mathbf{h}(t)} \tag{6e}$$

where $\mathbf{P}(t) = \hat{\mathbf{C}}_{h\xi}^{-1}(t)$. In the above we have assumed that initial values $\mathbf{W}(0), \mathbf{P}(0)$ are given. These initial values only affect the transient behavior and are not important for the steady-state performance of the algorithm. They can for example be taken as any full-rank matrices.

Due to the introduced approximations, the columns of $\mathbf{W}(t)$ will not be orthonormal. However, simulations show that they are 'nearly' orthonormal. Some applications may require orthonormal columns, which may call for a reorthogonalization scheme such as Gram-Schmidt. However, in our simulations no orthogonalization is performed.

Note that we have constrained the dimension of the IV-vector $\boldsymbol{\xi}(t)$ to $l = n$, which implies that no rank-reduction of the sample cross covariance matrix is performed. So, why not take $\mathbf{W}(t) = \hat{\mathbf{C}}_{z\xi}(t)$? The main motivation is that the matrix $\hat{\mathbf{C}}_{h\xi}^{-1}(t)$ post-multiplying in (5) forces the columns of $\mathbf{W}(t)$ to be 'nearly' orthonormal, resulting in good conditioning. Thus, IV-PAST can be thought of as a simple way to approximately orthogonolize the columns of $\hat{\mathbf{C}}_{z\xi}(t)$. The basic IV-algorithm will also serve as a preview of a more general rank-reducing IV-approach described in the following section.

4 Extended IV-algorithm (EIV-PAST)

A straightforward extension of the previous discussion, $l > n$, leads to the following criterion

$$V(\mathbf{W}(t)) = \left\|\hat{\mathbf{C}}_{z\xi}(t) - \mathbf{W}(t)\mathbf{W}^H(t)\hat{\mathbf{C}}_{z\xi}(t)\right\|_F^2 \tag{7}$$

where $\mathbf{W}(t) \in \mathcal{C}^{m\times n}$, $\hat{\mathbf{C}}_{z\xi}(t) \in \mathcal{C}^{m\times l}$. This approach corresponds to what in [3, Section 8] is called the *Extended* IV estimate. Without loss of generality we assume that $\rho(\mathbf{W}(t)) = n$. With probability 1 (w.p.1), $\rho(\hat{\mathbf{C}}_{z\xi}(t)) = \min(m,l) = \tilde{n}$, but $\rho(\mathbf{C}_{z\xi}) = n < l$. Consequently, a low-rank approximation of $\hat{\mathbf{C}}_{z\xi}(t)$ is desired. Thus, the following theorem is needed.

Theorem 1 *Let $\hat{\mathbf{C}}_{z\xi}(t)$ have the SVD*

$$\hat{\mathbf{C}}_{z\xi}(t) = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^H = \begin{bmatrix}\hat{\mathbf{U}}_s & \hat{\mathbf{U}}_n\end{bmatrix}\begin{bmatrix}\hat{\mathbf{\Sigma}}_s & \mathbf{0}\\ \mathbf{0} & \hat{\mathbf{\Sigma}}_n\end{bmatrix}\begin{bmatrix}\hat{\mathbf{V}}_s^H\\ \hat{\mathbf{V}}_n^H\end{bmatrix} \tag{8}$$

where $\hat{\mathbf{U}}_s \in \mathcal{C}^{m\times n}$. The remaining partitions are of appropriate dimensions. $\mathbf{W}(t)$ is a stationary point of (7) iff $\mathbf{W}(t) = \hat{\mathbf{U}}\mathbf{T}$, where $\hat{\mathbf{U}}$ denotes any n left singular vectors of $\hat{\mathbf{U}}$ and $\mathbf{T} \in \mathcal{C}^{n\times n}$ denotes an arbitrary unitary matrix. All stationary points of $V(\mathbf{W}(t))$ are saddle points except when $\hat{\mathbf{U}} = \hat{\mathbf{U}}_s$. In this case $V(\mathbf{W}(t))$ attains the global minimum. Note that for this choice, $\mathbf{W}(t)\mathbf{W}^H(t)\hat{\mathbf{C}}_{z\xi}(t) = \hat{\mathbf{U}}_s\hat{\mathbf{\Sigma}}_s\hat{\mathbf{V}}_s^H$, which in the sense of the Frobenius norm is the best possible rank n approximation of $\hat{\mathbf{C}}_{z\xi}(t)$.

Proof: *See [2].* □

Once again the projection approximation is applied:

$$\mathbf{W}^H(t)\hat{\mathbf{C}}_{z\xi}(t) = \mathbf{W}^H(t)\sum_{k=1}^{t}\gamma^{t-k}\mathbf{z}(k)\boldsymbol{\xi}^H(k) \tag{9}$$

$$\approx \sum_{k=1}^{t}\gamma^{t-k}\underbrace{\mathbf{W}^H(k-1)\mathbf{z}(k)}_{\mathbf{h}(k)}\boldsymbol{\xi}^H(k) \triangleq \hat{\mathbf{C}}_{h\xi}(t)$$

which gives the (quadratic) criterion

$$\bar{V}(\mathbf{W}(t)) = \left\|\hat{\mathbf{C}}_{z\xi}(t) - \mathbf{W}(t)\hat{\mathbf{C}}_{h\xi}(t)\right\|_F^2. \tag{10}$$

The minimizing argument of (10) is given by

$$\mathbf{W}(t) = \hat{\mathbf{C}}_{z\xi}(t)\hat{\mathbf{C}}_{h\xi}^{\dagger}(t) \tag{11}$$

where $(.)^{\dagger}$ denotes the Moore-Penrose pseudo-inverse. This approach will in most cases improve the accuracy of the estimates. For example, in Section 5 we will see that the tracking capabilities are much more 'well-behaved' in this case. In Appendix A an efficient $(3ml + \mathcal{O}(mn)$ complex multiplications) recursive updating formula of (11) is given. Note that the matrix inversion that arises in (14d) is of size (2×2), so it is a simple matter to invert it.

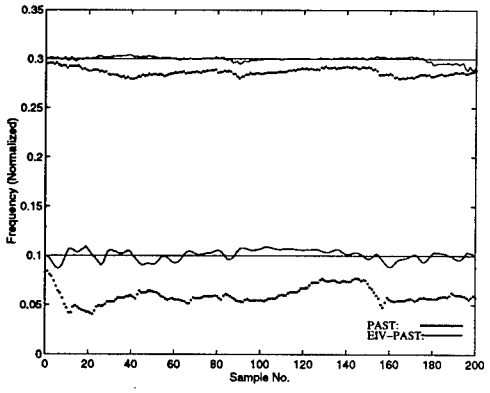


Figure 1. One realization of the frequency estimates. SNR=5 dB, $e(t) = \frac{1}{1-0.8q^{-1}}\varepsilon(t)$. $\gamma = 0.97$

5 Examples

Consider the scalar signal

$$z(t) = \sum_{j=1}^{2} a_j \cos(2\pi f_j(t)t + \varphi_j) + e(t) \tag{12}$$

where $a_1 = a_2 = \sqrt{2}$. The random phases φ_j are independent and uniformly distributed in $(-\pi, \pi)$. Thus, $n = 4$. Chose $m = 8$ which gives $\mathbf{z}(t) = [z(t), \ldots, z(t+7)]^T$. The (temporal) IV-vector is chosen as $\boldsymbol{\xi}(t) = [z(t-M), \ldots, z(t-M-l+1)]^T$ with $M = 11$. The number of instruments l is for IV-PAST $l = 4$, and for EIV-PAST $l = m = 8$. The frequencies are estimated using the ESPRIT-approach, *i.e.* the angles of the eigenvalues of $\mathbf{W}_{2:m}^{\dagger}(t)\mathbf{W}_{1:m-1}(t)$, where $\mathbf{W}_{i:j}$ denotes rows i to j of $\mathbf{W}$. For all algorithms, the following initial values were used:

$$\mathbf{P}(0) = \mathbf{I}_n \quad \mathbf{W}(0) = [\mathbf{I}_n\ \mathbf{0}_{(m-n)\times n}^T]^T. \tag{13}$$

However, the transient is typically not shown. In the simulations, $e(t) = \frac{1}{1-0.8q^{-1}}\varepsilon(t)$, where q^{-1} is the delay operator and $\varepsilon(t)$ is white Gaussian noise. Note that for this noise process, condition **A1** is violated: $E[\mathbf{e}(t)\boldsymbol{\xi}^H(t)] \neq \mathbf{0}$. In the simulations, $\gamma = 0.97$. From Fig.1 we see that the IV based approach clearly reduces the bias compared with PAST. Based on this observation, further simulations with PAST are omitted. The next example illustrates the tracking performance of the algorithms. The performance is also compared with the performance obtained with the frequency estimates obtained from the n dominant left singular vectors of $\hat{\mathbf{C}}_{z\xi}(t)$. We consider a step-change in a frequency, see Fig.2 and Fig.3. In Fig.3 we have

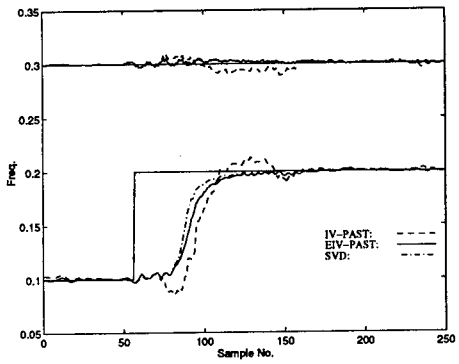


Figure 2. One realization of the frequency estimates. SNR=8 dB, $e(t) = \frac{1}{1-0.8q^{-1}}\varepsilon(t)$**.** $\gamma = 0.97$

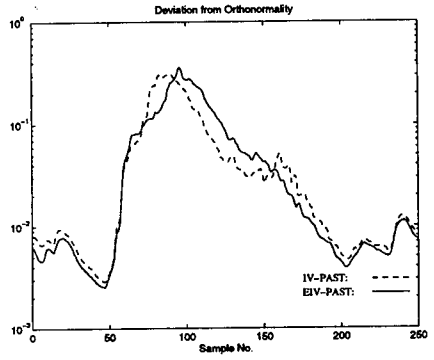


Figure 3. Deviation from orthonormality for a step-change. SNR=8 dB, $e(t) = \frac{1}{1-0.8q^{-1}}\varepsilon(t)$**.** $\gamma = 0.97$

used the following measure of deviation from orthonormality: $\|\mathbf{W}^H(t)\mathbf{W}(t) - \mathbf{I}_n\|_F$. The basic IV-algorithm shows a tendency to 'over-shoot', but this behavior is reduced by EIV-PAST. This is perhaps the major improvement offered by the Extended IV method. Note also that the estimates of the constant frequency are less affected for EIV-PAST than those of IV-PAST. Note also that it is almost impossible to distinguish the EIV-PAST estimates from those of the SVD, except during the transient phase.

6 Conclusions

In this paper Instrumental Variable generalizations of the subspace tracking algorithm PAST have been proposed. The presented algorithms are able to track slowly time-varying subspaces in colored noise fields. One requirement is that we must be able to find an IV-vector that is uncorrelated with the noise vector. Additionally, a certain rank requirement must be fulfilled. The conclusions are that an IV approach in our examples improves the results when the noise is not spatially white.

A Appendix

In this appendix we give the recursive updating formulas for the Extended IV-PAST algorithm, $\mathbf{P}(t) = \left(\hat{\mathbf{C}}_{h\xi}(t)\hat{\mathbf{C}}_{h\xi}^H(t)\right)^{-1}$. See [2] for a derivation.

$$\mathbf{W}(t) = \mathbf{W}(t-1) + \varepsilon(t)\mathbf{K}(t) \quad (14a)$$

$$\varepsilon(t) = \mathbf{v}(t) - \mathbf{W}(t-1)\mathbf{\Phi}(t) \quad (14b)$$

$$\mathbf{X}(t) = \gamma^2\mathbf{\Lambda}(t) + \mathbf{\Phi}^H(t)\mathbf{P}(t-1)\mathbf{\Phi}(t) \quad (14c)$$

$$\mathbf{K}(t) = \mathbf{X}^{-1}(t)\mathbf{\Phi}^H(t)\mathbf{P}(t-1) \quad (14d)$$

$$\mathbf{\Phi}(t) = [\mathbf{w}(t)\ \mathbf{h}(t)] \quad (14e)$$

$$\mathbf{w}(t) = \hat{\mathbf{C}}_{h\xi}(t-1)\boldsymbol{\xi}(t) \quad (14f)$$

$$\gamma^2\mathbf{\Lambda}(t) = \begin{bmatrix} -\boldsymbol{\xi}^H(t)\boldsymbol{\xi}(t) & \gamma \\ \gamma & 0 \end{bmatrix} \quad (14g)$$

$$\mathbf{v}(t) = \left[\hat{\mathbf{C}}_{z\xi}(t-1)\boldsymbol{\xi}(t)\ \ \mathbf{z}(t)\right] \quad (14h)$$

$$\hat{\mathbf{C}}_{h\xi}(t) = \gamma\hat{\mathbf{C}}_{h\xi}(t-1) + \mathbf{h}(t)\boldsymbol{\xi}^H(t) \quad (14i)$$

$$\hat{\mathbf{C}}_{z\xi}(t) = \gamma\hat{\mathbf{C}}_{z\xi}(t-1) + \mathbf{z}(t)\boldsymbol{\xi}^H(t) \quad (14j)$$

$$\mathbf{h}(t) = \mathbf{W}^H(t-1)\mathbf{z}(t) \quad (14k)$$

$$\mathbf{P}(t) = \frac{1}{\gamma^2}\left(\mathbf{P}(t-1) - \mathbf{P}(t-1)\mathbf{\Phi}(t)\mathbf{K}(t)\right) \quad (14l)$$

References

[1] P. Comon and G. H. Golub. "Tracking a few Extreme Singular Values and Vectors in Signal Processing". *Proceedings of the IEEE*, 78:1327–1343, Aug. 1990.

[2] T. Gustafsson. Instrumental variable subspace tracking using projection approximation. Submitted to IEEE Trans SP. Technical report CTH-TE-42, available from the author, Jan 1996.

[3] T. Söderström and P. Stoica. *System Identification.* Prentice-Hall, London, U.K., 1989.

[4] M. Viberg, P. Stoica, and B. Ottersten. "Array Processing in Correlated Noise Fields Based on Instrumental Variables and Subspace Fitting". *IEEE Trans. SP*, May 1995.

[5] B. Yang. "Projection Approximation Subspace Tracking". *IEEE Trans. SP*, 43(1):95–107, Jan . 1995.

A Maximum a Posteriori Approach to Beamforming in the Presence of Calibration Errors

A. SWINDLEHURST
Dept. of Elec. & Comp. Engineering
Brigham Young University
Provo, UT 84602

Abstract

The performance of DF-based beamformers is seriously degraded in situations where the array is imprecisely calibrated, or when the spatial coherence of the signal wavefronts is perturbed. When the calibration errors or perturbation may be characterized by a set of parameters drawn from a known Gaussian distribution, a maximum a posteriori *(MAP) estimator may be used to separately estimate the directions of arrival and the perturbation parameters, resulting in essentially an on-line auto-calibration. This paper examines the improvement that results from using the MAP auto-calibrated steering vectors in standard DF-based beamformers to estimate the received signal waveforms and suppress unwanted interference. For the special case of additive unstructured calibration errors and uncorrelated signals, it is shown that the MAP beamformer is similar in form to so-called "subspace corrected" approaches.*

1. Introduction

All methods for direction-finding (DF) and DF-based beamforming rely on the availability of information about the array response, and assume the signal wavefronts have perfect spatial coherence. Depending on the degree to which the actual response or wavefronts differ from their nominal values, DF and beamformer performance may be significantly degraded. To account for these types of perturbations, a slightly generalized model for the array response will be considered in this paper. The response will be parameterized not only by the directions of arrival (DOAs) of the signals, but also by a vector of perturbation or "nuisance" parameters that describe deviations of the response from its nominal value. These parameters can include, for example, displacements of the antenna elements from their nominal positions, uncalibrated receiver gain and phase offsets, *etc.*. With such a model, a natural approach is to attempt to estimate the unknown nuisance parameters simultaneously with the signal parameters. Such methods are referred to as *auto-calibration* techniques, and have been proposed by a number of authors, including [1, 2, 3, 4] among many others.

When auto-calibration techniques are employed, it is critical to determine whether both the signal and nuisance parameters are identifiable. In certain cases they are not; for example, one cannot uniquely estimate both DOAs and sensor phase characteristics (unless of course additional information is available, such as sources in known locations, *etc.*). The identifiability problem can be alleviated if the perturbation parameters are assumed to be drawn from some known *a priori* distribution. While this itself represents a form of additional information, it has the advantage of allowing an optimal maximum *a posteriori* (MAP) solution to the problem to be formulated. In [4] it is shown that, by using an asymptotically equivalent approximation to the resulting MAP criterion, the estimation of the signal and nuisance parameters can be decoupled, leading to a significant simplification of the problem.

Presumably, any of the above auto-calibration methods would provide not only improved DOA estimates, but also calibration information that would be useful in beamformer implementation. In this paper, beamformer performance is investigated for the case where the optimal MAP perturbation parameter estimates of [4] are used to update the array calibration. Simulations demonstrate that such an approach can result in a significant performance improvement, measured using either interference rejection capability or mean-squared error. In addition, for simple additive unstructured calibration errors, the MAP approach is shown in certain cases to yield a beamformer similar to the *subspace corrected* algorithms described in [5, 6].

2. Mathematical Model and Algorithms

The response of an arbitrary array of m sensors for a given DOA θ will be denoted by the m-vector $a(\theta, \rho)$, which is parameterized by a vector $\rho \in \mathbb{R}^p$ that describes the array perturbation. The array output is then modeled by the following familiar equation:

$$\mathbf{x}(t) = [a(\theta_1, \rho) \,\vdots\, \cdots \,\vdots\, a(\theta_d, \rho)] \begin{bmatrix} s_1(t) \\ \vdots \\ s_d(t) \end{bmatrix} + \mathbf{n}(t) \quad (1)$$

$$= \mathbf{A}(\theta, \rho)\mathbf{s}(t) + \mathbf{n}(t) \,, \quad (2)$$

where $\mathbf{s}(t)$ and $\mathbf{n}(t)$ represent the received signals and noise, respectively. It will be assumed that for a given collect, N samples are taken from the array. Both $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are assumed to be temporally white zero-mean complex Gaussian random processes, with covariances given by $\sigma^2\mathbf{I}$ and $\mathbf{P}$, respectively. The perturbation term ρ is also assumed to be drawn from a Gaussian distri-

0-8186-7576-4/96 $5.00

bution with known mean ρ_0 (corresponding to the nominal, unperturbed array response) and covariance $\mathbf{\Omega}$.

Given the above, the covariance of the array output and its eigendecomposition may be written as

$$\mathbf{R} = \mathbf{A}(\theta,\rho)\mathbf{P}\mathbf{A}^*(\theta,\rho) + \sigma^2\,\mathbf{I} = \mathbf{E}_s\mathbf{\Lambda}_s\mathbf{E}_s^* + \sigma^2\,\mathbf{E}_n\mathbf{E}_n^*,$$

where $\mathbf{\Lambda}_s$ contains the d largest eigenvalues, and the columns of the $m \times d$ matrix $\mathbf{E}_s$ are the corresponding unit-norm eigenvectors. Similarly, the columns of $\mathbf{E}_n$ are the $m-d$ eigenvectors corresponding to σ^2.

2.1. An Asymptotic MAP Estimator

In [4], it is shown that estimates of θ and ρ asymptotically equivalent to those from the exact MAP estimator may be obtained by setting[1]

$$\hat{\theta} = \arg\min_{\theta}\ \mathbf{a}_0^*\hat{\mathbf{M}}\mathbf{a}_0 - \hat{\mathbf{f}}^T\hat{\mathbf{\Gamma}}^{-1}\hat{\mathbf{f}} \tag{3}$$

$$\hat{\rho} = \rho_0 - \mathbf{\Gamma}^{-1}\mathbf{f}\,, \tag{4}$$

where

$$\mathbf{a}_0 = \mathrm{vec}(\mathbf{A}(\theta,\rho_0))\ ,\ \hat{\mathbf{M}} = \hat{\mathbf{U}}^T \otimes (\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^*) \tag{5}$$

$$\hat{\mathbf{U}} = \hat{\sigma}^{-2}\hat{\mathbf{A}}^\dagger\hat{\mathbf{E}}_s\tilde{\mathbf{A}}^2\hat{\mathbf{A}}_s^{-1}\hat{\mathbf{E}}_s^*\hat{\mathbf{A}}^{\dagger *}\ ,\ \hat{\mathbf{f}} = \mathrm{Re}\{\hat{\mathbf{D}}_\rho^*\hat{\mathbf{M}}\mathbf{a}_0\} \tag{6}$$

$$\hat{\mathbf{\Gamma}} = \mathrm{Re}\left\{\hat{\mathbf{D}}_\rho^*\hat{\mathbf{M}}\hat{\mathbf{D}}_\rho + \frac{1}{2N}\mathbf{\Omega}^{-1}\right\} \tag{7}$$

$$\mathbf{D}_\rho = \left[\frac{\partial\mathbf{a}(\theta,\rho)}{\partial\rho_1},\ldots,\frac{\partial\mathbf{a}(\theta,\rho)}{\partial\rho_p}\right]\Bigg|_{\theta,\rho_0}\,, \tag{8}$$

and where $\hat{\sigma}^2$ and $\hat{\mathbf{A}}$ are "consistent" estimates determined from some initial estimation step. The above approach is quite general in that, by proper choice of ρ, it can be applied to arbitrary types of model errors. Another key advantage is that estimation of θ and ρ is decoupled; a search is required only for the d DOA parameters in θ, and not for ρ (which is calculated directly given $\hat{\theta}$). Other properties of the algorithm are outlined in [4].

2.2. Optimal Beamformers

The minimum mean squared error (MSE) beamformer weights are easily shown to be

$$\mathbf{W}_{MSE} = \mathbf{R}^{-1}\mathbf{R}_{xs} = \mathbf{R}^{-1}\mathbf{A}(\theta,\rho)\mathbf{P}\,. \tag{9}$$

When the desired signal is uncorrelated with the interference, $\mathbf{P}$ is diagonal and the minimum MSE solution is just a scaled version of the so-called minimum variance distortionless response (MVDR) beamformer:

$$\mathbf{W} = \frac{\mathbf{R}^{-1}a(\hat{\theta})}{a^*(\hat{\theta})\mathbf{R}^{-1}a(\hat{\theta})}\,. \tag{10}$$

In the general case where the signal and interference are correlated, the optimal weights depend on the signals themselves through $\mathbf{R}_{xs}$ or $\mathbf{P}$, and thus they cannot be used directly (*i.e.*, without a training sequence, for example). In the approach of [8], the quantities $\mathbf{P}$ and $\mathbf{R}$ in (9) are replaced by their *structured* ML estimates:

$$\hat{\mathbf{P}}_s = \mathbf{A}_0^\dagger(\hat{\mathbf{R}} - \hat{\sigma}^2\mathbf{I})\mathbf{A}_0^{\dagger *}\ ,\ \hat{\mathbf{R}}_s = \mathbf{A}_0\hat{\mathbf{P}}_s\mathbf{A}_0^* + \hat{\sigma}^2\mathbf{I}\,,$$

where $\mathbf{A}_0 = \mathbf{A}(\hat{\theta},\rho_0)$, $(\cdot)^\dagger$ denotes a (left) pseudo-inverse, and $\hat{\mathbf{R}}$ is a sample estimate of $\mathbf{R}$.

Since calibration errors were not addressed in [8], the nominal model ρ_0 was used to calculate the beamformer weights. Nevertheless, the method performs well when calibration errors are present, as recently demonstrated in [9]. On the other hand, the MVDR approach is well known to be hyper-sensitive to array perturbations, especially at high SNR. While *ad hoc* methods employing artificial noise injection have been used to combat this problem, other techniques based on *subspace corrected* (SC) weights have found success in experimental systems [5, 6]. In these approaches, the $\mathbf{R}^{-1}$ term in (10) is replaced by $\mathbf{E}_s\mathbf{\Lambda}_s^{-1}\mathbf{E}_s^*$. This is equivalent to projecting $a(\hat{\theta})$ onto the signal subspace prior to forming the MVDR weights.

One of the goals of this paper is to study the improvement that results from using the method of [8] with $\mathbf{A}(\hat{\theta},\hat{\rho})$ rather than $\mathbf{A}(\hat{\theta},\rho_0)$, where $\hat{\rho}$ is obtained from the MAP estimator in (4). This approach will be referred to as the MAP beamformer in the sequel. In the next section, an interesting connection is made between the MAP beamformer and the SC-MVDR method. In particular, it is shown that for simple unstructured array errors and uncorrelated signals, the SC-MVDR and MAP weights have a very similar form.

3. Some Special Cases

For the moment, consider the following simple unstructured model for the perturbed array response:

$$\mathbf{A}(\theta,\rho) = \mathbf{A}(\theta) + \tilde{\mathbf{A}} \tag{11}$$

$$\rho = \begin{bmatrix} \mathrm{Re}\{\mathrm{vec}(\tilde{\mathbf{A}})\} \\ \mathrm{Im}\{\mathrm{vec}(\tilde{\mathbf{A}})\} \end{bmatrix}\,, \tag{12}$$

where the columns of $\tilde{\mathbf{A}}$, denoted $\tilde{a}_i$, are modeled as zero mean Gaussian random vectors with moments

$$\mathbf{E}[\tilde{a}_i\tilde{a}_k^*] = \nu_{ik}\,\mathbf{I}\ ,\ \mathbf{E}[\tilde{a}_i\tilde{a}_k^T] = 0 \qquad i,k = 1,\cdots,d. \tag{13}$$

[1] Strictly speaking, the equivalence of the above estimator and the optimal MAP approach only holds for first order errors $\rho - \rho_0$ that are "of the same order" as the finite sample effects of the noise. In other cases (particularly those model errors are dominant), a different approach should be used. For more details, see [4, 7].

This model corresponds to an additive, circularly symmetric complex array perturbation that is uncorrelated from sensor to sensor, but possibly θ-dependent. It is easy to verify that under these assumptions, the covariance of $\boldsymbol{\rho}$ is given by

$$\boldsymbol{\Omega} = \frac{1}{2}\begin{bmatrix} \mathrm{Re}\{\boldsymbol{\Upsilon}\}\otimes\mathbf{I} & -\mathrm{Im}\{\boldsymbol{\Upsilon}\}\otimes\mathbf{I} \\ \mathrm{Im}\{\boldsymbol{\Upsilon}\}\otimes\mathbf{I} & \mathrm{Re}\{\boldsymbol{\Upsilon}\}\otimes\mathbf{I} \end{bmatrix}, \qquad (14)$$

where the i,k^{th} element of the matrix $\boldsymbol{\Upsilon}$ is ν_{ik}.

It is interesting to examine the form of the MAP estimate $\hat{\boldsymbol{\rho}}$ for this case. To begin with, note that for the above model $\boldsymbol{\rho}_0 = 0$ and $\hat{\mathbf{D}}_{\boldsymbol{\rho}} = [\mathbf{I} \;\; j\mathbf{I}]$, where $\mathbf{I}$ is $md \times md$. Thus, $\hat{\boldsymbol{\rho}} = -\boldsymbol{\Gamma}^{-1}\mathbf{f}$, and

$$\boldsymbol{\Gamma} = \begin{bmatrix} \mathrm{Re}\{\hat{\mathbf{M}} + \frac{1}{N}\boldsymbol{\Upsilon}^{-1}\otimes\mathbf{I}\} & -\mathrm{Im}\{\hat{\mathbf{M}} + \frac{1}{N}\boldsymbol{\Upsilon}^{-1}\otimes\mathbf{I}\} \\ \mathrm{Im}\{\hat{\mathbf{M}} + \frac{1}{N}\boldsymbol{\Upsilon}^{-1}\otimes\mathbf{I}\} & \mathrm{Re}\{\hat{\mathbf{M}} + \frac{1}{N}\boldsymbol{\Upsilon}^{-1}\otimes\mathbf{I}\} \end{bmatrix}$$

$$\mathbf{f} = \begin{bmatrix} \mathrm{Re}\{\hat{\mathbf{M}}\mathbf{a}_0\} \\ \mathrm{Im}\{\hat{\mathbf{M}}\mathbf{a}_0\} \end{bmatrix}.$$

Using the fact that, for any invertible matrix $\mathbf{Z}$,

$$\begin{bmatrix} \mathrm{Re}\{\mathbf{Z}\} & -\mathrm{Im}\{\mathbf{Z}\} \\ \mathrm{Im}\{\mathbf{Z}\} & \mathrm{Re}\{\mathbf{Z}\} \end{bmatrix}^{-1} = \begin{bmatrix} \mathrm{Re}\{\mathbf{Z}^{-1}\} & -\mathrm{Im}\{\mathbf{Z}^{-1}\} \\ \mathrm{Im}\{\mathbf{Z}^{-1}\} & \mathrm{Re}\{\mathbf{Z}^{-1}\} \end{bmatrix}, \qquad (15)$$

it is easy to show that

$$\hat{\boldsymbol{\rho}} = -\begin{bmatrix} \mathrm{Re}\left\{\left(\hat{\mathbf{M}} + \frac{1}{N}\boldsymbol{\Upsilon}^{-1}\otimes\mathbf{I}\right)^{-1}\hat{\mathbf{M}}\mathbf{a}_0\right\} \\ \mathrm{Im}\left\{\left(\hat{\mathbf{M}} + \frac{1}{N}\boldsymbol{\Upsilon}^{-1}\otimes\mathbf{I}\right)^{-1}\hat{\mathbf{M}}\mathbf{a}_0\right\} \end{bmatrix}. \qquad (16)$$

A further simplification of (16) is possible that is quite revealing. Using the definition of $\hat{\mathbf{M}}$ in (5), note that

$$\left(\hat{\mathbf{M}} + \tfrac{1}{N}\boldsymbol{\Upsilon}^{-1}\otimes\mathbf{I}\right)^{-1} = \left(\hat{\mathbf{U}}^T + \tfrac{1}{N}\boldsymbol{\Upsilon}^{-1}\right)^{-1}\otimes(\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^*) + N\boldsymbol{\Upsilon}\otimes(\hat{\mathbf{E}}_s\hat{\mathbf{E}}_s^*).$$

Multiplying this equation on the right by $\hat{\mathbf{M}}\mathbf{a}_0$ and simplifying then yields

$$\hat{\boldsymbol{\rho}} = -\begin{bmatrix} \mathrm{Re}\left\{\left[\left(\mathbf{I} + \frac{1}{N}(\boldsymbol{\Upsilon}\mathbf{U}^T)^{-1}\right)^{-1}\otimes(\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^*)\right]\mathbf{a}_0\right\} \\ \mathrm{Im}\left\{\left[\left(\mathbf{I} + \frac{1}{N}(\boldsymbol{\Upsilon}\mathbf{U}^T)^{-1}\right)^{-1}\otimes(\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^*)\right]\mathbf{a}_0\right\} \end{bmatrix}. \qquad (17)$$

Finally, using (12) and properties of the Kronecker product, the MAP estimate of the array response becomes

$$\mathbf{A}(\hat{\theta},\hat{\boldsymbol{\rho}}) = \mathbf{A}(\hat{\theta}) - \hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^*\mathbf{A}(\hat{\theta})\left(\mathbf{I} + \frac{1}{N}(\boldsymbol{\Upsilon}\mathbf{U}^T)^{-1}\right)^{-1}. \qquad (18)$$

The key point of interest is that, if $\boldsymbol{\Upsilon}^{-1}/N \to 0$, then the MAP estimate of the array response converges to a subspace corrected version of the nominal response:

$$\lim_{\boldsymbol{\Upsilon}^{-1}/N\to 0}\mathbf{A}(\hat{\theta},\hat{\boldsymbol{\rho}}) = \hat{\mathbf{E}}_s\hat{\mathbf{E}}_s^*\mathbf{A}(\hat{\theta}).$$

Furthermore, if the estimated MAP array response is used in (10), the MVDR beamformer (10) will converge to the SC-MVDR approach. The condition $\boldsymbol{\Upsilon}^{-1}/N \to 0$ occurs either with a large data sample, or when the array perturbation is large. In either case, the information provided by the prior distribution of $\boldsymbol{\rho}$ is of little value, and is essentially ignored by the MAP criterion. This observation provides some theoretical justification for the SC-MVDR technique, which previously had been derived using *ad hoc* (but well motivated) reasoning. However, in cases where the prior cannot be neglected, using SC response vectors for beamforming will not be optimal and significant degradation can result. This is seen in the simulation examples described later.

3.1. Gain and Phase Errors

For arrays composed of nominally identical elements, a common approach used to describe deviations in the array response attempts to model the non-uniform gain and phase effects of the receiver electronics behind each antenna element. In this model, the nominal response is perturbed by an unknown complex diagonal matrix:

$$\mathbf{A}(\theta,\boldsymbol{\rho}) = \mathbf{G}\mathbf{A}(\theta), \quad \boldsymbol{\rho} = \begin{bmatrix} \mathrm{Re}\{\mathbf{g}\} \\ \mathrm{Im}\{\mathbf{g}\} \end{bmatrix}, \qquad (19)$$

where $\mathbf{g} = \mathrm{diag}\{\mathbf{G}\}$. The mean of the distribution for $\boldsymbol{\rho}$ in this case is given by $\boldsymbol{\rho}_0 = [\mathbf{e}^T \;\; 0]^T$, where $\mathbf{e}$ is an $m \times 1$ vector of ones. For simplicity, in this discussion the covariance of $\boldsymbol{\rho}$ will be assumed to be $\boldsymbol{\Omega} = (\sigma_a^2/2)\mathbf{I}$, which implies that the individual gain and phase errors are all mutually independent and identically distributed.

The derivation of the MAP estimate of $\boldsymbol{\rho}$ and hence $\mathbf{g}$ is straightforward but somewhat cumbersome, and thus will not be presented here. However, the result is quite simple, and is given by

$$\hat{\mathbf{g}} = \left(\mathbf{I} + \sigma_a^2 N\mathbf{Z}\right)^{-1}\mathbf{e} \qquad (20)$$

$$\mathbf{Z} = \left[\sum_{i,k=1}^{d} u_{ki}\overline{a}(\hat{\theta}_i)a^T(\hat{\theta}_k)\right]\odot(\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^*), \qquad (21)$$

where u_{ki} is the k,i^{th} element of $\hat{\mathbf{U}}$, $\overline{(\cdot)}$ denotes conjugation, and $\odot$ an element-wise (Hadamard) product. Note that for very small gain/phase errors where $\sigma_a \to 0$, $\hat{\mathbf{g}} \to \mathbf{e}$ and hence $\hat{\mathbf{G}} \to \mathbf{I}$ as expected.

4. Simulation Results

In this section, the performance of the MAP beamformer is studied by means of a number of simulation examples. The first example involves a nominally unit-gain uniform linear array perturbed by an unstructured calibration error in the form of equation (11)-(14) with $\boldsymbol{\Upsilon} = \sigma_a^2\mathbf{I}$ and $\sigma_a = 0.2$. The array receives 100 samples of two 20dB SNR uncorrelated Gaussian signals with

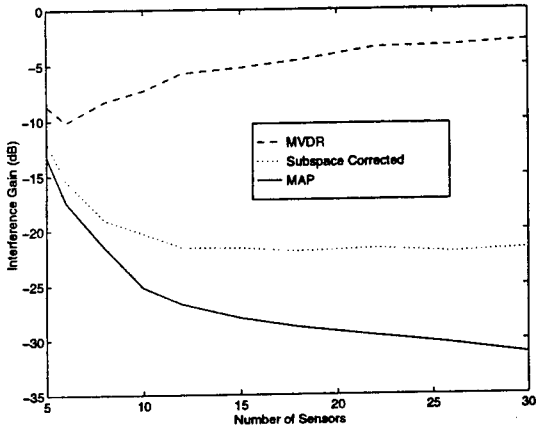


Figure 1: A Comparison of Beamformer Performance, Unstructured Calibration Errors

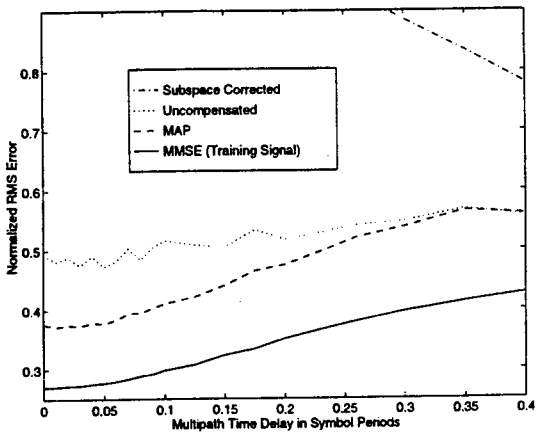


Figure 2: Root MSE Performance of Various Beamformers for a Multipath Channel

arrival angles of 5° and 15°. Using DOA estimates from the optimal MAP estimator, the relative interference rejection capability of the MVDR, SC-MVDR, and MAP beamformers was calculated for various array sizes. The results are plotted in Figure 1 based on 500 independent trials. The plot shows the gain of the beamformer weights for the 5° source in the direction of the 15° interferer (normalized for a unit gain response at 5°). The subspace correction eliminates the signal cancelation effect of the MVDR approach, but the MAP beamformer provides a significant advantage, especially for larger arrays. The above simulation was repeated assuming receiver gain/phase errors as described by (19), also with $\sigma_a = 0.2$, and a plot almost identical to Figure 1 was obtained. Algorithm performance is seen in this case to depend very little on the type of calibration error encountered.

When the signals arriving at the array are highly correlated, interference rejection is no longer an appropriate performance criterion. In such cases, an optimal beamformer will attempt to combine correlated arrivals with the desired signal to improve the quality of the resulting estimate, as measured using (for example) mean-squared error. To examine beamformer performance for the case of correlated signals, a two-ray multipath channel was simulated for various relative delays. A miscalibrated 5-element linear array was assumed to receive a random QPSK signal from $-6°$, as well as a slightly delayed copy of the signal from 6°. Both arrivals had an SNR of 0 dB, and the array was again perturbed according to (11)-(14) with $\Upsilon = \sigma_a^2 \mathbf{I}$ and $\sigma_a = 0.15$. For each trial, MAP DOA estimates were obtained based on 75 samples from the array, and normalized RMS signal errors were computed. The results are plotted in Figure 2 for various relative delays between the two arrivals. The "uncompensated" approach corresponds to the method of [8] implemented with $\mathbf{A}(\hat{\theta}, \rho_0)$ rather than $\mathbf{A}(\hat{\theta}, \hat{\rho})$ as in the MAP beamformer. The minimum MSE curve was obtained using a known 75-sample training sequence to compute the optimal weights, and was included to give an idea of the "best possible" performance.

While the SC-MVDR approach can to some degree compensate for array perturbations, it cannot eliminate signal cancelation due to the presence of a correlated arrival, and its performance in this case is quite poor. For small delays, correcting for calibration errors yields a 25-30% improvement in RMS error, which translates into a reduction in symbol error rate of approximately a factor of 6 (from .041 to .007) for this example.

References

[1] A. Paulraj and T. Kailath, "Direction-of-Arrival Estimation by Eigenstructure Methods with Unknown Sensor Gain and Phase", In *Proc. ICASSP*, pages 17.7.1–17.7.4, 1985.

[2] Y. Rockah and P. M. Schultheiss, "Array Shape Calibration Using Sources in Unknown Locations – Part I: Far-Field Sources", *IEEE Trans. on ASSP*, **35**:286–299, March 1987.

[3] A. J. Weiss and B. Friedlander, "Array Shape Calibration Using Sources in Unknown Locations - A Maximum Likelihood Approach", *IEEE Trans. on ASSP*, **37**:1958–1966, Dec. 1989.

[4] M. Viberg and A. Swindlehurst, "A Bayesian Approach to Auto-Calibration for Parametric Array Signal Processing", *IEEE Trans. on Sig. Proc.*, **42**:3495–3507, Dec. 1994.

[5] R. Schmidt and R. Franks, "Multiple Source DF Signal Processing: An Experimental System", *IEEE Trans. on Ant. and Prop.*, **34**:281–290, March 1986.

[6] B. Wahlberg, I. Mareels, and I. Webster, "Experimental and Theoretical Comparison of some Algorithms for Beamforming in Single Receiver Adaptive Arrays", *IEEE Trans. on Ant. and Prop.*, **39**:21–28, Jan. 1991.

[7] M. Viberg and A. Swindlehurst, "Analysis of the Combined Effects of Finite Samples and Model Errors on Array Processing Performance", *IEEE Trans. on Sig. Proc.*, **42**:3073–3083, Nov. 1994.

[8] B. Ottersten, R. Roy, and T. Kailath, "Signal Waveform Estimation in Sensor Array Processing", In *Proc. 23rd Asilomar Conference*, pages 787–791, Nov. 1989.

[9] J. Yang and A. Swindlehurst, "The Effects of Array Calibration Errors on DF-Based Signal Copy Performance", *IEEE Trans. on Sig. Proc.*, **43**:2724–2732, Nov. 1995.

An Iterative Virtual-ESPRIT Algorithm (IVESPA)

Egemen Gönen and Jerry M. Mendel
Signal and Image Processing Institute, Department of Electrical Engineering-Systems,
University of Southern California, Los Angeles, CA 90089-2564
Phone: (213)-740-4445, Fax: (213)-740-4651, E-Mail: gonen@sipi.usc.edu, mendel@sipi.usc.edu

Abstract

An iterative algorithm (IVESPA) for narrow-band direction finding and waveform recovery is presented which is based on the virtual-ESPRIT (VESPA) of [1]. IVESPA can handle the case where the data length is short and some of the sources have very small higher-order statistics compared to others, in which case VESPA needs more data to localize the weak sources. IVESPA can be applied to uncalibrated and arbitrary-shape arrays provided the array has two sensors having identical response–the same requirement as in VESPA. Results of a real data experiment demonstrating IVESPA are presented.

1. Introduction

Estimating the parameters of narrow-band signals using an array of sensors has been a very attractive problem of research. Typically, the parameters of interest are the directions of arrival, polarizations and the waveforms of the signals. Existing approaches to this problem can be classified into two main categories as the so-called subspace- and nonsubspace-based ones. The subspace-based methods are usually preferred, because they yield high resolution results. These methods require eigendecomposition or singular value decomposition of an array covariance or cumulant matrix, depending on the particular subspace-method used. From configuration point of view, subspace methods based on the array covariance matrix are applicable to arrays which have either analytically-known response or identical but displaced subarrays, or calibrated arrays. Among the subspace methods, VESPA of Dogan and Mendel [1], which is based on a cumulant matrix, has the lightest configuration requirements: two sensors having identical response are needed; other sensors in the array may have arbitrary and unknown responses and configurations.

Like all subspace-based methods, VESPA relies on sample statistics of the array measurements which suffer from cross terms due to the presence of multiple sources. When some of the sources have very small powers and cumulants compared to those of other sources, undesirable cross terms are present in the sample statistics of the weak sources due to the other sources for small numbers of samples. In this case, VESPA fails to accurately localize the weak sources. In practice, this case occurs when the source signals have different constellations and significantly different power levels. Note that the denser the source constellation becomes, the smaller the cumulant of the signal becomes, because the signal looks more Gaussian. For example, fourth-order cumulants of unit-power BPSK, 4QAM and 16QAM signals are -2, -1 and -0.68, respectively. In addition, sources having small powers are deemphasized during the calculation of sample higher-order statistics, because higher than second-order powers of the data are computed. As an example of this case, we will present the results of a real data experiment, in Section 3, that involves three sources: a BPSK signal, another BPSK signal with power -11.23 dB below the first one and a 16QAM signal with power -22.10 dB below the first BPSK signal.

The problem is formulated and the solution is presented in Section 2. A real data experiment is presented in Section 3. Conclusions are provided in Section 4. Throughout this paper, lower-case boldface letters represent vectors; upper-case boldface letters represent matrices; and, lower and upper-case letters represent scalars. $\mathbf{A}(i,j)$ denotes the ij-th element of $\mathbf{A}$.

2. Problem Formulation and the Proposed Algorithm

Suppose that we have an M element array containing an identical response pair of sensors. The other elements of the array may have arbitrary and unknown configuration and responses. Consider a signal scenario where there are P independent narrowband signals having nonzero fourth-order cumulants which are received by the array from directions $\{\phi_1, \cdots, \phi_P\}$. Let $\mathbf{r}(t) = [r_1(t), \cdots, r_M(t)]^T$ be the re-

0-8186-7576-4/96 $5.00

ceived signal vector which can be expressed as

$$\mathbf{r}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \quad (1)$$

where $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_P]^T$ is the $M \times P$ steering matrix, $\mathbf{s}(t) = [s_1(t), \cdots, s_P(t)]$ is the P-vector of independent sources and $\mathbf{n}(t)$ is a Gaussian noise process independent of the signals. The problem of interest is to estimate the directions $\{\phi_1, \cdots, \phi_P\}$, and to recover the sources $\{s_1(t), \cdots, s_P(t)\}$.

Before presenting the solution, we adopt the following notation for fourth-order cumulant matrices. Given two scalar processes $x_1(t)$ and $x_2(t)$ and an M-vector process $\mathbf{y}(t)$, we define $\text{cum}(x_1(t), x_2(t), \mathbf{y}(t), \mathbf{y}(t)^H)$ as the $M \times M$ matrix whose ij-th entry is $\text{cum}(x_1(t), x_2(t), y_i(t), y_j^*(t))$ where $y_i(t)$ and $y_j(t)$ are the i-th and j-th components of $\mathbf{y}(t)$, respectively.

We propose the following iterative algorithm for estimating the directions of the signals:

Step 1: Estimate the following two fourth-order cumulant matrices:

$$\begin{aligned}\mathbf{C}_{11} &\triangleq \text{cum}(r_1(t), r_1^*(t), \mathbf{r}(t), \mathbf{r}(t)^H) \\ &= \sum_{p=1}^{P} \gamma_{4,p} |\mathbf{A}(1,p)|^2 \mathbf{a}_p \mathbf{a}_p{}^H \\ &= \mathbf{A}\mathbf{\Lambda}\mathbf{A}^H \end{aligned} \quad (2)$$

and,

$$\begin{aligned}\mathbf{C}_{12} &\triangleq \text{cum}(r_2(t), r_1^*(t), \mathbf{r}(t), \mathbf{r}(t)^H) \\ &= \sum_{p=1}^{P} \gamma_{4,p} |\mathbf{A}(1,p)|^2 e^{-j\frac{2\pi d}{\lambda}\sin\phi_p} \mathbf{a}_p \mathbf{a}_p{}^H \\ &= \mathbf{A}\mathbf{\Phi}\mathbf{\Lambda}\mathbf{A}^H \end{aligned} \quad (3)$$

where $\{\gamma_{4,p}\}_{p=1}^P$ are the fourth-order cumulants of the sources; $\mathbf{\Lambda} \triangleq diag\{|\mathbf{A}(1,1)|^2\gamma_{4,1}, \cdots, |\mathbf{A}(1,P)|^2\gamma_{4,P}\}$ and $\mathbf{\Phi} \triangleq diag\{e^{-j\frac{2\pi d}{\lambda}\sin\phi_1}, \cdots, e^{-j\frac{2\pi d}{\lambda}\sin\phi_P}\}$. (2) is derived using cumulant properties [CP1], [CP3],[CP5], [CP6] in [2]. Note that the fourth-order cumulant of the additive Gaussian measurement noise is zero.

Having estimated the matrices $\mathbf{C}_{11}$ and $\mathbf{C}_{12}$, and assuming only one source is present, the arrival angle ϕ_1 of the most powerful source is obtained by following the ESPRIT solution described in the Appendix, as $\phi_1 = -\sin^{-1}\left(\frac{\lambda}{2\pi d}\angle(f_x/f_y)\right)$. Note that this step is the same as VESPA except that we assume there is only one source. The procedure in the Appendix also gives the steering vector $\mathbf{a}_1$ of the most powerful source ($i = 1$, $\mathbf{a}_1 = \mathbf{b}_1$). Then proceed by repeating the following steps for $i = 2, \cdots, P$:

Step 2: Form a modified signal, $\mathbf{r}_i(t) = \mathbf{N}_i\mathbf{r}(t)$ where $\mathbf{N}_i$ is the left null-space of the $M \times (i-1)$ matrix $\mathbf{A}_i = [\mathbf{a}_{i-1}, \cdots, \mathbf{a}_1]$, ($\mathbf{A}_2 = \mathbf{a}_1$). Doing so suppresses the most powerful $(i-1)$ sources in $\mathbf{r}(t)$.

Step 3: Estimate the following two $(M-i+1)\times(M-i+1)$ cumulant matrices:

$$\mathbf{C}_{i1} \triangleq \text{cum}(r_{i1}(t), r_{i1}^*(t), \mathbf{r}_i(t), \mathbf{r}_i(t)^H) \quad (4)$$

$$\mathbf{C}_{i2} \triangleq \text{cum}(r_{i2}(t), r_{i1}^*(t), \mathbf{r}_i(t), \mathbf{r}_i(t)^H) \quad (5)$$

where $r_{ik}(t)$ is the kth element of $(M-i+1)$-vector $\mathbf{r}_i(t)$.

Assuming only one source is present, find the modified steering vector $\mathbf{b}_i$ of that source following the procedure in the Appendix.

Step 5: Compute $\mathbf{a}_i = pinv(\mathbf{N}_i)\mathbf{b}_i$, where $pinv$ denotes pseudoinverse.

Step 6: Use the elements of $\mathbf{a}_i$ corresponding to the identical response pair of sensors to find the direction ϕ_i of the ith source. This is done as follows:

Let the identical response pair be the m-th and $(m+1)$th sensors. Then the responses of these sensors to the i-th wavefront, i.e. the m-th and $m+1$-th elements of $\mathbf{a}_i$, are in the form $a_{im} = c_i$ and $a_{i(m+1)} = c_i e^{-j\frac{2\pi d}{\lambda}\sin\phi_i}$, where d is the separation between the m-th and $(m+1)$th sensors. Consequently, ϕ_i can be found from a_{im} and $a_{i(m+1)}$.

Step 7: Recover the ith source using $\mathbf{a}_i$ in an MVDR beamformer.

3. Experimental Results

In this section we demonstrate IVESPA and compare it with VESPA by means of the following experiment, using a set of data provided by our sponsor, CRASP.

Three signals of 1000 symbols each are generated. The signal types are BPSK, BPSK and 16QAM, and they occupy a bandwidth of 350 KHz. These signals were used to modulate wavefronts designed to approximate uniform plane waves impinging upon an 8-element uniform linear array with an element spacing of one half wavelength at 900 MHz. The arrival directions are: BPSK1 at 6.3^o, BPSK2 at 25.2^o and 16QAM at 40^o. The 900 MHz 8-channel measurements were downconverted and sampled at 5.12 MHz.

The eigenvalues of the estimated 8×8 array covariance matrix are as follows:

$$10^4 * [6.25, 0.47, 0.03, 0.00, 0.00, 0.00, 0.00, 0.00] \quad (6)$$

First, VESPA was applied to this data. VESPA starts by choosing a guiding sensor pair and estimating two cumulant

matrices. In our case, any two of the sensor measurements can be used as the guiding sensor pair since the array is uniform and linear. We used the first two sensors for this purpose, and estimated the following fourth-order cumulants:

$$\begin{aligned} \mathbf{C}_1 &\triangleq \text{cum}(r_1(t), r_1^*(t), \mathbf{r}(t), \mathbf{r}(t)^H) \\ \mathbf{C}_2 &\triangleq \text{cum}(r_2(t), r_1^*(t), \mathbf{r}(t), \mathbf{r}(t)^H) \end{aligned} \quad (7)$$

Before applying the rest of the VESPA steps we first checked the singular values of $\mathbf{C}_1$ and $\mathbf{C}_2$; e.g., the singular values of $\mathbf{C}_1$ are found to be:

$$10^8 * [3.73, 0.06, 0.004, 0.00, 0.00, 0.00, 0.00, 0.00] \quad (8)$$

Observe that the the second and third signal singular values which belong to the second BPSK source and the 16QAM source, respectively, are very small compared to the first singular value, which belongs to the first BPSK signal. One reason why the singular values of the cumulant matrix are more separated than the eigenvalues of the covariance matrix is that, the computation of fourth-order cumulant estimates requires fourth powers of the data, and these increase faster than the second powers for high signal levels. Yet another reason is the difference between the fourth-order cumulants of equal-power BPSK and 16QAM signals, as mentioned in Section 1. Applying VESPA, we obtained the following angle estimates:

$$6.37^o, 6.30^o, 7.43^o \quad (9)$$

which shows that VESPA is biased towards the most powerful source.

Second, we applied IVESPA to this data, and obtained the following angle estimates:

$$6.34^o, 25.86^o, 40.59^o \quad (10)$$

It is seen that the arrival angles are estimated correctly with IVESPA.

Finally, we show that as the sample size is increased, VESPA gives accurate estimates. To show this, we simulated the same real data experiment in the computer paying particular attention to the signal conditions. We increased the sample size by 500 steps in this range, and for each sample size, we ran both VESPA and IVESPA on the simulated data for 10 realizations of the experiment. The averaged direction-of-arrival estimates obtained from VESPA and the actual values of DOAs are plotted as a function of data length in Figure 1. It is observed that for short data lengths VESPA fails to give reliable estimates; however, as the data length increases, the estimates converge to their actual values. On the other hand, IVESPA worked fine for all the values of the data length.

4. Conclusions

We presented an iterative high-resolution cumulant-based algorithm (IVESPA) for direction finding and waveform recovery. Our algorithm is based on VESPA of [1]; however, IVESPA can handle some signal scenarios for which VESPA fails to localize all the sources accurately. IVESPA is more general than VESPA in terms of applicability, but computationally more intensive. We demonstrated IVESPA by means of a real-data experiment.

5. Appendix: A procedure for estimating the arrival angle and steering vector of the most dominant source

A modified form of TLS ESPRIT [3] for one source:
Step 1: Stack $\mathbf{C}_{i1}$ and $\mathbf{C}_{i2}$ into a $2(M-i+1) \times (M-i+1)$ matrix $\mathbf{C}$ as follows:

$$\mathbf{C} \triangleq \begin{bmatrix} \mathbf{C}_{i1} \\ \mathbf{C}_{i2} \end{bmatrix} \quad (11)$$

and, perform the SVD of $\mathbf{C}$; keep the first left singular $2(M-i+1)$-vector of $\mathbf{C}$. Let this vector be $\mathbf{u}_1$.

Step 2: Partition $\mathbf{u}_1$ into two $(M-i+1)$-vectors $\mathbf{u}_{11}$ and $\mathbf{u}_{12}$.

Step 3: Perform the SVD of $[\mathbf{u}_{11}, \mathbf{u}_{12}]$; keep the last right singular vector of $[\mathbf{u}_{11}, \mathbf{u}_{12}]$. Let this 2-vector be $\mathbf{f}$.

Step 4: Partition $\mathbf{f}$ as $\mathbf{f} \triangleq \begin{bmatrix} f_x \\ f_y \end{bmatrix}$.

Step 5: An estimate of the modified steering vector of the source is obtained to within a scalar, as $\mathbf{b}_i = \mathbf{u}_{11} - \frac{f_y}{f_x}\mathbf{u}_{12}$.

6. Acknowledgement

This work was supported by the Center for Research on Applied Signal Processing at the University of Southern California.

References

[1] M. C. Dogan and J. M. Mendel, "Applications of cumulants to array processing, Part I: aperture extension and array calibration," *IEEE Trans. on Signal Processing*, vol. 42, no. 5, pp. 1200–1216, May 1995.

[2] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications," *Proc. IEEE*, vol.79, no.3, pp.278–305, March 1991.

[3] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *Optical Engineering*, vol. 29, no. 4, pp. 296–313, April, 1990

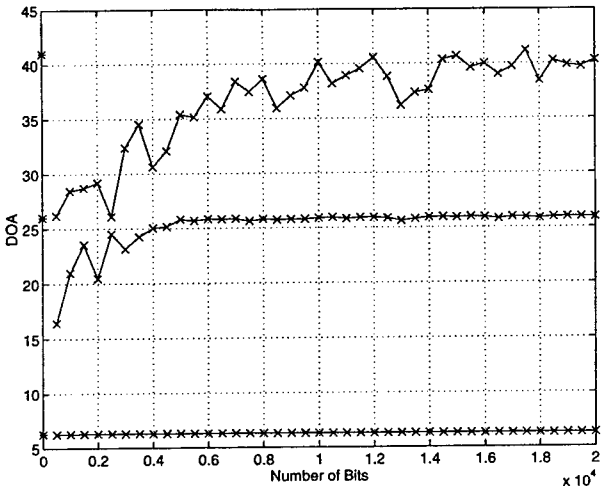


Figure 1: Average value of the direction-of-arrival estimates obtained from 10 realizations of VESPA as the data length is varied. The actual values of the directions-of-arrival are marked as "*" on the plot.

Directivity pattern measurement of moving acoustic sources

Franck Poisson and Jean Christophe Valière
Institut d'Acoustique et de Mécanique de l'Université du Maine LAUM
URA CNRS 1101 Av. O Messiaen, BP 535, 72017 LE MANS Cedex, France
Tel : (33) 43 83 32 78
poisson@laum.univ-lemans.fr, valiere@laum.univ-lemans.fr

Olivier Coste
SNCF Direction de la Recherche
45 rue de Londres, 75379 PARIS Cedex 08, France
Tel : (33) 1 40 08 99 06
coste@dr.sncf.fr

Abstract

Noise reduction of transportation is of major concern for environmental topics. As regards the railway, high speed creates new noise sources. This paper describes the last step of acoustic moving sources study. Localization methods using microphones arrays provide positions, acoustic powers, and spectrums of sources. The proposed one computes the directivity pattern of sources and gives a time-frequency representation of the emited signal while the sources pass the measurement point. Experiments are carried out to characterize acoustic sources of a high speed train (TGV) in real operating conditions.

1. Introduction

Localization techniques, using an array of microphones, provide the acoustic power, the position and the spectrum of the source. The beamforming can not be used [4] to localize high speed moving noise sources without modifications. The dedopplerisation [1] is a method currently used. Another technique [5] needs a time-frequency analysis of the output of the array. In this case, the array is focused at the end-fire. The time localization property of bilinear time-frequency representation and the spatial selectivity of the array perform the localization. The directivity pattern of moving acoustic sources is much difficult to estimate. Its computation must be performed while the source passes the measurement area.
In this paper, a method to measure the directivity pattern of a moving noise source is presented. Our interest only concerns linear trajectories and constant speed movement. Sources are supposed localized thanks to a method described above. First, the principle of the technique is described. Then, the array processing technique is presented. The choice of the bilinear time-frequency distribution is achieved. Some simulations are carried out to test the method with several directivity patterns. The directivity pattern of a moving source of a high speed train (TGV) is computed.

2. Proposed method

2.1 Principle

In order to estimate the directivity pattern of a noise source, its level is measured for several observation angles around. When the source is moving, it is difficult to turn around. On the other hand, the passing source can be observed through several angles. The array processing performs the tracking. Then, the evolution of its level may be computed by a time-frequency representation. The result is presented according to the observation angle in the source space in a polar plot. The figure 1 presents the three steps of the technique described below.

2.2 Source tracking

Firstly, an array processing focuses the moving source during its passby. The employed method is very similar to the dedopplerisation method used in the context of localization by [1]. Here, the position of the source is considered to be choosen. The Doppler effect is removed by re-building

0-8186-7576-4/96 $5.00

the emitted signal which would be received on sensor in the case of a non-moving source. The array output of N microphones is computed with the following equation :

$$S(t) = \frac{\sum_{i=1}^{M} \alpha_i P_i(t - \frac{R_i}{c})}{\sum_{i=1}^{M} \frac{\alpha_i}{R_i}} \quad (1)$$

where $P_i(t)$ is the pressure on the microphone i at time t, R_i is the distance between the microphone i and the focused source at time t, α_i is a coefficient of a weighting window and c is the sound velocity. The computation of the acoustic pressure at time $t + \frac{R_i}{c}$ needs an interpolation between two samples. The output signal $S(t)$ corresponds to the emited signal of the focused source, windowed by a spatial filter moving around it. The continuous estimation of the signal level provides the directivity pattern.

2.3 Bilinear transformation of the reconstructed signal

During the previous tracking, the Doppler effect is not perfectly removed because of the interpolations, so a frequency modulation remains. An efficient analysis tool must be able to track the signal level of the output of the array processing. Flandrin [2] has shown that some transformations belonging to the Cohen's class are well suited to track frequency shifts :

$$CO_S(t,f)\} = \int\int\int S(t' + \frac{\tau}{2}) S^*(t' - \frac{\tau}{2}) e^{-j2\pi f\tau} e^{j2\pi\eta(t'-t)} f(\eta,\tau) d\eta dt' d\tau. \quad (2)$$

where $f(\eta,\tau)$ allows to build the transformation suited to a given frequency evolution.
In a first approximation, the frequency evolves slowly according to a linear law. Among Cohen's transformations (equation 2), the Wigner-Ville (WV) transformation is optimal to follow linear frequency modulation :

$$WV_S(t,f) = \int S(t + \frac{\tau}{2}) S^*(t - \frac{\tau}{2}) e^{-j2\pi f\tau} d\tau. \quad (3)$$

Another property of this transformation, in contrast with the Fourier transform, is the conservation of the time support of the signal. This property allows to localize the apparition time of the signal with more accuracy.
In practice, the Pseudo-Smoothed Wigner-Ville (PSWV) transformation is used to reduce the interference terms due to the bilinear structure of the WV distribution.

2.4 Directivity pattern representation

The previous step of the proposed method provides a time-frequency diagram of the output of the array processing. It corresponds to the levels of the focused point at each frequency. An algorithm permits to follow the maximum level along the modulation curve around the emitted frequency. A reference of the source position is taken at the begining of the tracking. The time axis of the time-frequency diagram is converted into observation angles in the source space. The directivity pattern corresponding to the position and the frequency selected can be drawn in the source space.

3. Simulations

In order to test the described method, some simulations are carried out. A sine-wave source at frequency f, considered to be localized, is moving along a linear trajectory at constant speed v. Its directivity pattern is a $\cos\theta$ shape, like a dipole source. The received pressure on the microphone i is [3] :

$$P_i(t) = \frac{D(\theta_i)}{R_i(1 - M_a\cos(\theta_i))^2} \sin\left(2\pi f(t - \frac{R_i}{c})\right), \quad (4)$$

where

$$\cos(\theta_i) = \frac{vt}{\sqrt{D^2 + (vt)^2}},$$

and where $M_a = \frac{v}{c}$ is the Mach number, D is the distance between the trajectory and the receiver and $D(\theta_i)$ describes the source directivity.
The output of a linear array of 29 microphones spaced out 6cm and located at 6.5m away from the trajectory is computed. Figure 3 shows the directivity pattern $D(\theta)$ of the simulated source in dotted line. The tracking of the supposed source position is achieved with the equation 1. The PSWV transformation of the re-builded signal is computed and presented in figure 2. In this simulated case, the Doppler effect is perfectly removed and the constant source frequency appears. The time axis is converted into a θ axis. At time $0s$, the source is at the end-fire of the array. The measured directivity pattern drawn in the source space is presented in solid line in figure 3. This result can be compared with the directivity pattern of the simulated source presented on the same picture.
The proposed method has been tested for several directivity patterns with different shapes and directions and also for different source speeds. In all cases, it measures a directivity pattern corresponding to the simulated one.

4. Measurements

This method is applied on acoustic sources in a real situation. The previous linear array configuration is used. Its frequency range is $[2000Hz, 4000Hz]$. An experiment with a moving acoustic source for which the position and the frequency are known is carried out. A loudspeaker is fixed on a high speed train (TGV) and generates two sine-waves at

$3000Hz$ and $3400Hz$.
The figure 4 shows the result of the localization [5] of the acoustic source on the train. The position along the train is -11.5m. This area is selected to be analyzed with the proposed method. The PSWV of the tracking signal is presented on figure 5. The Doppler effect is not perfectly removed. The energy of the signal is concentrated round two frequencies corresponding to the emitted ones. The following of maximum levels along modulation curves permits to extract two directivity patterns at frequency $3000Hz$ and $3400Hz$ shown in figures 6 and 7 in the source space. The frequency evolution of the source during the tracking shows that the algorithm follows a single source.
For both frequencies, the shapes of directivity patterns are similar. The movement introduces a small rotation of the diagrams. This technique has been successfuly applied on noise sources of the train.

5. Conclusion

The proposed method is the final step of the study of moving acoustic sources. A localization technique provides for each source, the acoustic power, the position along the train, the height and the spectrum. The time-frequency representation of the source tracking permits to characterize the emited signal. If it is localized in position and frequency, the directivity pattern of the source can be computed.
The main advantage of this technique is that the measurement is performed in real operating conditions. In this case, the rotation of the directivity pattern probably due to an aeroacoustic effect can be observed. Some other effects of the movement, depending on the source speed for example, have been noticed. Then, this method improves the understanding of phenomena responsible for noise generation.

References

[1] B. BARSIKOV. On removing the doppler frequency shift from array measurements of railway noise. *Journal of Sound and Vibration*, 120(1):190–196, 1988.

[2] P. FLANDRIN. *Temps Fréquence*. HERMES, 1993.

[3] MORSE AND INGARD. *Theoretical Acoustics*. Series in electrical engineering. MC GRAW HILL BOOK COMPANY, Publishers, Inc, New York, 1982.

[4] M. PALLAS. Localization et quantification en champ proche de sources mobiles a grande vitesse. *14 -ieme colloque GRETSI Juan les Pins*, 1:1179–1182, September 13-16 1993.

[5] F. POISSON, J. VALIÈRE, P. FODIMAN, and O. COSTE. Localization of moving acoustic sources from a tgv high-speed train by a time frequency analysis. In *15 th International Congress on Acoutics*, Trondheim, Norway, June 1995. ASA.

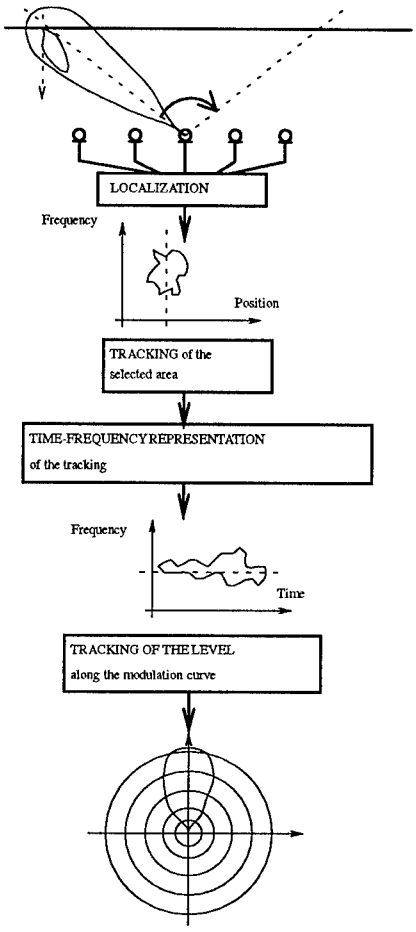


Figure 1. Diagrammatic representation of the method.

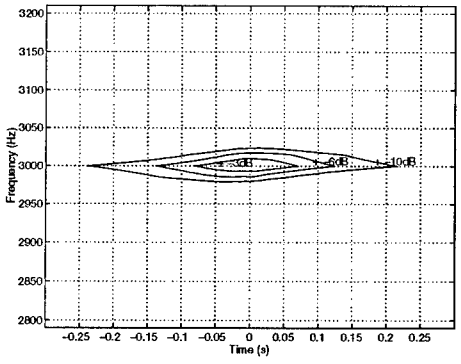


Figure 2. PSWV transformation of the signal corresponding to the tracking of the selected source.

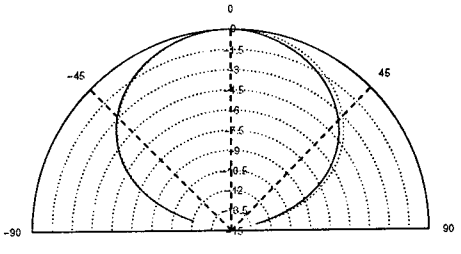


Figure 3. Directivity pattern in $\cos\theta$ (dB) of the simulated source (dotted line) and directivity pattern in dB measured by the proposed method (solid line).

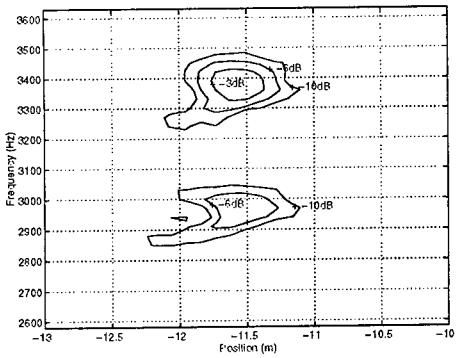


Figure 4. Localization of acoustic sources on the train.

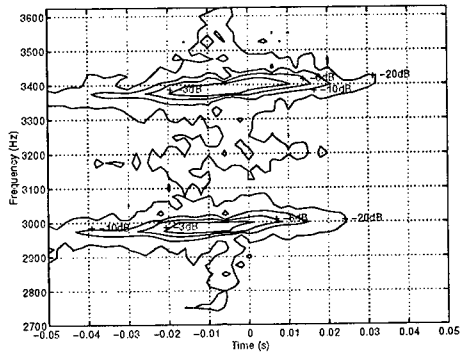


Figure 5. PSWV transformation of the tracking signal.

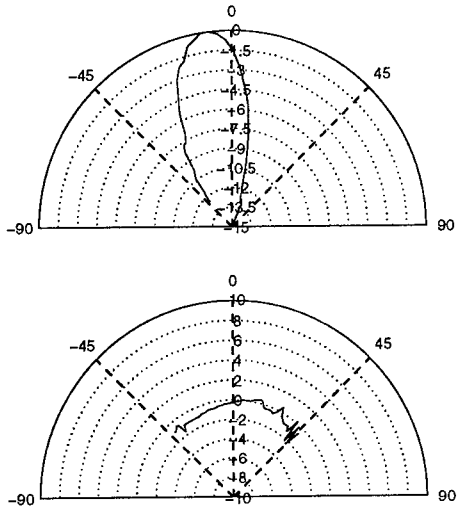


Figure 6. Directivity pattern in dB of the source at frequency 3000 Hz and frequency error in percent during the tracking of the maximum level.

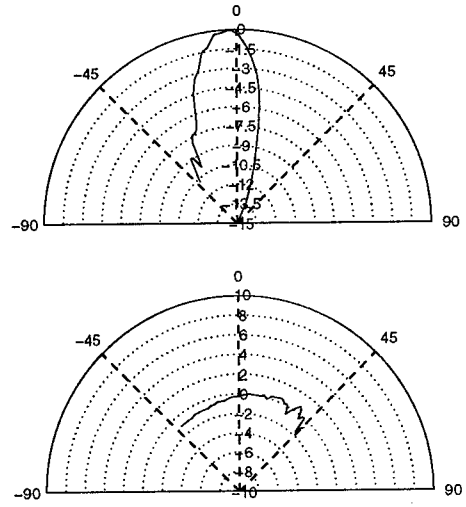


Figure 7. Directivity pattern in dB of the source at frequency 3400 Hz and frequency error in percent during the tracking of the maximum level.

Spatial and Temporal Frequency Estimation of Uncorrelated Signals Using Subspace Fitting

Bo Göransson, Magnus Jansson and Björn Ottersten
Signals, Sensors and Systems
Royal Institute of Technology (KTH)
100 44 Stockholm, Sweden
bogor@s3.kth.se

Abstract

In this paper we present a novel method for spatial and temporal frequency estimation in the case of uncorrelated sources. By imposing the diagonal structure given in the signal covariance matrix, it is possible to improve the performance of subspace based estimators. The proposed method combines ideas from subspace and covariance matching methods to yield a non-iterative frequency estimation algorithm. In a numerical example we show that the estimator has a lower small sample resolution threshold than root-MUSIC and similar large sample performance.

1. Introduction

Estimating frequencies from uniformly sampled data has been an active research area for decades. A number of, so called, high resolution algorithms or eigenstructure methods have been presented and analyzed in the literature, *e.g.*, [4, 6, 7, 8]. One disadvantage with these subspace based methods is that it is difficult to incorporate knowledge of the source correlation into the eigendecomposition. In this paper we propose an estimator which combines ideas from subspace and covariance matching methods. The objective is to find a frequency estimator which uses the knowledge of the signal correlation without significantly increasing the estimator complexity. In a numerical example we show that the proposed method has promising small sample performance.

2. Problem Formulation

The well known problem of estimating temporal or spatial frequencies from uniformly sampled data corrupted by additive white noise can be reduced to the problem of determining the parameters in the following model of the data covariance matrix

$$\mathbf{R} = \mathbf{A}(\boldsymbol{\omega})\mathbf{S}\mathbf{A}^*(\boldsymbol{\omega}) + \sigma^2\mathbf{I} \ , \tag{1}$$

where d is the number of frequencies and where $\boldsymbol{\omega} = [\omega_1, \cdots, \omega_d]^{\mathrm{T}}$. In what follows we assume that d is known, if unknown, it can be estimated from the data by techniques described in [2, 9].

In (1), the $d \times d$ matrix $\mathbf{S}$ denotes the unknown diagonal signal covariance matrix, σ^2 is the unknown noise variance and the $m \times d$ Vandermonde matrix $\mathbf{A}(\boldsymbol{\omega})$ is defined by,

$$\mathbf{A}(\boldsymbol{\omega}) = \begin{pmatrix} 1 & \cdots & 1 \\ e^{i\omega_1} & \cdots & e^{i\omega_d} \\ \vdots & & \vdots \\ e^{i(m-1)\omega_1} & \cdots & e^{i(m-1)\omega_d} \end{pmatrix} \tag{2}$$

where m is the number of sensors in the array processing case and the data window length in the temporal case.

In the spatial frequency estimation problem, the matrix $\mathbf{A}$ is often parameterized by the direction of arrivals (DOA) denoted by $\boldsymbol{\theta}$. For a linear array with uniformly spaced elements, the relationship between $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$ is given by

$$\omega_k = \frac{2\pi\Delta \sin(\theta_k)}{c} \tag{3}$$

where Δ is the element spacing and c denotes the speed of propagation of the impinging wave, and where θ_k is measured relative array broadside.

3. Frequency Estimation

The focus of this paper is on how to estimate the frequencies in the vector $\boldsymbol{\omega} = [\omega_1, \cdots, \omega_d]^{\mathrm{T}}$. In particular we would like to use the knowledge that the signals are uncorrelated without increasing the estimator complexity considerably.

0-8186-7576-4/96 $5.00

The so called subspace estimation techniques rely on the properties of the eigendecomposition of (1). Let $\{\lambda_k\}$ denote the eigenvalues of $\mathbf{R}$ arranged in descending order, *i.e.*, $\lambda_1 > \lambda_2 > \cdots > \lambda_m$. Since $\mathbf{A}$ is full rank, due to its Vandermonde structure, and $\mathbf{S}$ is positive definite, it follows that

$$\begin{aligned} &\lambda_k > \sigma^2 \qquad \text{for } k = 1, \ldots, d \\ &\lambda_{d+1} = \cdots = \lambda_m = \sigma^2 . \end{aligned} \tag{4}$$

The eigenvectors of $\mathbf{R}$ corresponding to $\{\lambda_1, \cdots, \lambda_m\}$ are denoted with $\{\mathbf{e}_1, \cdots, \mathbf{e}_m\}$. Define

$$\mathbf{E}_s = [\mathbf{e}_1, \cdots, \mathbf{e}_d], \tag{5}$$

$$\mathbf{E}_n = [\mathbf{e}_{d+1}, \cdots, \mathbf{e}_m], \tag{6}$$

$$\mathbf{\Lambda}_s = \text{diag}[\lambda_1, \cdots, \lambda_d], \tag{7}$$

$$\mathbf{\Lambda}_n = \text{diag}[\lambda_{d+1}, \cdots, \lambda_m] = \sigma^2 \mathbf{I}, \tag{8}$$

where the notation diag[·] refers to a diagonal matrix with the arguments as diagonal elements. With the notation introduced above we have

$$\mathbf{R} = \mathbf{E}_s \mathbf{\Lambda}_s \mathbf{E}_s^* + \sigma^2 \mathbf{E}_n \mathbf{E}_n^* . \tag{9}$$

Combining the two expressions for $\mathbf{R}$ in (1) and (9) yields the following equality

$$\mathbf{A}\mathbf{S}\mathbf{A}^* + \sigma^2 \mathbf{I} = \mathbf{E}_s \mathbf{\Lambda}_s \mathbf{E}_s^* + \sigma^2 \mathbf{E}_n \mathbf{E}_n^* . \tag{10}$$

Since $\mathbf{E}_n \mathbf{E}_n^* = \mathbf{I} - \mathbf{E}_s \mathbf{E}_s^*$, it follows that

$$\mathbf{A}\mathbf{S}\mathbf{A}^* = \mathbf{E}_s \mathbf{\Lambda} \mathbf{E}_s^* , \tag{11}$$

where $\mathbf{\Lambda} = \mathbf{\Lambda}_s - \sigma^2 \mathbf{I}$. By using the vec-operator ($\mathbf{d} = \text{vec}(\mathbf{D})$ is a vector obtained by stacking the columns of $\mathbf{D}$) (11) can be written as $\left(\text{vec}(\mathbf{XYZ}) = \left(\mathbf{Z}^{\mathrm{T}} \otimes \mathbf{X}\right) \text{vec}(\mathbf{Y})\right)$

$$(\mathbf{A}^{\mathrm{c}} \otimes \mathbf{A})\, \text{vec}(\mathbf{S}) = (\mathbf{E}_s^{\mathrm{c}} \otimes \mathbf{E}_s)\, \text{vec}(\mathbf{\Lambda}) \tag{12}$$

where $\otimes$ denotes the Kronecker matrix product and where $(\cdot)^{\mathrm{c}}$ denotes complex conjugation. Since $\mathbf{S}$ and $\mathbf{\Lambda}$ are diagonal matrices with real-valued entries, there exists a $(d^2 \times d)$ selection matrix $\mathbf{L}$ such that

$$\text{vec}(\mathbf{S}) = \mathbf{L}\mathbf{s}, \qquad \text{vec}(\mathbf{\Lambda}) = \mathbf{L}\boldsymbol{\lambda}, \tag{13}$$

where $\mathbf{s}$ and $\boldsymbol{\lambda}$ are vectors consisting of the diagonal entries of $\mathbf{S}$ and $\mathbf{\Lambda}$, respectively.

Let $\hat{\mathbf{R}}$ denote a sample estimate of the theoretical covariance matrix, and let $\hat{\mathbf{E}}_s$ be the estimated "signal subspace" obtained from an eigendecomposition of $\hat{\mathbf{R}}$ similar to (9). Replacing $\mathbf{E}_s$ and $\mathbf{\Lambda}$ with its estimates in (12) we have

$$\mathbf{B}(\boldsymbol{\omega})\mathbf{s} \approx \hat{\mathbf{F}}\hat{\boldsymbol{\lambda}}, \tag{14}$$

where $\mathbf{B}(\boldsymbol{\omega}) = (\mathbf{A}^{\mathrm{c}} \otimes \mathbf{A})\,\mathbf{L}$ and $\hat{\mathbf{F}} = \left(\hat{\mathbf{E}}_s^{\mathrm{c}} \otimes \hat{\mathbf{E}}_s\right)\mathbf{L}$. We now suggest to estimate the unknown parameters by minimizing the following least squares criterion

$$\|\mathbf{B}(\boldsymbol{\omega})\mathbf{s} - \hat{\mathbf{F}}\hat{\boldsymbol{\lambda}}\|^2 . \tag{15}$$

Minimizing with respect to $\mathbf{s}$ yields

$$\hat{\mathbf{s}} = \mathbf{B}^{\dagger}(\boldsymbol{\omega})\hat{\mathbf{F}}\hat{\boldsymbol{\lambda}}, \tag{16}$$

where $\mathbf{B}^{\dagger}$ denotes the pseudo-inverse of $\mathbf{B}$. Substituting this back into the criterion we arrive at the following criterion for finding the frequency estimates

$$\min_{\boldsymbol{\omega}} \|\mathbf{P}^{\perp}_{\mathbf{B}(\boldsymbol{\omega})}\hat{\mathbf{F}}\hat{\boldsymbol{\lambda}}\|^2 , \tag{17}$$

where $\mathbf{P}^{\perp}_{\mathbf{B}} = \mathbf{I} - \mathbf{B}\mathbf{B}^{\dagger}$ is the orthogonal projector onto the null space of $\mathbf{B}^*$. The criterion (17) is in general multimodal, rendering the multidimensional search for a global extremum computationally expensive. In the following we will use the ideas in [1, 5, 6] to rewrite the minimization in (17) in a computationally much more attractive form. From the definition of $\mathbf{B}(\boldsymbol{\omega})$ it follows that the k^{th} column of $\mathbf{B}$ is given by

$$\begin{aligned} \mathbf{B}_k = [1\ z_k \cdots z_k^{m-1} &\vdots z_k^{-1}\ 1 \cdots z_k^{m-2} \vdots z_k^{-2} \cdots z_k^{m-3} \vdots \cdots \\ &\cdots \vdots z_k^{-(m-1)} \cdots 1]^{\mathrm{T}} \end{aligned} \tag{18}$$

where $z_k = \mathrm{e}^{i\omega_k}$. Observing the shift structure in (18) it is possible to parameterize the nullspace of $\mathbf{B}^*$ by the coefficients in the following polynomial

$$g_0 z^d + g_1 z^{d-1} + \cdots + g_d = g_0 \prod_{k=1}^{d} \left(z - \mathrm{e}^{i\omega_k}\right) \tag{19}$$
$$g_0 \neq 0 .$$

Define a full rank matrix $\mathbf{G}$ of dimension $m^2 \times (m^2 - d)$, which depends linearly on the coefficients in (19), such that

$$\mathbf{G}^*\mathbf{B} = 0 . \tag{20}$$

This implies that the columns of $\mathbf{G}$ constitutes a basis for the nullspace of $\mathbf{B}^*$ and

$$\mathbf{P}_{\mathbf{G}} = \mathbf{G}(\mathbf{G}^*\mathbf{G})^{-1}\mathbf{G}^* = \mathbf{P}^{\perp}_{\mathbf{B}} . \tag{21}$$

In order to illustrate the parameterization a simple example is provided.

Example: Assume $m = 2$ and $d = 1$, which implies that the matrix $\mathbf{B}$ consists of one column only. The polynomial (19) will in this case be given by

$$g_0 z + g_1 = 0 . \tag{22}$$

Using (22) and the shift structure in (18), we can write (20) as

$$\mathbf{G}^*\mathbf{B} = \begin{bmatrix} g_1 & g_0 & 0 & 0 \\ 0 & g_0 & 0 & g_1 \\ 0 & 0 & g_1 & g_0 \end{bmatrix} \begin{bmatrix} 1 \\ z \\ z^{-1} \\ 1 \end{bmatrix} = 0$$

By using the parameterization described above it follows that the criterion (17) can be reformulated as

$$\min \|(\mathbf{G}^*\mathbf{G})^{-1/2}\mathbf{G}^*\hat{\mathbf{F}}\hat{\boldsymbol{\lambda}}\|^2 \ , \qquad (23)$$

where the minimization is over the polynomial coefficients in (19). A two-step estimation procedure can now be devised as follows:

1. Obtain a consistent estimate of $\{g_k\}$ by minimizing the quadratic function obtained by replacing $(\mathbf{G}^*\mathbf{G})^{-1/2}$ in (23) by some positive definite matrix $\mathbf{W}$.

2. Use the estimate of $\{g_k\}$ from step 1 to construct a consistent estimate of $(\mathbf{G}^*\mathbf{G})^{-1/2}$. Insert this in (23) and solve a new quadratic problem. The frequency estimates are then given by rooting the polynomial (19).

It can be shown that this two-step procedure has the same large sample accuracy as the estimates obtained by minimizing (17). The main advantage is that we avoid the non-linear parameter search. For small sample scenarios it can be useful to reiterate step 2 a few times to improve the accuracy.

4. Numerical Example

Here a numerical example is provided to demonstrate the performance of the proposed method. We consider the direction of arrival estimation of two waves impinging from angles $\theta_1 = 10°$ and $\theta_2 = 20°$ on a ULA with 5 elements separated by a half wavelength. The uncorrelated signal sources are modeled as white and complex Gaussian distributed, each with SNR = 3dB. The MSE errors for different data lengths are calculated for the proposed method and for root-MUSIC [3], each MSE are based on 200 independent trials. The MSE for θ_1 is depicted in Fig. 1. This example demonstrates the superior performance for small sample scenarios compared to root-MUSIC.

5. Conclusions

The main idea in this paper was to present a novel method for spatial and temporal frequency estimation in the case of uncorrelated sources. By imposing the diagonal structure given in the signal covariance matrix, it is possible to improve the performance of subspaced based estimators.

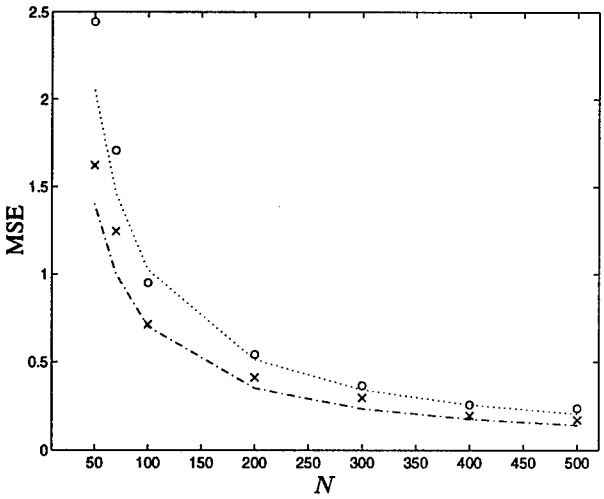


Figure 1. MSE values for θ_1 versus the number of snapshots, N: 'x' – proposed method, 'o' – root-MUSIC. The dash-dotted line represents the CRB when the correlation structure of the sources is known and the dotted line is the CRB without this knowledge.

References

[1] Y. Bresler and A. Macovski. Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-34(5):1081–1089, October 1986.

[2] J.-J. Fuchs. Estimating the Number of Sinusoids in Additive White Noise. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-36(12):1846–1853, December 1988.

[3] B. D. Rao and K. V. S. Hari. Performance analysis of Root-Music. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-37(12):1939–1949, December 1989.

[4] P. Stoica and A. Nehorai. MUSIC, Maximum Likelihood, and Cramér-Rao Bound. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-37(5), May 1989.

[5] P. Stoica and K. C. Sharman. Maximum Likelihood Methods for Direction-of-Arrival Estimation. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-38(7):1132–1143, July 1990.

[6] P. Stoica and K. C. Sharman. Novel Eigenanalysis Method for Direction Estimation. *Proc. IEE*, F(1):19–26, February 1990.

[7] P. Stoica and T. Söderström. Statistical Analysis of MUSIC and Subspace Rotation Estimates of Sinusoidal Frequencies. *IEEE Trans. on Signal Processing*, SP-39(8):1836–1847, August 1991.

[8] M. Viberg and B. Ottersten. Sensor Array Processing Based on Subspace Fitting. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-39(5):1110–1121, May 1991.

[9] M. Wax and T. Kailath. Detection of Signals by Information Theoretic Criteria. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-33(2):387–392, April 1985.

Localization of the Wideband Sources with Estimation of the Antenna Shape

S. Bourennane and M. Frikel

CMCS CNRS-URA 2053, Quartier Grossetti, B.P. 52, 20250 Corte - France
e-mail : bourenna@univ-corse.fr frikel@univ-corse.fr

Abstract

The purpose of this paper is the passive angular location of the wideband sources using an array of sensors. The interest of the knowledge of the antenna shape, when the treatment is applied on the received data, is illustrated by the improvement of the signal to noise ratio and by the increase in of the antenna directivity.

In this paper,the extension of the propagator method is presented: an algebraic operator is extracted from the cross-spectral matrices of the data or from the received signals. This technic avoids the rather expensive eigendecomposition of cross-spectral matrices at each frequency of the analysis bandwidth used in the known methods. The results of simulations support the theoretical predictions.

1. Introduction

The estimation of direction of arrival (DOA) of multiple wideband signals is a recent problem in array signal processing. Many techniques have been reported in the literature [1-3], of which eigenstructure methods are among the most established. The concept of the signal subspace processing have been used in the wideband case [1]. The basic idea is to use a coherent signal subspace estimate obtained by the eigendecomposition of a frequency domain combination of modified narrow-band cross-spectral matrix estimates. It is shown that the coherent subspace method is an alternative to incoherent subspace method that improves the efficiency of the estimation by focusing the energy of the analysis bandwidth into a focusing frequency. Similar technic have been proposed in [2], the originality of this method is the construction of the focusing operators; which used to transform the signal subspaces. Generally, these methods have better performances than the classical methods but their rather expensive computational load limits their implementation. To avoid this difficult, several papers [4-9] have been published in the aim to reduce the computational load for the eigendecomposition or to estimate the signal subspace without eigendecomposition.The propagator method [4-6] is one of these methods which is considered as an alternative of the MUSIC method.

In this paper, we introduce an extension of the propagator method for broadband sources. The transformation of the incoherent propagators is performed through focusing matrices. The obtained coherent propagator is used to estimate the antenna shape and the DOA of the sources.

2. Problem formulation

We consider an array of N sensors which received the wavefield generated by P wideband sources in the presence of an additive noise. The array geometry is arbitrary. The received signal vector, in the frequency domain, is given by :

$\mathbf{r}(f_j) = \mathbf{A}(f_j)\mathbf{s}(f_j) + \mathbf{n}(f_j)$

Where $\mathbf{r}(f_j)$ is the Fourier transform of the array ouput vector, $\mathbf{s}(f_j)$ is a source vector, $\mathbf{n}(f_j)$ is a sensor noise, and $\mathbf{A}(f_j)$ is the $N \times P$ transfer matrix of the source-sensor array systems with respect to some chosen reference point.

It is assumed that the array is unambiguous and calibrated, so that the rank of $\mathbf{A}(f_j)$ is equal to P for any frequency. The sensor noise is assumed to be independent of the source signals and spatially white or the cross-spectral matrix is known but for a scale factor. In this case, a prewhitening step is required to create diagonal noise cross-spectral matrix. The sources are not fully correlated.

The cross-spectral matrix of the observation vector at frequency f_j is given by :

$\mathbf{\Gamma}(f_j) = \mathbf{A}(f_j)\mathbf{\Gamma}_s(f_j)\mathbf{A}^+(f_j) + \sigma^2(f_j)\mathbf{I}$

Where the superscript $^+$ represents the Hermitian transpose. $\mathbf{\Gamma}_s(f_j)$ is the source cross-spectral matrix.

Our aim is to estimate the angles $\theta_i, i = 1, ..., P$ and the antenna shape from the received data. In this paper, the detection of the sources is not treated. We assume that the number of the sources P is known or can be estimated [10].

For locating the wideband sources several solutions have been proposed in the literature and are summarized

0-8186-7576-4/96 $5.00

as following :
-The incoherent subspace methods: the analysis bandwidth is divided into several frequency bins and then at each frequency the treatment is applied and the obtained results are combined to obtain the final result.
- The coherent subspace methods: the different subspaces are transformed in a predefined subsapce using the focusing matrices.
For estimating the antenna shape, the existing methods treat the narrowband case [5-11]. The temporal methods have been proposed for the wideband signals but they have not any success, because they have low spatial resolution.

3. Narrow-band propagator method

In this section, we recall briefly the propagtor method. We consider the no noisy situation, e.d.:
$\mathbf{r}(f_j) = \mathbf{A}(f_j)\mathbf{s}(f_j)$.
The direction of the sources matrix is partitioned [4-6] in two block matrices, let :

$$\mathbf{A}(f_j) = \begin{bmatrix} \mathbf{X}(f_j) \\ - - - \\ \mathbf{Y}(f_j) \end{bmatrix}$$

Where $\mathbf{X}(f_j)$ is a $P \times P$ matrix and $\mathbf{Y}(f_j)$ is a $(N - P) \times P$ matrix. We assume that the model propagation vector is such that $\mathbf{X}(f_j)$ is nonsingular for example the P first sensors are linear and equispaced then $\mathbf{X}(f_j)$ is a Vandermonde matrix.
The (N-P) last rows of $\mathbf{A}(f_j)$ are linearly dependent of the P first rows, we can write $\mathbf{Y}(f_j)$ as :

$\mathbf{Y}(f_j) = \mathbf{\Pi}^+(f_j)\mathbf{X}(f_j)$
or $\mathbf{\Pi}^+(f_j) = \mathbf{Y}(f_j)\mathbf{X}^-1(f_j)$
The $P \times (N - P)$ matrix, $\mathbf{\Pi}(f_j)$, is called the propagator [4-6].
We define the matrix $\mathbf{Q}(f_j)$ as :
$\mathbf{Q}^+(f_j) = [\mathbf{\Pi}^+(f_j) \mid -\mathbf{I}]$
It is easy to see that :

$\mathbf{Q}^+(f_j)\mathbf{A}(f_j) = \mathbf{\Pi}^+(f_j)\mathbf{X}(f_j) - \mathbf{Y}(f_j) = 0$
or $\mathbf{Q}^+(f_j)\mathbf{a}_p(f_j) = 0$ for $p = 1, ..., P$ and $j = 1, ..., M$
The construction of $\mathbf{Q}(f_j)$ needs the knowledge of the directions of the sources and the geometry of the antenna. For this we can not use directly the previous result. With the former partition the cross-spectral matrix is :

$$\mathbf{\Gamma}(f_j) = \begin{bmatrix} \mathbf{\Gamma_{11}}(f_j) & \mathbf{\Gamma_{11}}(f_j)\mathbf{\Pi}(f_j) \\ \mathbf{\Pi}^+(f_j)\mathbf{\Gamma_{11}}(f_j) & \mathbf{\Pi}^+(f_j)\mathbf{\Gamma_{11}}(f_j)\mathbf{\Pi}^+(f_j) \end{bmatrix}$$

$$\mathbf{\Gamma}(f_j) = \begin{bmatrix} \mathbf{\Gamma_{11}}(f_j) & \mathbf{\Gamma_{12}}(f_j) \\ \mathbf{\Gamma_{21}}(f_j) & \mathbf{\Gamma_{22}}(f_j) \end{bmatrix}$$

$\mathbf{\Gamma_{11}}(f_j) = \mathbf{X}(f_j)\mathbf{\Gamma_S}(f_j)\mathbf{X}^+(f_j)$
We have for example :

$\mathbf{\Gamma_{12}}(f_j) = \mathbf{\Gamma_{11}}(f_j)\mathbf{\Pi}(f_j)$

it follows, the estimate of the propagator is:
$\mathbf{\Pi}(f_j) = \mathbf{\Gamma_{11}}^{-1}(f_j)\mathbf{\Gamma_{12}}(f_j)$
Other partitions of the cross-spectral matrix are given in [4-6]. We note that the partition in [4-6] can be lead to the computational complexity important for $P << N$.
In a noisy situation, the optimal propagator is obtained by the constrained minimization problem

$$\min_{\mathbf{\Pi}(f_j)} \left\| \hat{\mathbf{\Gamma}}_{\mathbf{12}}(f_j) - \hat{\mathbf{\Gamma}}_{\mathbf{11}}(f_j)\mathbf{\Pi}(f_j) \right\| \text{ for } j = 1, ..., M.$$

They used the Frobenius matrix norm.

4. Extension of the propagator method to the wideband signals

4.1. Incoherent propagator

The analysis bandwidth is divided into M frequencies. The narrowband propagator is applied at each frequency bin. The final result is obtained by averaging the different results. The directions of the sources are estimated by plotting, as function of θ, the following measure :

$$J(\theta) = \frac{1}{M}\sum_{j=1}^{M} \frac{1}{\mid \mathbf{Q}^+(f_j)\mathbf{a}(\theta, f_j) \mid}$$

4.2. Coherent propagator

The transformation matrices are used at each frequency bin such that we obtain the focused propagtor at the center frequency and, then all the transformed propagators are coherently averaged to obtain the coherent propagator, i.e
$\mathbf{D}(f_j)\mathbf{\Pi}(f_j) = \hat{\mathbf{\Pi}}(f_o)$
Where $\mathbf{D}(f_j)$ is the focusing matrix, and $\hat{\mathbf{\Pi}}(f_o)$ is given by:

$$\hat{\mathbf{\Pi}}(f_o) = \left[\hat{\mathbf{X}}(f_o)\tilde{\mathbf{\Gamma}}_{\mathbf{S}}(f_o)\hat{\mathbf{X}}^+(f_o)\right]^{-1} \mathbf{\Gamma_{12}}(f_o)$$

with $\tilde{\mathbf{\Gamma}}_{\mathbf{S}}(f_o) = \frac{1}{M}\sum_{j=1}^{M} \hat{\mathbf{X}}^{-1}(f_j)\mathbf{\Gamma_{11}}(f_j)\left(\hat{\mathbf{X}}^+(f_j)\right)^{-1}$
$\hat{\mathbf{X}}(f_j)$ is constructed by using an initial directions of the sources. $\mathbf{\Gamma_{12}}(f_o)$ is a block matrix of the cross-spectral matrix at the focusing frequency f_o.
The transformation matrix is given by :
$\mathbf{D}(f_j) = \hat{\mathbf{\Pi}}(f_o)\mathbf{\Gamma}^+_{\mathbf{12}}(f_j)\left[\mathbf{\Gamma_{12}}(f_j)\mathbf{\Gamma}^+_{\mathbf{12}}(f_j)\right]^{-1}\hat{\mathbf{\Gamma}}_{\mathbf{11}}(f_j)$
Using these transformation matrices, all the propagator can be combined to find the focused propagator, in the following manner,

$\tilde{\mathbf{\Pi}}(f_o) = \frac{1}{M}\sum_{j=1}^{M} \mathbf{D}(f_j)\mathbf{\Pi}(f_j)$

The obtained propagator is, now, used to construct the coherent matrix, given by :

$\tilde{\mathbf{Q}}^+(f_o) = \left[\tilde{\mathbf{\Pi}}^+(f_o) \mid -\mathbf{I}\right]$

We have : $\tilde{\mathbf{Q}}^{+}(f_o)\mathbf{a_p}(f_o) = 0$, for $p = 1, ..., P$, using this result, the directions of the sources are given by the values of θ for which the function $J(\theta)$ is maximized,

$$J(\theta) = \frac{1}{\mid \tilde{\mathbf{Q}}^{+}(f_o)\mathbf{a}(f_o, \theta) \mid} \qquad for \qquad \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

5. Algorithm for estimating the antenna shape

The source vectors contain $2(N-1)$ unknown parameters corresponding to $(N-1)$ modulus and $(N-1)$ phases; however, there are only $2(N-1)$ free parameters per source, which permit to treat only one source, to overcome this difficult two-step algorithm are used :
- In the first step the modulus are eliminated by using the coherency matrix of the received data, and then by using the conjugate gradient algorithm the phase distribution along the antenna is estimated.
- In the second step the phase estimates obtained in the first step are introduced in the cross-spectral matrix and then modulus can be estimated.

Note that for seperating different contributions of the phase estimates to obtain the antenna shape, an algorithm such that multidimensional Wiener filter can be used.

6. Simulations results

In the following simulations, an antenna of $N = 20$ equispaced sensors with an arbitrary distorsion compared to a linear antenna. The source signals are temporally stationary zero-mean bandpass white Gaussian processes with the same bandwidth $[100, 131Hz]$. Two source signals impinge on the array at $\theta_1 = 10°$ and $\theta_2 = 12°$ respectively, with $SNR = 3dB$. The analysis bandwidth is decomposed into $M = 32$ narrowband components via FFT.

Fig. 1 gives the arithmetic mean of the obtained results with the incoherent propagators method with $P = 2$.
Fig. 2 shows the direction of arrival of the sources with the coherent propagator described above. We remark that, in the two cases, the sources are not perfectely localized, however the coherent method gives a better results that the incoherent method.
Fig. 3 shows the estimation of the antenna shape using the proposed method.

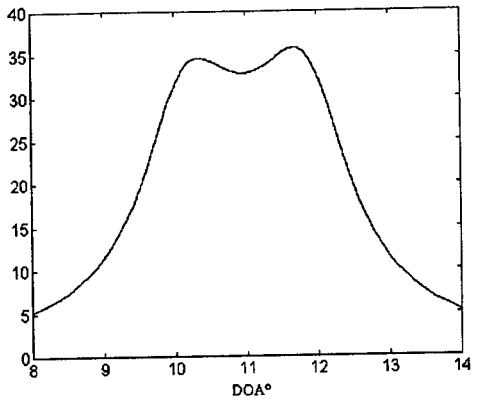


Figure 1 : DOA of the sources using incoherent method without antenna correction.

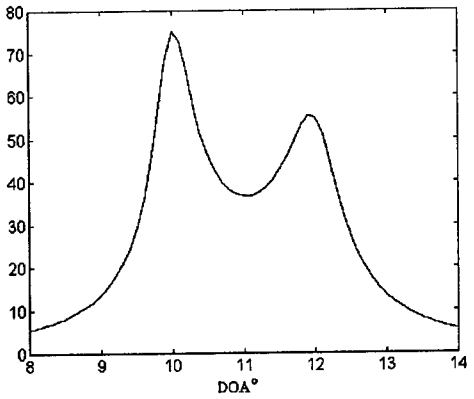


Figure 2 : DOA of the sources using coherent method without antenna correction.

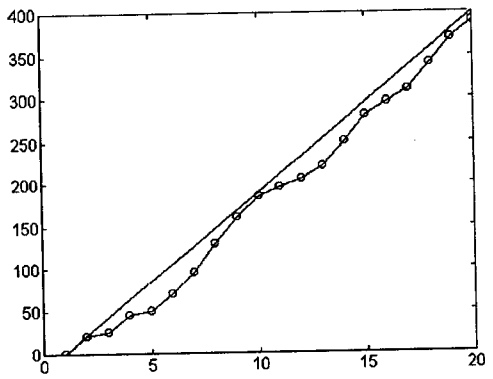


Figure 3 : Estimation of the antenna shape.

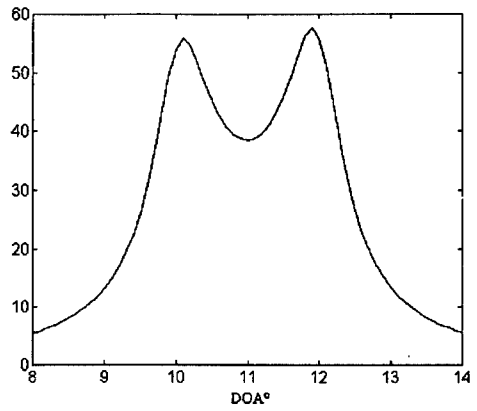


Figure 4 : DOA of the sources using incoherent method after antenna correction.

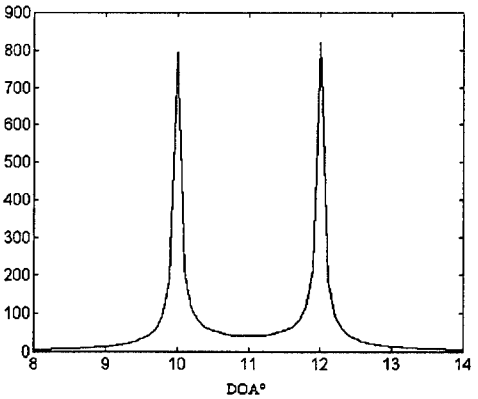


Figure 5 : DOA of the sources using coherent method after antenna correction.

Fig. 4 and Fig. 5 present the bearing estimation of the sources after the compensation of the phase due to the sensors displaced. these results show that after the antenna correction, the sources are exactely localized, however the coherent propagator is efficient compared to the incoherent treatment.
This numerical example shows the interest of the estimation of the antenna shape.

7. Conclusion

In this paper, we have extended the propagator method to the localization of the wideband signals using an array of sensors. This method avoids the eigendecomposition of the cross-spectral matrix. It is based on the transformation of the narrowband propagators. We have shown that the knowledge of the antenna shape permits the compensation of the fluctuations of the phase along the antenna which improves the localization.

References

[1] H. Wang and M. Kaveh, Coherent Signal-Subspace Processing for the detection and estimation of angles of arrival of multiple wideband sources, IEEE Trans. ASSP, Vol. 33, n° 4, Aug. 1985.

[2] S. Valaee and P. Kabal, Wideband array processing using a two-sided correlation transformation, IEEE trans. on sig. Proc. Vol. 43, n° 1, Janvier 1995.

[3] S. Bourennane, B. Faure and J. L. Lacoume, Traitement d'antenne pour des sources large bande, Ann. Télécom., 1990.

[4] J. Munier and G.Y. Delisle, Spatial analysis using new properties of the cross-spectral matrix, IEEE Trans. Signal Proc., vol. 39, March 1991.

[5] J. Munier, G. Jourdain and G.Y. Delisle, A new algorithm for the identification of distorted wavefronts, Underwater Acoustic Data Processing, NATO Advanced study Institute, Kingston, Ontario, Canada 1988.

[6] S.Marcos, A. Marsal and M. Benidir, The propagator method for source bearing estimation, Signal Processing, vol. 42, n° 2, March 1995.

[7] G. Xu, T. Kailath, Fast subspace decomposition, IEEE Trans. SP, vol. 42, n° 3, March 1994.

[8] V. T. Ermolaev and A. B. Gershman, Fast algorithm for minimum norm direction-of-arrival estimation, IEEE Trans. on Sig. Proc., vol. 42, n° 9, Sep. 1994.

[9] P. Stoica and T. Söderström, On Di's and subspace fitting approaches to direction estimation, IEEE trans. on Sig. proc., vol.42, n° 3, March 1994.

[10] M. Frikel and S.Bourennane, Detection method using eigenvalues, Acta Acustica, vol. 82, Jan.-Feb. 1996.

[11] B. Faure, S. Bourennane, and J. L. Lacoume, Analyse temporelle et spatiale de l'onde reçue sur une grande antenne à longue distance en ASM, Revue Traitement du Signal, 1987.

[12] J. Munier, Identification de fronts d'ondes corrélées et distordus, Revue Traitement du signal, vol. 4, n° 4, 1987.

[13] C. E. Davila, M. Azmoodeh, Efficient estimation of the signal subspace without eigendecomposition, IEEE Trans. on Sig. Proc., 1994.

3-D SPHERICAL LOCALIZATION OF MULTIPLE NON-GAUSSIAN SOURCES USING CUMULANTS

Raghu N. Challa *and* Sanyogita Shamsunder
Dept. of Elec. Engr., Colorado State University, Fort Collins,CO 80523

1. ABSTRACT

In this paper an eigen decomposition technique based on cumulant matrices is proposed to passively localize narrowband non-Gaussian sources in the spherical coordinates viz., azimuth, elevation, range, using signals recorded by a centro-symmetric cross array. The multiple degrees of freedom available from cumulants are exploited to transform the near-field data into pseudo-data collected by a virtual rectangular array observing virtual far-field sources. The centro-symmetric array structure is preserved in the pseudo-data thus allowing efficient real-valued processing via Unitary ESPRIT.

2. INTRODUCTION

In recent years, several eigen decomposition algorithms have been proposed for passive source localization using sensor arrays. However, most approaches operate under the assumption of far-field sources and consequently can only estimate the azimuth (1-D) or the azimuth and elevation (2-D) using passive sensing (see for e.g. [4]-[7]) and are based on the planar wave-front approximation. In many a situation, sources are close to the array and the inherent curvature of the waveforms can no longer be neglected. Recent works on near-field source localization concentrated on estimating the azimuth and range only. The algorithms in [2, 3] either involved multiple 1-D searches of a 2-D MUSIC cost function or Wigner-Ville distributions and provided poor resolution, while in [1] the invariant properties of cumulant matrices were used for range and azimuth estimation. None of the existing works address passive 3-D localization which involves the estimation of the spherical coordinates, namely *azimuth, elevation*, and *range*. This paper proposes a 3-D localization algorithm, which employs cross-cumulants of signals recorded by a 2-D cross array.

Consider a near-field scenario in which co-channel, narrowband signals from L sources located at azimuth, elevation, and range given by the vector $[\alpha_l, \theta_l, r_l]$, impinge upon a cross-array aligned with the X and Y axes (Figure 1). Although for simplicity, it is assumed that each of these branches consists of uniformly spaced omnidirectional sensors (with spacing d), the algorithm is applicable even when the sensor responses are not identical, as long as the array is conjugate centro-symmetric (see e.g. [5], for a description of centro-symmetric arrays). With inter-sensor spacing d and $m, n \in \{-M-1, \ldots, -1, 0, 1, \ldots, M\}$, the output of the sensor located at coordinates (md, nd) is :

$$u_{m,n}(t) = \sum_{l=1}^{L} s_l(t) e^{j\tau_{m,n}(l)} + v_{m,n}(t), \qquad (1)$$

where, $\tau_{m,n}(l) \triangleq [\omega_{xl} m + \phi_{xl} m^2 + \omega_{yl} n + \phi_{yl} n^2 + \beta_l mn]$ is the phase difference between the lth source signal at sensor $\{m, n\}$ and that recorded by the sensor located at $\{0, 0\}$. The parameters $\{\omega_{xl}, \phi_{xl}, \omega_{yl}, \phi_{yl}, \beta_l\}$ are nonlinear functions of $[\alpha_l, \theta_l, r_l]$. With $\beta_l = -\frac{\pi d^2}{\lambda r_l} \sin^2\theta_l \sin 2\alpha_l$:

$$\begin{aligned} &\omega_{xl} = -\tfrac{2\pi d}{\lambda} \sin\theta_l \cos\alpha_l, \quad \omega_{yl} = -\tfrac{2\pi d}{\lambda} \sin\theta_l \sin\alpha_l, \\ &\phi_{xl} = \tfrac{\pi d^2}{\lambda r_l} \left(1 - \sin^2\theta_l \cos^2\alpha_l\right), \quad \text{and} \\ &\phi_{yl} = \tfrac{\pi d^2}{\lambda r_l} \left(1 - \sin^2\theta_l \sin^2\alpha_l\right). \end{aligned} \qquad (2)$$

The narrowband signals $s_l(t)$ are zero-mean, stationary, mutually independent, with non-zero fourth-order cumulants, while the sensor noise $v_{m,n}(t)$ is modeled as zero-mean, Gaussian and independent of $s_l(t)$. Localization involves the estimation of $[\alpha_l, \theta_l, r_l]$, given the observations $\{u_{m,n}(t)\}$ for $t \in \{0, \ldots, T-1\}$. The parameter vectors $\boldsymbol{\omega}_x \triangleq [\omega_{x1}, \ldots, \omega_{xL}]'$, $\boldsymbol{\phi}_x \triangleq [\phi_{x1}, \ldots, \phi_{xL}]'$, $\boldsymbol{\omega}_y \triangleq [\omega_{y1}, \ldots, \omega_{yL}]'$, and $\boldsymbol{\phi}_y \triangleq [\phi_{y1}, \ldots, \phi_{yL}]'$ are first estimated via subspace methods and then paired to yield the locations.

3. ALGORITHM DESCRIPTION

The proposed algorithm is implemented in two steps. Estimation of $\{\boldsymbol{\omega}_x, \boldsymbol{\phi}_x\}$ $(\{\boldsymbol{\omega}_y, \boldsymbol{\phi}_y\})$ using the signals from X (Y) subarray is considered in the first step. In Step 2, the elements of $\{\boldsymbol{\omega}_x, \boldsymbol{\phi}_x\}$ are

0-8186-7576-4/96 $5.00

paired with those in $\{\omega_y, \phi_y\}$. This last step is essential to obtain the final source spherical coordinates $[\alpha_l, \theta_l, r_l]$.

Step 1 : Estimation of ω_x, ϕ_x, ω_y, and ϕ_y

From (1), the signals collected by the sensors along the X subarray $u_{m,0}(t)$ are :

$$u_{m,0}(t) = \sum_{l=1}^{L} s_l(t) e^{j[\omega_{xl} m + \phi_{xl} m^2]} + v_{m,0}(t). \quad (3)$$

The model in (3) coincides with the signal model corresponding to azimuth and range only (2-D) estimation of near-field sources with a 1-D array [1].

In order to gain insight into the possible source-sensor configuration which gives rise to the signal model in (3), consider a 2-D plane $ABCD$ containing the X subarray and the lth source as shown in the Figure 2. From the figure we see that, as far as the X subarray is concerned, source l is at a distance r_l from the array center, at an angle $\mu_l = \sin^{-1}(\sin\theta_l \cos\alpha_l)$ with the perpendicular AB, passing through the array center B and lying in the plane $ABCD$. In [1], a HOS based solution that yields paired estimates of r_l and μ_l was proposed. The same approach can be applied to the model in (3) to obtain ω_x and ϕ_x.

However, here we take a different view point and propose a solution which exploits the centro-symmetry of fourth-order cross-cumulant matrices. This new algorithm is based on the observation that as a result of the nonlinear operations involved in cumulant computation, the data collected from a 1-D linear array (the X subarray) observing near-field sources can be transformed into pseudo-data collected from a virtual rectangular array observing virtual far-field sources. The azimuth and elevation of these virtual far-field sources will turn out to be functions of $\{\omega_{xl}, \phi_{xl}\}$ in the original data.

3.1 Transformation of Data

Under the model assumptions, using (3), and cumulant properties, the fourth-order cross-cumulant of the signals from sensors at $\{-m, 0\}$, $\{m, 0\}$, $\{n-1, 0\}$, $\{n, 0\}$ simplify to :

$$c_{4m,n}(\tau) \triangleq \text{cum}\{u^*_{-m,0}(t+\tau), u_{m,0}(t), u^*_{n-1,0}(t), u_{n,0}(t)\} = \sum_{l=1}^{L} c_{4s_l}(\tau) e^{j[2\omega_{xl} m + 2\phi_{xl} n + (\omega_{xl} - \phi_{xl})]}, \quad (4)$$

where $c_{4s_l}(\tau) \triangleq \text{cum}\{s^*_l(t+\tau), s_l(t), s^*_l(t), s_l(t)\}$ is the fourth-order cumulant of the source signal $s_l(t)$. Notice that the cumulants of the noise term $v_{m,0}(t)$ do not appear owing to the fact that cumulants of order greater than two vanish when the process is Gaussian. By collecting the cumulants $c_{4m,n}(\tau)$ for $-M \le m, n \le M$ the matrix consisting of cross-cumulants is obtained as :

$$\mathbf{C}_{4x}(\tau) = \sum_{l=1}^{L} c_{4s_l}(\tau) e^{j(\omega_{xl} - \phi_{xl})} \mathbf{a}(\omega_{xl}) \mathbf{a}'(\phi_{xl}). \quad (5)$$

Note that the arrangement in (5) resembles the response of a rectangular array observing far-field sources. With ρ representing either ω_{xl} or ϕ_{xl}, the partial steering vectors $\mathbf{a}(\rho) \triangleq [e^{-j2\rho M}, \ldots, e^{-j2\rho}, 1, e^{j2\rho}, \ldots, e^{j2\rho M}]'$ are centro-symmetric with respect to the array center. Then, $\mathcal{A}(\omega_{xl}, \phi_{xl}) \triangleq \mathbf{a}(\omega_{xl})\mathbf{a}'(\phi_{xl})$ is the array steering matrix for the lth source observed by a virtual rectangular array of size $K \times K$ (where $K = 2M+1$) with elements uniformly spaced at $\{md, nd\}$ for $m, n \in \{-M, \ldots, M\}$. Consequently, $\mathbf{C}_{4x}(\tau)$ in (5) can be thought of as the data collected by an array which observes virtual far-field sources with direction cosines proportional to $2\omega_{xl}$ and $2\phi_{xl}$.

Instead of arranging the cumulants in a rectangular array we can collect them in a single $K^2 \times 1$ vector to obtain :

$$\mathbf{c}_{4x}(\tau) = \sum_{l=1}^{L} c_{s_l}(\tau) e^{j(\omega_{xl} - \phi_{xl})} \mathbf{a}_{\otimes}(\omega_{xl}, \phi_{xl}), \quad (6)$$

where, $\mathbf{a}_{\otimes}(\omega_{xl}, \phi_{xl}) \triangleq vec[\mathcal{A}(\omega_{xl}, \phi_{xl})]$ is obtained by column stacking elements of $\mathcal{A}(\omega_{xl}, \phi_{xl})$. For the sake of notational convenience we henceforth denote $\mathcal{A}(\omega_{xl}, \phi_{xl})$ as $\mathcal{A}_l$. Assuming that the source cumulants $c_{4s_l}(\tau)$ are non-zero for lags $\tau \in \{0, 1 \ldots, \tau_{max} - 1\}$, vectors $\mathbf{c}_{4x}(\tau)$ are collected in a matrix of size $(K^2 \times \tau_{max})$:

$$\mathbf{C}_{\mathbf{X}} = [\mathbf{c}_{4x}(0), \mathbf{c}_{4x}(1), \ldots, \mathbf{c}_{4x}(\tau_{max} - 1)], \quad (7)$$

so as to obtain τ_{max} "snapshots" from the virtual rectangular array. Each vector $\mathbf{c}_{4x}(\tau)$ belongs to $\mathcal{S}$, the signal subspace spanned by the virtual array steering vectors $\{\mathbf{a}_{\otimes}(\omega_{xl}, \phi_{xl})\}_{l=1}^{L}$. Thus, the cumulant based preprocessing maps the 2-D near-field *azimuth-range* estimation (using a linear array) to an equivalent 2-D far-field *azimuth-elevation* problem (arising from a rectangular array).

3.2 2-D Unitary ESPRIT

As mentioned earlier, several algorithms have been proposed to solve this 2-D far-field problem (see for e.g., [5] - [7]). From (6), we observe that the problem is two-fold : (i) estimation of the model parameters ω_{xl}, ϕ_{xl} and (ii) pairing the parameters. We adopt the principle behind the Unitary ESPRIT

algorithm in [5] since it not only results in efficient real-valued processing but also automatically yields paired estimates of the model parameters. Owing to the symmetry in the cumulant definition of (4), the centro-symmetric property of the original array carries over to the new virtual rectangular array $\mathcal{A}_l$. The description of the algorithm in the sequel closely follows that in [5] but with pseudo-data $\mathbf{C_X}$ instead of the original data.

The partial steering vectors $\mathbf{a}(\omega_{xl})$ and $\mathbf{a}(\phi_{xl})$ satisfy the following invariance relationship

$$e^{j\rho}\mathbf{J}_1\mathbf{a}(\rho) = \mathbf{J}_2\mathbf{a}(\rho), \quad \text{for } \rho = \omega_{xl} \text{ or } \phi_{xl}. \tag{8}$$

With $\mathbf{I}_{(K-1)}$ denoting an identity matrix of size $(K-1)$, the selection matrices $\mathbf{J}_1 = [\mathbf{I}_{(K-1)}\ \mathbf{0}_{(K-1)\times 1}]$ and $\mathbf{J}_2 = [\mathbf{0}_{(K-1)\times 1}\ \mathbf{I}_{(K-1)}]$ select the first and last $(K-1)$ rows of $\mathbf{a}(\rho)$. Let us next define a unitary matrix with conjugate centro-symmetric rows as follows :

$$\mathbf{Q}_K \triangleq \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{I}_M & \mathbf{0} & j\mathbf{I}_M \\ \mathbf{0}' & \sqrt{2} & \mathbf{0}' \\ \mathbf{\Pi}_M & \mathbf{0} & -j\mathbf{\Pi}_M \end{bmatrix}. \tag{9}$$

In (9), $\mathbf{\Pi}_M$ is the permutation matrix with ones on the anti-diagonal. It can be shown that $\mathbf{Q}_K$ transforms conjugate centro-symmetric vectors into real-valued vectors. Thus, the real-valued manifold corresponding to $\mathcal{A}_l$ is

$$\mathbf{d}_l = \text{vec}\left[\mathbf{Q}_K^H \mathcal{A}_l \mathbf{Q}_K^*\right]. \tag{10}$$

Consequently, a real-valued signal subspace can be generated by the columns of $\mathbf{D} \triangleq [\mathbf{d}_1, \ldots, \mathbf{d}_L]$. Let $\mathbf{E}_s$ be the orthonormal basis for this subspace. Consequently, there exists a real, non-singular $\mathbf{T}$ such that $\mathbf{E}_s = \mathbf{DT}$. Using (8), the following invariance relations can be shown to hold [5] :

$$\mathbf{K}_{\omega 1}\mathbf{E}_s\mathbf{\Psi}_\omega = \mathbf{K}_{\omega 2}\mathbf{E}_s, \quad \mathbf{K}_{\phi 1}\mathbf{E}_s\mathbf{\Psi}_\phi = \mathbf{K}_{\phi 2}\mathbf{E}_s \tag{11}$$

$$\text{where, } \mathbf{\Psi}_\omega = \mathbf{T}^{-1}\mathbf{\Omega}_\omega\mathbf{T}, \quad \mathbf{\Psi}_\phi = \mathbf{T}^{-1}\mathbf{\Omega}_\phi\mathbf{T}, \tag{12}$$

and $\mathbf{K}_{\omega 1} = \mathbf{I}_K \otimes \mathbf{K}_1$, $\mathbf{K}_{\omega 2} = \mathbf{I}_K \otimes \mathbf{K}_2$, $\mathbf{K}_{\phi 1} = \mathbf{K}_1 \otimes \mathbf{I}_K$, $\mathbf{K}_{\phi 2} = \mathbf{K}_2 \otimes \mathbf{I}_K$, $\mathbf{K}_1 = \text{Re}\left\{\mathbf{Q}_{K-1}^H \mathbf{J}_2 \mathbf{Q}_K\right\}$, $\mathbf{K}_2 = \text{Im}\left\{\mathbf{Q}_{K-1}^H \mathbf{J}_2 \mathbf{Q}_K\right\}$.

Equations (11) and (12) are similar to those that show up in the classical ESPRIT. We can solve for $\mathbf{\Psi}_\omega$ and $\mathbf{\Psi}_\phi$ via the TLS solution in the presence of estimation errors which arise when finite data are used in practice. The matrices $\mathbf{\Omega}_\omega = \text{diag}[\tan(\omega_{x1}), \ldots, \tan(\omega_{xL})]$ and $\mathbf{\Omega}_\phi = \text{diag}[\tan(\phi_{x1}), \ldots, \tan(\phi_{xL})]$ are thus obtained as the eigen values of $\mathbf{\Psi}_\omega$ and $\mathbf{\Psi}_\phi$ respectively.

The two real-valued eigen decompositions in (12) can be replaced by the following complex valued eigen decomposition which also yields an automatic pairing of $\{\omega_x, \phi_x\}$

$$\mathbf{\Psi} \triangleq \mathbf{\Psi}_\omega + j\mathbf{\Psi}_\phi = \mathbf{T}^{-1}(\mathbf{\Omega}_\omega + j\mathbf{\Omega}_\phi)\mathbf{T}. \tag{13}$$

Thus, the lth parameter pair $\{\omega_{xl}, \phi_{xl}\}$ can be obtained from the {real, imaginary} parts of the lth eigen value of $\mathbf{\Psi}$ in (13).

The basis $\mathbf{E}_s$ needed in the preceding relations, can be obtained from the transformed data $\mathbf{C}_{\mathrm{Xr}} = (\mathbf{Q}_K^H \otimes \mathbf{Q}_K^H)\mathbf{C_X}$ which can be shown to lie in the subspace spanned by $\mathbf{D}$. Thus $\mathbf{E}_s$ can be extracted as the L left singular vectors of $[\text{Re}(\mathbf{C}_{\mathrm{Xr}})\ \text{Im}(\mathbf{C}_{\mathrm{Xr}})]$ corresponding to the L largest singular values, provided $L < \min\{K^2, 2\tau_{max}+1\}$.

The signal model for the Y subarray data is similar to (3) with $\{\omega_{xl}, \phi_{xl}\}$ replaced by $\{\omega_{yl}, \phi_{yl}\}$. Hence, we can obtain the paired estimates of the parameters $\{\omega_y, \phi_y\}$ by following the same steps but using $\mathbf{C_Y}$ instead of $\mathbf{C_X}$. Since we can only obtain the paired-parameter estimates with a unknown permutation ambiguity, we denote this as : $\{\tilde{\omega}_x, \tilde{\phi}_x\}$ and $\{\tilde{\omega}_y, \tilde{\phi}_y\}$.

Step 2 : Pairing of $\{\tilde{\omega}_x, \tilde{\phi}_x\}$ and $\{\tilde{\omega}_y, \tilde{\phi}_y\}$

In computing the location coordinates $[\alpha_l, \theta_l, r_l]$ for the sources, the source parameter from the $\{\tilde{\omega}_x, \tilde{\phi}_x\}$ set has to be paired with the right one from the $\{\tilde{\omega}_y, \tilde{\phi}_y\}$ set so that the nonlinear equations in (2) can be solved. There are $L!$ possible pairings for the L sources. Let

$$\mathbf{u}(t) \triangleq [u_{-M-1,0}(t), \ldots, u_{M,0}(t), u_{0,-M-1}(t), \ldots, u_{0,-1}(t), u_{0,1}(t), \ldots, u_{0,M}(t)]' \tag{14}$$

denote the $(2K+1) \times 1$ data vector obtained by stacking the signals collected from the X and Y subarray. Let $\mathbf{B}_{(2K+1)\times L}$ be the corresponding array steering matrix, $\mathbf{s}(t)_{L\times 1}$ the source vector and $\mathbf{v}(t)$ the noise vector, then the matrix form of (1) is :

$$\mathbf{u}(t) = \mathbf{Bs}(t) + \mathbf{v}(t). \tag{15}$$

For each combination $\{\omega_{xp}, \phi_{xp}, \omega_{yq}, \phi_{yq}\}$, the "possible" array steering matrix $\mathbf{B}_p$ is constructed and the model mismatch error e_p,

$$e_p \triangleq \sum_t \left\|\mathbf{B}_p^\perp \mathbf{u}(t)\right\|^2, \tag{16}$$

is evaluated where $\mathbf{B}_p^\perp$ is the projection matrix onto the null space of $\mathbf{B}_p$. The combination which minimizes e_p is then the correct pairing. In contrast to the least-squares estimation of the source coordinates directly from the data, the minimization here is only over a parameter space of size $L!$ (finite set). Finally, the spherical coordinates of the sources are obtained from $\{\omega_x, \phi_x, \omega_y, \phi_y\}$ via (2).

4. SIMULATION RESULTS

Example 1: The algorithm is tested with two non-Gaussian sources at $[\alpha, \theta, r] = [45^0, 45^0, 1.5\lambda]$, and $[-20^0, 10^0, 2\lambda]$ and with $2M+2 = 4$ elements in each of the subarrays. The source signals are generated as BPSK signals filtered with first-order Butterworth filter with cutoff frequency of 0.4π. The sensor noise is circular white Gaussian (SNR= 30dB). The cumulant matrices are estimated from $T = 2048$ data samples ($\tau_{max} = 5$). Figure 3 shows the sensor array configuration, with thick circles representing the sensor locations. The actual directions of arrival and ranges as seen by the reference sensor and the estimates obtained from the proposed method (result of 500 trials) are also indicated. The estimates (mean $\pm$ std. dev.) are: $[\hat{\alpha}_1, \hat{\theta}_1, \hat{r}_1] = [44.76^0 \pm 0.6, 44.48^0 \pm 1.65, 1.54\lambda \pm 0.23]$ and $[\hat{\alpha}_2, \hat{\theta}_2, \hat{r}_2] = [-17.55^0 \pm 6.66, 10.19^0 \pm 0.67, 2.03\lambda \pm 0.31]$ confirm the superior performance of the proposed approach.

Example 2: The algorithm is tested on a second set of sources at $[\alpha, \theta, r] = [20^0, 45^0, 1.5\lambda]$, and $[20^0, 10^0, 2\lambda]$ respectively (same azimuth). The estimates $[\hat{\alpha}_1, \hat{\theta}_1, \hat{r}_1] = [20.19^0 \pm 1.68, 44.69^0 \pm 1.53, 1.67\lambda \pm 0.66]$ and $[\hat{\alpha}_2, \hat{\theta}_2, \hat{r}_2] = [18.7^0 \pm 7.11, 10.41^0 \pm 1.18, 12.1\lambda \pm 0.55]$ are close to the true values (Figure 4).

Example 3: As a final example we consider two sources arriving at the same elevation angle. Table 1 shows the true parameters along with the estimates for two sample sizes. As expected, the estimation variance decreases with larger sample size.

Table 1

Source 1	α_1	θ_1	r_1
True	45^0	20^0	1.5λ
Est :(T=2048)	47.6 ± 9.8	19.4 ± 2.1	$1.58\lambda \pm 0.19$
Est :(T=4096)	46.7 ± 7.3	19.5 ± 1.9	$1.58\lambda \pm 0.18$
Source 2	α_2	θ_2	r_2
True	-10^0	20^0	1.5λ
Est :(T=2048)	-9.1 ± 2.8	21.7 ± 4.5	$1.93\lambda \pm 0.58$
Est :(T=4096)	-9.4 ± 2.2	21.3 ± 3.8	$1.91\lambda \pm 0.31$

References

[1] R. N. Challa, and S. Shamsunder, "High-order subspace based algorithms for passive localization of near-field sources," 29th Asilomar Conference, 1995. 1995.

[2] D. Starer and A. Nehorai, "Passive localization of near-field sources by path following," IEEE T-SP, March 1994.

[3] A. L. Swindlehurst and T. Kailath,"Passive direction of arrival and range estimation for near-field sources," IEEE Spec. Est. and Mod. Workshop, 1988.

[4] R. Roy and T. Kailath, "ESPRIT-Estimation of Signal Parameters by Rotational Invariance Techniques," IEEE T-ASSP, pp. 984-995, July 1989.

[5] M. D. Zoltowski, M. Haardt, and C. P. Mathews, "Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT," IEEE T-SP, 1996 (to appear).

[6] A. L. Swindlehurst and T. Kailath, "Azimuth/elevation direction finding using regular array geometries", IEEE T-AES, Jan. 1993.

[7] A. J. van der Veen, P. B. Ober, E. D. Deprettere, "Azimuth and elevation computation in High Resolution DOA Estimation", IEEE T-SP, July 1992.

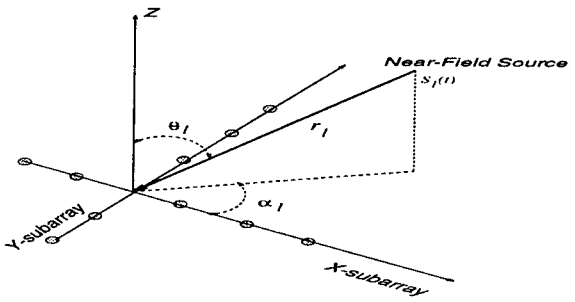


Figure 1. Near-field scenario.

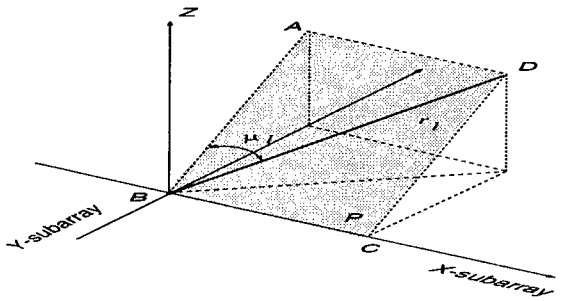


Figure 2. Source as seen by the X subarray.

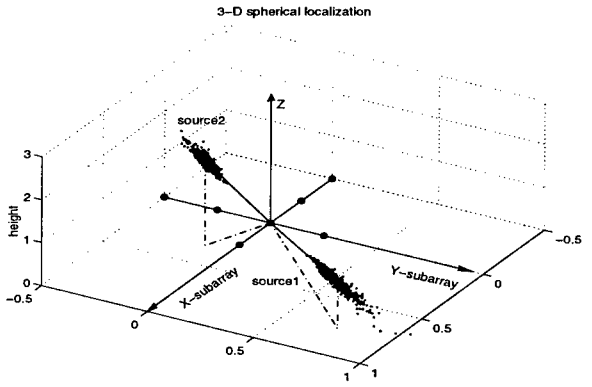


Figure 3. Example 1

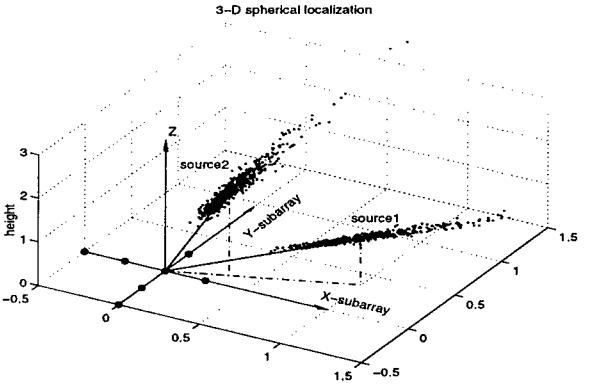


Figure 4. Example 2

Frequency Invariant Beamforming with Exact Null Design[†]

Peter J. Kootsookos
(CR)²RASys, RSISE, ANU
Canberra 0200, AUSTRALIA

Darren B. Ward
Dept of Engineering, FEIT, ANU,
Canberra 0200, AUSTRALIA

Robert C. Williamson
Dept of Engineering, FEIT, ANU,
Canberra 0200, AUSTRALIA

Abstract

This paper extends earlier results by Ward, Kennedy and Williamson [1,2] for the design of broadband arrays with frequency-invariant (FI) beam patterns to the case where it is desired to place an exact null in a given direction. The beamforming is done using appropriately selected FIR filters.

First, the previous results for generating FI beam patterns using FIR filters are briefly summarised. Second, new results which give the conditions required for exact nulls in the beam pattern for all frequencies in any, possibly non-FI, beam pattern are given. Third, a method of generating beam patterns which possess an exact null and which are close, in an L_2 sense, to an arbitrary FI pattern is presented. Finally, some preliminary experimental results which corroborate the theoretical findings are presented.

1. Problems Addressed

Consider an array of N spatially separated omni-directional microphones. The array has a nominal aperture of P half-wavelengths at a given frequency. The signals from each sensor are sampled at sampling frequency f_s and are filtered using L-tap finite impulse response filters with frequency responses

$$H_n(f) := \sum_{m=0}^{L-1} h_n[m] e^{-j2\pi f m}, \quad n = 1, 2, \ldots, N.$$

This is illustrated in Figure 1.

We wish to select the filter coefficients, $h_n[m]$, and the sensor locations, x_n, so that the farfield array response from direction θ

$$A(\theta, f) := \sum_{n=1}^{N} H_n(f) \exp\left(j2\pi f \frac{f_s}{c} x_n \sin\theta\right) = \boldsymbol{h}^T \boldsymbol{d}(\theta, f) \tag{1}$$

[†]The authors wish to acknowledge the funding of the activities of the Cooperative Research Centre for Robust and Adaptive Systems by the Australian Commonwealth Government under the Cooperative Research Centres Program.
Supported in part by the Australian Research Council.

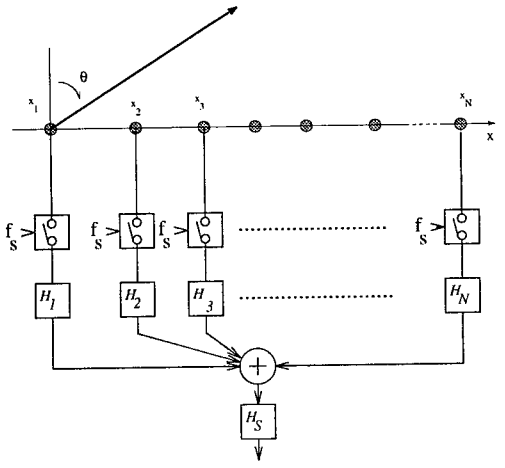


Figure 1. The array geometry assumed.

possesses certain properties over the frequency range $f \in [f_L, f_U]$. The velocity of wave propagation is denoted c. The NL-dimensional vector of FIR coefficients is

$$\boldsymbol{h}^T = [h_1[0] \ldots h_N[0] \ldots h_1[L-1] \ldots h_N[L-1]]$$

and

$$\boldsymbol{d}(\theta, f) = \begin{bmatrix} e^{j2\pi f \tau_1(\theta)} \\ \vdots \\ e^{j2\pi f \tau_N(\theta)} \\ \vdots \\ e^{j2\pi f [\tau_1(\theta) - L + 1]} \\ \vdots \\ e^{j2\pi f [\tau_N(\theta) - L + 1]} \end{bmatrix}$$

is the NL-dimensional delay vector with

$$\tau_n(\theta) := \frac{f_s}{c} x_n \sin\theta.$$

The property of frequency invariance has been investigated previously [1–3]. This paper examines obtaining exact nulls in a beam pattern and the interaction of this property with frequency invariance.

0-8186-7576-4/96 $5.00 © 1996 IEEE

1.1. Problem One: FI

Suppose it is desired that

$$A^{FI}(\theta, f) = A(\theta) \quad f \in [f_L, f_U].$$

It was shown in [1] how to choose the x_n locations and that, when chosen appropriately, the array filters should have the dilation property

$$H_n^{\mathrm{P}}(f) = H_{\mathrm{ref}}^{\mathrm{P}}\left(\frac{x_n}{x_{\mathrm{ref}}} f\right)$$

where $H_n^{\mathrm{P}}(f)$ is the filter response of the n^{th} filter and $H_{\mathrm{ref}}^{\mathrm{P}}(f)$ is the *primary filter* response at some reference location x_{ref}.

Ward, Kennedy and Williamson [2] considered two possible filter bank implementations with this property: a multirate approach and a single rate approach. This paper follows the single rate approach with the primary filter coefficients given by

$$h_n^{\mathrm{P}}[m] = \frac{1}{\gamma_n} \sum_{k=-(L-1)/2}^{(L-1)/2} h_{\mathrm{ref}}^{\mathrm{P}}[m] \mathrm{sinc}\left(\frac{m}{\gamma_n} - k\right)$$

where $\gamma_n = \frac{x_n}{x_{\mathrm{ref}}}$. The secondary filter coefficients are chosen so that $H^{\mathrm{S}}(f)$ is a differentiator over $f \in [f_L, f_U]$.

The full frequency invariant shading filters are then

$$h_n^{FI}[m] = g_n h_n^{\mathrm{P}}[m] * h^{\mathrm{S}}[m]$$

where $*$ denotes convolution in the m index and g_n is a spatial weighting term to account for the (possibly) nonlinear array spacing.

1.2. Problem Two: END

The first new results of this paper are Proposition 1 and its Corollary, which are the conditions for Exact Null Design (END).

Proposition 1 — Condition for a Broadband Null
A broadband null at θ_0 will be available if and only if either

$$A^{END}(\theta_0, f) := \sum_{n=1}^{N} H_n^{END}(f) e^{j2\pi f \tau_n(\theta_0)} = 0, \ \forall f \quad (2)$$

or, equivalently,

$$\sum_{n=1}^{N} \left[h_n^{END}[m] * \frac{\sin\left(\pi[m + \tau_n(\theta_0)]\right)}{\pi[m + \tau_n(\theta_0)]} \right] = 0, \ \forall m \quad (3)$$

where $*$ again denotes convolution in the m index. □

Proof. From (1), the array response in direction θ_0 is

$$\begin{aligned} A(\theta_0, f) & \quad (4) \\ &= \sum_{n=1}^{N} H_n(f) \exp\left(j2\pi f \frac{f_s}{c} x_n \sin\theta_0\right) \\ &= \sum_{n=1}^{N} H_n(f) e^{j2\pi f \tau_n(\theta_0)} \quad (5) \\ &= \sum_{n=1}^{N} \sum_{m=0}^{L-1} h_n[m] e^{-j2\pi f m} e^{j2\pi f \tau_n(\theta_0)} \\ &= \sum_{m=0}^{L-1} \left[\sum_{n=1}^{N} h_n[m] * \frac{\sin\left(\pi[m + \tau_n(\theta_0)]\right)}{\pi[m + \tau_n(\theta_0)]} \right] e^{-j2\pi f m}. \quad (6) \end{aligned}$$

Equation (5) yields (2) and the inverse discrete Fourier transform of (6) gives (3). ■

It is not immediately clear how (2) and (3) may be easily enforced. The following result shows this.

Corollary 1 — Integer Delay Property
If $\tau_n(\theta_0)$ is an integer then

$$\sum_{n=1}^{N} h_n[m + \tau_n(\theta_0)] = 0, \quad \forall m \quad (7)$$

is a sufficient condition for (3) being satisfied. □

Proof. If $\tau_n(\theta_0) \in \mathbb{Z}$ then (3) becomes

$$\sum_{n=1}^{N} h_n[m] * \delta[m + \tau_n(\theta_0)] = 0, \quad \forall m$$

where δ is the Kronecker delta. ■

Remark 1: *Note that it is always possible to place a null at $\theta_0 = 0$ because in this case*

$$\tau_n = \frac{f_s}{c} x_n \sin\theta_0 = 0, \ \forall n$$

and (3) reduces to requiring

$$\sum_{n=0}^{N} h_n^{END}[l] = 0, \ \forall l.$$

□

Using this idea, the END condition for a null at broadside may be written as

$$\boldsymbol{C}^T \boldsymbol{h} = 0$$

where

$$\boldsymbol{C} := I_L \otimes \mathbf{1}_N$$

where I_L is the $L \times L$ identity matrix and $\mathbf{1}_N$ is the N-vector of ones.

1.3. Problem Three: FI with END

Designing a frequency invariant array generates array filters h_n^{FI}. Placing an exact null imposes the underdetermined constraints (7). The question of whether it is possible to do both arises, taking into account the slack of the exact null constraints.

The approach we take is similar to that of Frost [4]. Assume we time-delay beamsteer the FI beamformer to direction $-\theta_0$. The effect of this is to move the null to broadside. In the remainder, we will use the tilde symbol ($\tilde{}$) to indicate a steered quantity.

The steered filter coefficient NL-vectors, with END conditions imposed, are given by $\tilde{\boldsymbol{h}}_{FIEND} = \tilde{\boldsymbol{h}}_{FI} + \tilde{\boldsymbol{h}}_c$ where $\tilde{\boldsymbol{h}}_c$ are the deviations from $\tilde{\boldsymbol{h}}_{FI}$ which allow for exact null design.

We approach the problem by imposing the exact null constraints (7) while minimising the cost functional

$$\begin{aligned} J &= \int_{-\pi}^{+\pi}\int_{f_L}^{f_U} \left|\tilde{A}^{FI}(\theta,f) - \tilde{A}^{FIEND}(\theta,f)\right|^2 df\,d\theta \\ &= \int_{-\pi}^{+\pi}\int_{f_L}^{f_U} |\tilde{\boldsymbol{h}}_c^T \boldsymbol{d}(\theta,f)|^2 df\,d\theta = \tilde{\boldsymbol{h}}_c^T \boldsymbol{D}\tilde{\boldsymbol{h}}_c \qquad (8) \end{aligned}$$

where

$$\boldsymbol{D} := \int_{-\pi}^{+\pi}\int_{f_L}^{f_U} \boldsymbol{d}(\theta,f)\boldsymbol{d}^H(\theta,f) df\,d\theta.$$

The best $\tilde{\boldsymbol{h}}_c$ is then found as the solution to the optimisation problem

$$\min_{\tilde{\boldsymbol{h}}_c} \tilde{\boldsymbol{h}}_c^T \boldsymbol{D}\tilde{\boldsymbol{h}}_c$$

subject to $\boldsymbol{C}^T(\tilde{\boldsymbol{h}}_{FI} + \tilde{\boldsymbol{h}}_c) = 0$. The solution to this problem is

$$\tilde{\boldsymbol{h}}_c^{\text{opt}} = \boldsymbol{D}^{-1}\boldsymbol{C}\left[\boldsymbol{C}^T\boldsymbol{D}^{-1}\boldsymbol{C}\right]^{-1}\boldsymbol{a}$$

where $\boldsymbol{a} := -\boldsymbol{C}^T\tilde{\boldsymbol{h}}_{FI}$.

The matrix $\boldsymbol{D}$ is of full rank, provided no two sensor locations coincide; $\boldsymbol{C}$ is also of full rank. This solution has the same form as that presented in [4]. The optimum *unsteered* response is then

$$A^{FIEND}(\theta,f) = (\boldsymbol{h}_{FI} + \boldsymbol{h}_c^{\text{opt}})^T \boldsymbol{d}(\theta,f)$$

where

$$\boldsymbol{h}_c^{\text{opt}} = \begin{bmatrix} h_1^{\text{opt}}[0] \\ \vdots \\ h_N^{\text{opt}}[0] \\ \vdots \\ h_1^{\text{opt}}[L-1] \\ \vdots \\ h_N^{\text{opt}}[L-1] \end{bmatrix} = \begin{bmatrix} \tilde{h}_1^{\text{opt}}[-\tau_1(\theta_0)] \\ \vdots \\ \tilde{h}_N^{\text{opt}}[-\tau_N(\theta_0)] \\ \vdots \\ \tilde{h}_1^{\text{opt}}[L-1-\tau_1(\theta_0)] \\ \vdots \\ \tilde{h}_N^{\text{opt}}[L-1-\tau_N(\theta_0)] \end{bmatrix}$$

2. Example Array Design

The figures show the resulting array responses for the following array parameters:

$$f_L = 1000\text{Hz},\quad f_U = 2000\text{Hz},\quad N = 7.$$

The primary filters in the FI design are chosen to be 5 taps, the secondary filters are 3 taps long. The END design required a null to be placed at $\theta = \pi/4 = 45°$. The array was 36cm long.

Because the array was to be tested in a small anechoic chamber, the farfield design methodology presented here was modified to allow for a *nearfield* design. For the results shown, the source was 2.8 metres from the array. Space precludes inclusion of the derivations for the nearfield case [5]. For an alternative technique, see [6,7].

Figures 2, 3 and 4 show array responses at 20 regularly spaced frequencies between 1000Hz and 2000Hz for the original FI design [2], an END design with no account taken of frequency invariance and the FIEND design where filters are adjusted to cater for the null while minimising the cost function (8).

Remark 2: *Clearly the FIEND responses more closely resemble the FI-only response than does the END response; the value of J for the END response plotted is 0.0487 and the value of J for the FIEND response is 0.0114.* □

Time constraints precluded measurement of the array responses of all designs; only the END design was tested empirically.

Good correspondence between theoretical and measured results were obtained for frequencies 995Hz, 1248Hz, 1505Hz, 1748Hz and 2004Hz are displayed in Figure 5. Some problems were encountered with the response measured at 1748Hz.

3. Conclusions

We have presented one new result which allows exact nulls to be incorporated into any broadband array design. If frequency invariance is required, another new result, presented here, allows exact nulls while minimising a mean square error cost between a frequency invariant design and the design which includes a null.

An exact null design was tested in the laboratory; theoretical and measured responses compared favourably.

Acknowledgements

The authors are grateful for the use of the anechoic chamber at the Department of Mechanical Engineering of the Australian Defence Force Academy. We thank Marion Burgess of the Acoustics and Vibration Centre for the use of their facilities.

Martin Buss, Bob Edwards, Marshall Shepard and Martin Stonebridge are also gratefully acknowledged for technical support for the design and implementation of the microphone array.

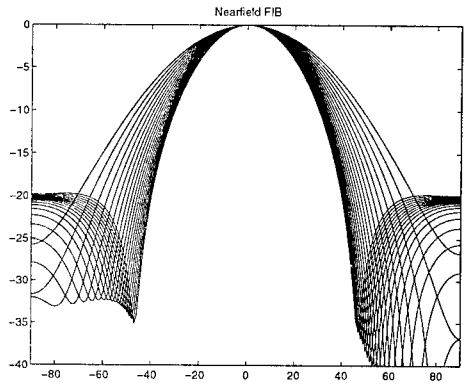


Figure 2. Array responses at various frequencies for FI array.

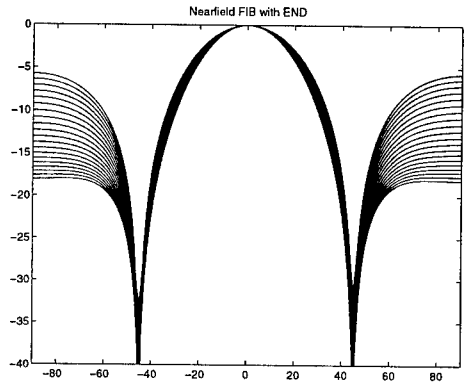


Figure 3. Array responses at various frequencies for END array.

References

[1] D.B. Ward, R.A. Kennedy & R.C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," *J. Acoust. Soc. Am.* 97(2) (1995), 1023–1034.

[2] D.B. Ward, R.A. Kennedy & R.C. Williamson, "FIR Filter Design For Frequency-Invariant Beamformers," *IEEE Signal Processing Letters, to appear* (1996).

[3] J.H. Doles III & F.D. Benedict, "Broad-Band Array Design Using the Asymptotic Theory of Unequally Spaced Arrays," *IEEE Transactions on Antennas and Propagation* 36(1) (January, 1988), 27–33.

[4] O.L. Frost III, "An Algorithm for Linearly Constrained Adaptive Array Processing," *Proc. IEEE* 60(8) (August, 1972), 926–935.

[5] F. Khalil, J.P. Jullien & A. Gilloire, "Microphone array for sound pickup in teleconference systems," *J. Audio Eng. Soc.* 42(9) (Sept., 1994), 691–700.

[6] Rodney A. Kennedy, Thushara Abhayapala, Darren B. Ward & Robert C. Williamson, "Nearfield Broadband Frequency Invariant Beamforming," *Proceedings of IEEE Conference on Acoustics, Speech and Signal Processing, to appear* (1996).

[7] Peter J. Kootsookos, Darren B. Ward & Robert C. Williamson, "Frequency Invariant Beamforming with Exact Null Design: Farfield and Nearfield ," *in preparation* (1996).

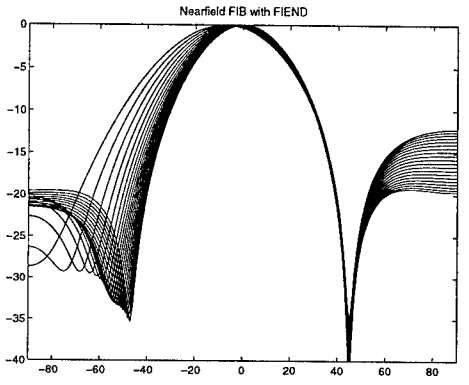


Figure 4. Array responses at various frequencies for FIEND array.

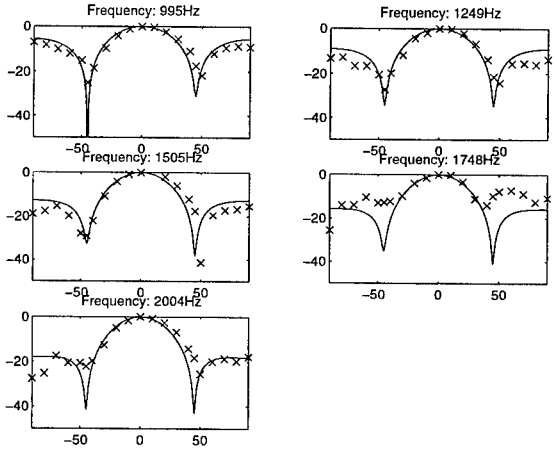


Figure 5. Predicted (—) and measured (x) array responses at various frequencies for END array.

Low-Complexity Angle of Arrival Estimation of Wideband Signals Using Small Arrays

Jari Yli-Hietanen[1], Kari Kalliojärvi[2], Jaakko Astola[1]

[1] Tampere University of Technology
Signal Processing Laboratory
P.O. Box 553
FIN-33101 Tampere, Finland
ypsilon@cs.tut.fi, jta@cs.tut.fi

[2] Nokia Research Center
Radio Communications Laboratory
P.O. Box 45
FIN-00211 Helsinki, Finland
kari.kalliojarvi@research.nokia.com

Abstract

When the signal to noise ratio is relatively high, the angle of arrival of the strongest signal can be estimated with a very simple method and a small 3-D sensor array. The differences in the arrival times of the wideband signal received by spatially separated sensors are estimated using the polarity coincidence correlation. These time differences, i.e. time delays, determine the angle of arrival. In this paper the effects of quantization of the time delays are studied. It is found out that this simple method gives comparable performance to the conventional, direct correlation based methods in the case of a relatively high signal to noise ratio.

1. Introduction

In this paper a low-complexity method for the estimation of the angle of arrival of a wideband signal with a small array is developed. The lower the complexity of the angle-of-arrival estimation method is, the simpler, the less expensive, and the more reliable is the hardware with which the method can be realized. A small array was set as the goal in order to make possible a less expensive array which could even be used as a part of portable equipment.

A small sensor array is defined here as follows: The number of sensors is less than 8 and the maximum distance between any two sensors is less than $20 \cdot T_s \cdot c$, where T_s is the sampling interval and c is the propagation velocity of the signal. There are no other restrictions to the locations of the sensors, so they can form a three-dimensional structure.

In order to achieve a low-complexity method only the angle of arrival of the signal from the strongest signal source is estimated. The differences in arrival times of a wideband signal received by spatially separated sensors can be estimated one by one if the sensor array is receiving one dominating wave propagating signal.

This work was sponsored by Nokia Foundation.

When the signal to noise ratio is relatively high (e.g. 5 dB in the case of a one signal in white gaussian noise) time delays with which the signal reaches different sensors can be estimated using polarity coincidence correlation [1], meaning that signals are quantized to 1-bit representation. The signal to noise ratio needed to achieve a certain variance of the time delay estimate depends on the signal and noise spectrums, cross-correlations of the signal and noise at different sensors, cross-correlations of noise at different sensors, the length of the estimation window, and the sampling rate. The variances of the time delay estimates are given in closed-form expressions for general signal and noise spectra in [2].

The main advantage of the polarity coincidence correlation is the possibility to use simple 1-bit quantization of signals. This results that simpler analog automatic gain control can be utilized when compared to multibit quantization. Polarity coincidence correlation is also computationally much simpler and less demanding than direct cross-correlation methods because it can be implemented without multiplications [3]. Also the handling of 1-bit signals requires considerably less memory than that of multibit signals.

The disadvantage of using 1-bit quantization is that interpolation of the signal values between sampled values isn't possible. Therefore the time delays can be estimated only with the accuracy of one sampling interval. The error of the time delay estimates caused by rounding to the nearest multiple of the sampling interval causes errors to the estimate of the propagation vector. This kind of error, which hasn't been considered earlier, is studied in this paper. The study is based on modeling the rounding error of the time delay estimates by independent white noise which is uniformly distributed in the interval $[-T_s/2, T_s/2)$.

2. Method

In principle the introduced method for the angle of arrival estimation is as follows: First, signals received by spatially separated sensors are quantized to 1-bit represen-

0-8186-7576-4/96 $5.00

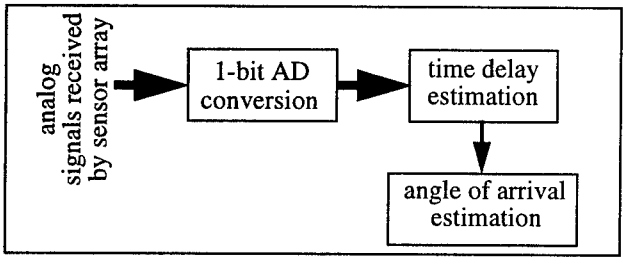


Figure 1. Block diagram of the angle of arrival estimation method

tation. Then, the time delays between signals in different sensors are estimated by polarity coincidence correlation. Finally, these time delays are used to determine the angle of arrival. The block diagram of the angle of arrival estimation method is presented in Figure 1.

Let us assume that the sensor array is receiving a wave propagating signal $s(t,\boldsymbol{x})$ caused by a distant event. t is the time and $\boldsymbol{x}$ is a three-dimensional vector representing a location in an orthogonal coordinate system. It is assumed that $s(t,\boldsymbol{x})$ can be modeled as a sum of plane waves with common direction of propagation but with different frequency and amplitude,

$$s(t,\boldsymbol{x}) = \sum_l A_l \exp\left(j\omega_l\left(t-\boldsymbol{k}^{\mathrm{T}}\boldsymbol{x}\right)\right), \tag{1}$$

where j is the imaginary unit, ω_l is the frequency, A_l is the amplitude of the lth component of the wideband signal, and $\boldsymbol{k}$ is a propagation vector which determines the direction and the velocity of propagation of the plane wave, $\|\boldsymbol{k}\| = 1/c$. T denotes matrix transpose.

Let the signal received by the ith sensor located at $\boldsymbol{x}_i$ be

$$y_i(t) = s(t,\boldsymbol{x}_i) + w_i(t) \tag{2}$$

where $w_i(t)$ is the noise component received by ith sensor. If sensors in the array are identical, then in the ideal noise-free case the only difference between signals received by different sensors is the time delay because $s(t,\boldsymbol{x})$ is a sum of plane waves with a common direction of propagation.

The time delay between the signals received by the ith and the mth sensor is

$$\tau_{im} = \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{x}_i-\boldsymbol{x}_m) = \boldsymbol{k}^{\mathrm{T}}\boldsymbol{x}_{im}, \tag{3}$$

where $\boldsymbol{x}_{im}$ is called a sensor vector. The propagation vector $\boldsymbol{k}$, which gives the angle of arrival, is determined by three time delays $\tau_{i_n m_n}$, $n = 1,2,3$ if corresponding sensor vectors $\boldsymbol{x}_{i_n m_n}$ are linearly independent. In general, $\boldsymbol{k}$ is the least squares solution of the matrix equation $\boldsymbol{t}_M = \boldsymbol{V}_M\boldsymbol{k}$ (where M is the number of the time delay estimates used in the estimation of $\boldsymbol{k}$), i.e.

$$\boldsymbol{k} = \boldsymbol{V}_M^{\mathrm{T}}\left(\boldsymbol{V}_M^{\mathrm{T}}\boldsymbol{V}_M\right)^{-1}\boldsymbol{t}_M, \tag{4}$$

where

$$\boldsymbol{V}_M = \left[\boldsymbol{x}_{i_1 m_1}\ \boldsymbol{x}_{i_2 m_2}\ \cdots\ \boldsymbol{x}_{i_M m_M}\right]^{\mathrm{T}}, \qquad \boldsymbol{t}_M = \left[\tau_{i_1 m_1}\ \tau_{i_2 m_2}\ \cdots\ \tau_{i_M m_M}\right]^{\mathrm{T}}, \tag{5}$$

provided that the rank of the matrix $\boldsymbol{V}_M$ is 3.

Time delays are estimated using polarity coincidence correlation: d_{im} is taken as an estimate of the time delay τ_{im} if it maximizes the sum

$$R_{im}(d_{im}) = \sum_n \operatorname{sign}(y_i(nT_s))\operatorname{sign}(y_m(nT_s-d_{im})) \ . \tag{6}$$

The estimated delays d_{im} can only get values which are multiple of sampling interval because interpolation of the signals quantized to 1-bit representation is not possible. The error of the time delay estimate caused by rounding to the nearest multiple of sampling interval can be modeled as white noise which is uniformly distributed in the interval $[-T_s/2, T_s/2)$. The error is assumed to be statistically independent of the time delay to be rounded and other time delays.

If all pairs of the vectors $\boldsymbol{x}_{i_n m_n}$ contained in the matrix $\boldsymbol{V}_M$ are linearly independent, meaning that no pair of vectors have equal direction, then the error of the time delay can also be assumed to be statistically independent of the errors of the other time delays. In this case the covariance matrix of the rounding error of the estimate of the time delay vector is

$$\operatorname{cov}\{\Delta\boldsymbol{t}_M\} = \frac{T_s^2}{12}\boldsymbol{I}, \tag{7}$$

where $\boldsymbol{I}$ is the identity matrix and

$$\Delta\boldsymbol{t}_M = \boldsymbol{d}_M - \boldsymbol{t}_M, \quad \boldsymbol{d}_M = \left[d_{i_1 m_1}\ d_{i_2 m_2}\ \cdots\ d_{i_M m_M}\right]^{\mathrm{T}}. \tag{8}$$

The rounding of the time delays to the nearest multiple of sampling interval causes errors to the estimate of propagation vector $\boldsymbol{k}$. The covariance matrix of that kind of error can now be estimated because $\boldsymbol{k}$ is calculated by matrix multiplication of the time delay vector,

$$\operatorname{cov}\{\Delta\boldsymbol{k}_M\} = \frac{T_s^2}{12}\left(\boldsymbol{V}_M^{\mathrm{T}}\boldsymbol{V}_M\right)^{-1}, \tag{9}$$

where

$$\Delta\boldsymbol{k}_M = \boldsymbol{k}_M - \boldsymbol{k}, \quad \boldsymbol{k}_M = \boldsymbol{V}_M^{\mathrm{T}}\left(\boldsymbol{V}_M^{\mathrm{T}}\boldsymbol{V}_M\right)^{-1}\boldsymbol{d}_M. \tag{10}$$

The more time delay estimates are taken into account in the calculation of the estimate of the propagation vector $\boldsymbol{k}$ the smaller are the variances of the error of the components of $\boldsymbol{k}_M$.

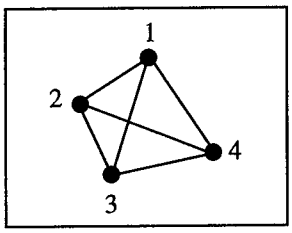


Figure 2. The 3-D grid of four sensors used in simulations

The computational load of the proposed angle-of-arrival method depends on the complexity of the estimation of the cross-correlation functions, the maximum allowed time delay τ_{max}, and the number of the time delay estimates M used. When nonoverlapping estimation windows and polarity coincidence correlation are used, one conditional counter increment per sample is needed to estimate one value of the cross-correlation function. When conventional direct correlation method is used, one multiplication and one addition per sample are needed. $(2d_{max}+1)$ estimated values of the cross-correlation function are needed for each of the M time delay estimate used, where

$$d_{max} = \text{round}\left(\frac{\tau_{max}}{T_s}\right). \tag{11}$$

After estimating the cross-correlation functions their maximal values are found out to estimate the time delays. Finally the propagation vector $\boldsymbol{k}$ which determines the angle of arrival is calculated by a multiplication of $3xM$ matrix and a vector of length M, see equation (10).

3. Simulation Results

A grid of four sensors with equal distance $20 \bullet T_s \bullet c$ between all sensors (see Figure 2) is used as the array in the simulations.

Distribution of the rounding errors of the time delay estimates was verified by simulations. Randomly generated vectors $\boldsymbol{k}(n)$, $n = 1, 2, \ldots, 100\,000$ were used as a test sequence. The vectors $\boldsymbol{k}(n)$ were generated as follows: First three-dimensional vectors were generated, whose components were independent and uniformly distributed in the interval [-1,1]. Then the vectors were scaled so that the norm of each vector was $1/c$. The true values of all the six possible time delays were formed with the equation

$$\boldsymbol{t}_6 = \boldsymbol{V}_6\boldsymbol{k}, \tag{12}$$

where

$$\begin{aligned} \boldsymbol{V}_6 &= \left[x_{12}\ x_{13}\ x_{14}\ x_{23}\ x_{24}\ x_{34}\right]^T \\ \boldsymbol{t}_6 &= \left[\tau_{12}\ \tau_{13}\ \tau_{14}\ \tau_{23}\ \tau_{24}\ \tau_{34}\right]^T \end{aligned}. \tag{13}$$

After that the time delays were rounded to the nearest multiple of the sampling interval. Then the error between the original and rounded time delays was estimated. The histogram of the error values is shown in Figure 3. The assumption of uniform distribution holds very well.

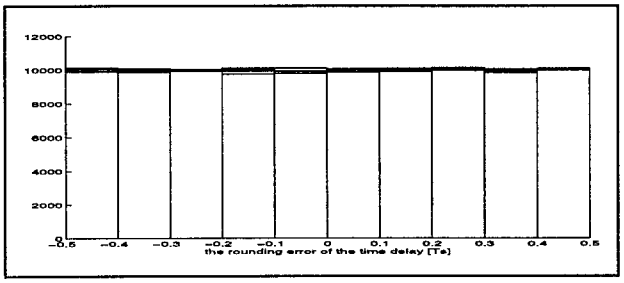


Figure 3. Histogram of rounding errors of time delays

The error of the estimate of $\boldsymbol{k}$ caused by the rounding of the time delays to the nearest multiple of the sampling interval were formed by using 3 and 6 rounded time delay estimates in calculating the estimate of $\boldsymbol{k}$. In the case of 3 time delays

$$\boldsymbol{V}_3 = \left[x_{12}\ x_{13}\ x_{14}\right]^T \qquad \boldsymbol{t}_3 = \left[\tau_{12}\ \tau_{13}\ \tau_{14}\right]^T \tag{14}$$

were used.

The sample covariance matrices and the covariance matrices of the errors $\Delta\boldsymbol{k}_3$ and $\Delta\boldsymbol{k}_6$ formed with equations (9) and (10) are shown in Table 1. The simulated values are quite close to the values given by the equation (9). It is noticed by comparing the covariance matrices of the errors $\Delta\boldsymbol{k}_3$ and $\Delta\boldsymbol{k}_6$ that in this case the use of 6 time delay estimates instead of 3 time delay estimates halved the variance of the error of the components of the estimated propagation vector.

The proposed angle-of-arrival method was then simulated. The test signal used was a sum of 5 sinusoids with frequencies of 0.02π, 0.06π, 0.1π, 0.26π, and 0.46π with respect to the sampling frequency 2π and with a common amplitude. The test signal was assumed to propagate as a plane wave. The same signal but propagating as a plane wave to a different direction was used as an interfering signal. The power of the interfering signal, $v_I(n)$, was increasing during simulation,

$$v_I(n) = 10^{\left(\frac{2n}{10^6}-1\right)} v_S, \tag{15}$$

where v_s is the power of the test signal. The signals received by the sensors were the sum of the test signal, the interfering signal and white independent gaussian noise with variance $0.1v_s$. All the six possible time delays were estimated using polarity coincidence correlation presented in equation (6) with nonoverlapping estimation windows

	sample covariance (divided by $10^{-3}\left(\frac{T_s}{c}\right)^2$)	covariance (divided by $10^{-3}\left(\frac{T_s}{c}\right)^2$)
Δk_3	$\begin{bmatrix} 0.2080 & -0.1207 & -0.0848 \\ -0.1207 & 0.3490 & -0.0482 \\ -0.0848 & -0.0482 & 0.3809 \end{bmatrix}$	$\begin{bmatrix} 0.2083 & -0.1203 & -0.0851 \\ -0.1203 & 0.3472 & -0.0491 \\ -0.0851 & -0.0491 & 0.3819 \end{bmatrix}$
Δk_6	$\begin{bmatrix} 0.1043 & 0.0001 & -0.0003 \\ 0.0001 & 0.1046 & 0.0001 \\ -0.0003 & 0.0001 & 0.1046 \end{bmatrix}$	$\begin{bmatrix} 0.1042 & 0.0000 & 0.0000 \\ 0.0000 & 0.1042 & 0.0000 \\ 0.0000 & 0.0000 & 0.1042 \end{bmatrix}$

Table 1. Covariance matrices of the error of the estimated propagation vectors

with a length of 10^4 samples. The length of the test signal was 10^6 samples. The estimate of the propagation vector $\boldsymbol{k}$ was formed with equation (10). The components of the vector ck_6 as a function of time are presented in Figure 4. In this simulation the method tracks the signal from the strongest signal source when the power of the stronger signal is about 2.5 times the power of the weaker signal. The results achieved by using direct correlation instead of polarity coincidence correlation, i.e. using

$$R_{im}(d_{im}) = \sum_n y_i(nT_s)\, y_m(nT_s - d_{im}) \qquad (16)$$

instead of equation (6), are also presented in Figure 4. The proposed method gives comparable performance to the conventional correlation based method.

4. Conclusions

A low-complexity method for estimation of the angle of arrival of the signal from the strongest signal source was introduced. The method is based on the differences in the arrival times of the signal at different sensors. These differences are estimated using polarity coincidence correlation and as a consequence, 1-bit quantization can be used. Because of this the required amount of calculation is significantly reduced when compared to conventional methods without noticeable differences in the performance when the signal to noise ratio is relatively high. The introduced method also makes possible quite complex 3-D sensor placements if necessary. These characteristics make the introduced method very easy and cheap to implement, and robust to operate. Therefore the method is suitable for low-cost applications, where it is sufficient to find out only the angle of arrival of the strongest signal source.

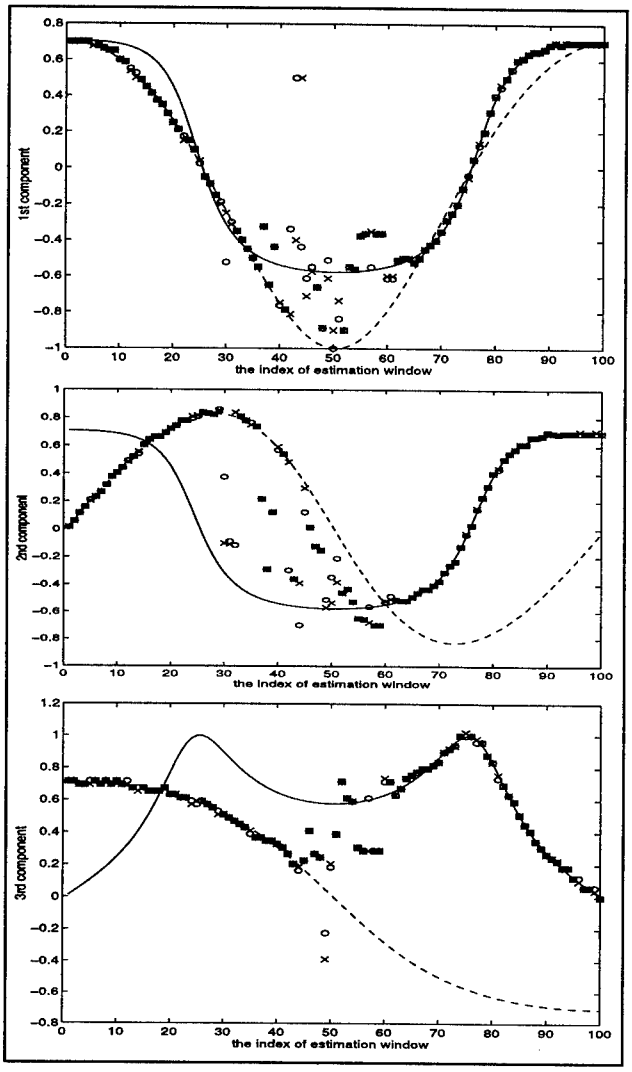


Figure 4. The components of the propagation vectors as a function of time, circles: estimated with polarity coincidence correlation; crosses: estimated with direct correlation; dashed line: the propagation vector of the test signal; solid line: the propagation vector of the interfering signal

References

[1]. A. Fertner and A. Sjölund, "Comparison of various time delay estimation methods by computer simulation," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-34, no 5, pp. 1329-1330, Oct. 1986.

[2]. R. Cusani, "Performance of fast time delay estimators," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, no 5, pp. 757-759, May 1989.

[3]. G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Transactions on Signal Processing,* vol. 41, no 2, pp. 525-533, Feb. 1993.

Improved MUSIC using Uniform Subarrays

Gaetano Scarano, Giampiero Panci, Giovanni Jacovitti
INFOCOM Dept., Università di Roma "La Sapienza"
via Eudossiana 18, I-00184 Roma, Italy
gaetano@infocom.ing.uniroma1.it

Abstract

Performance analysis shows the asymptotic optimality of the MUSIC technique applied to bearing estimation problems for a sufficiently large number of sensors and not fully-coherent sources (see for instance [1, 2]). This implies that a large number of covariance lags has to be computed; moreover, the computational load of the eigen–decomposition of large covariance matrices may be too severe for practical applications.

With reference to uniformly spaced linear arrays (ULA's), in this paper we show that the accuracy gain associated to an increased number of sensors can be alternatively obtained by applying the MUSIC technique to particular configurations of pairs of ULA's, referenced to as subarrays, using a significantly smaller number of sensors.

It is also shown that the accuracy loss of the proposed method, w.r.t. a full ULA covering the same array aperture, can be minimized by varying the distance between the two subarrays.

The provided simulation results shows the applicability of the proposed method.

1 Introduction

The classical problem in array signal processing is to estimate the directions of arrival (DOA) of plane waves with an array of sensors. Among others, the MUSIC technique [4] has became popular due to its simple formulation, easy implementation and high statistical efficiency.

Moreover, performance analysis (see for instance [1, 2]) shows the asymptotic optimality of the MUSIC technique for a sufficiently large number of sensors and not fully-coherent sources. On the other hand, a larger array implies the computation of more covariance lags and a more heavy computational load of the eigen–decomposition.

Looking at the application of high–resolution DOA estimation methods based on subspace decomposition, such as MUSIC, ESPRIT, *etc.*, to *reduced* covariance matrices, we analize here a sensors arrangements in pairs of ULA's; each ULA will be referred to as a *subarray*. We will show that the accuracy of the MUSIC technique is basically saved when the two subarrays are *optimally* displaced each other, having significantly reduced the number of array elements.

This feature is substantially due to the fact that the array manifold associated to the proposed configurations of pairs of subarrays retains the same *slope* near the intersections with the signal subspace, when the two subarrays are optimally displaced. In fact, the loss of the estimation variance is limited by the effective aperture enlargement.

To demonstrate this circumstance, we employ the analytical expression of the estimation variance given in [1], particularized to the specific array manifold, as a function of the subarrays splitting distance. Moreover, by following the guidelines indicated in [2], we evaluate the angular distance between the estimated and the true signal–subspaces. Interestingly, the splitting distance yielding the minumum angular distance between the estimated and the true signal–subspaces does not coincides with the splitting distance yielding the minimum estimation variance. Following the idea presented in [3], where a linear prediction method is applied to a pair of subarrays where the *reference* subarray is constituted by a single sensor, we show that the *optimum* splitting distance is still obtained by using a generalized minimum prediction error variance criterion also when the reference subarray is formed by more than one sensor.

Finally, simulation results are provided to show the applicability of the MUSIC to the proposed configuration of pairs of subarrays.

2 Performance Analysis of MUSIC applied to subarrays

For reference, let us refer to a ULA of M sensors spaced d meters apart. A general configuration of a pair of subarrays consists in forming the first subarray with the first

0-8186-7576-4/96 $5.00

K_1 sensors and the second subarray with the last K_2 sensors at the other endpoint. Let us pose $K=K_1+K_2$. The subarrays distance is $\mathcal{D}=\Delta \cdot d$ where $\Delta=(M-K+1)$.

In other words, the subarray configuration is obtained by powering off $\Delta-1$ sensors in the middle of an M sensors ULA, so as to mantain the overall array aperture $(M-1)d$. The manifold associated to this sensors configuration is described by the steering vector

$$\mathbf{a}(\omega)=\left[1, e^{j\omega}\,;\cdots,\, e^{j(K_1-1)\omega}, e^{j(K_1+\Delta-1)\omega}\,;\cdots,\, e^{j(M-1)\omega}\right]^{\mathrm{T}}$$

where $\omega = 2\pi d \sin(\theta)/\lambda$, θ is the generic DOA[1] and λ is the wavelength.

When L sources are considered, the matrix of the steering vectors (sometimes referred to as the *mixing* matrix) is[2]

$$\mathbf{A} = [\mathbf{a}(\omega_1)\,;\cdots,\, \mathbf{a}(\omega_L)] \tag{1}$$

and the vector of the array sensors measurements $\mathbf{x}=[x_1\,;\cdots,\, x_{K_1}\; x_{K_1+\Delta}\,;\cdots,\, x_M]^{\mathrm{T}}$ is given by

$$\mathbf{x} = \mathbf{A}\cdot\mathbf{f} + \mathbf{w} \tag{2}$$

where $\mathbf{f} = [f_1\,;\cdots,\, f_L]^{\mathrm{T}}$ is the vector of $L < M$ independent, zero-mean, circularly complex, Gaussian distributed sources, and $\mathbf{w} = [w_1\,;\cdots,\, w_M]^{\mathrm{T}}$ is the vector of observation noises, circularly complex, zero-mean and Gaussian distributed, independent of the sources $\mathbf{f}$. Naturally, the observation model (2) is formally equivalent to the observation model drawn from the full array configuration; the mixing matrix $\mathbf{A}$ takes into account the actual form of the array manifold, *i.e.* how the sensors are located along the receiver.

The main advantage of using the subarrays configuration consists in mantaining the same end–to–end array aperture $(M-1)d$ while using K sensors instead of M, so allowing for a significant computational saving. This is paid with an estimate accuracy loss.

To evaluate the accuracy of the estimation carried out by using the MUSIC technique to the subarrays configuration, we start by recalling here the (approximate) expression of the mean squared value of the distance between the *signal subspace*, *i.e.* the range of the mixing matrix $\mathbf{A}$, and an estimate of the signal subspace obtained from the eigen–decomposition of the sample covariance matrix

$$\widehat{\mathbf{R}}_{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}(i)\cdot\mathbf{x}^{\mathrm{H}}(i)$$

where N is the number of available independent snapshots $\mathbf{x}(i)$.

Said α the angle between the subspaces, we report here the following espression of the mean squared value $\mathrm{E}\left\{(1-\cos\alpha)^2\right\}$ of the subspaces distance, drawn from [6] (with some rearrangements of terms), in the case of two uncorrelated sources and white noises

$$\mathrm{E}\left\{(1-\cos\alpha)^2\right\} \simeq \frac{1}{N}\frac{2(K-2)}{K^2}\frac{1}{\mathrm{SNR}^2} \cdot\left(\frac{1+K\cdot\mathrm{SNR}\cdot(1+x)}{(1+x)^2}+\frac{1+K\cdot\mathrm{SNR}(1-x)}{(1-x)^2}\right) \tag{3}$$

where $\mathrm{SNR} = (P_1+P_2)/2\sigma^2$ is the signal–to–noise ratio for sources with power P_1 and P_2 and white Gaussian noise with variance σ^2.

The parameter x depends on the array manifold as follows:

$$x = \sqrt{(1-A)+A\cdot|\phi|} \tag{4}$$

where $A=4P_1P_2/(P_1+P_2)^2$ depends only on the source powers and $\phi=\mathbf{a}^{\mathrm{H}}(\omega_1)\cdot\mathbf{a}(\omega_2)/K$ is the (normalized) scalar product of the steering vectors $\mathbf{a}(\omega)$ evaluated at the true DOA's ω_1 and ω_2, spaced $\delta\omega=(\omega_2-\omega_1)/2$ apart.

The correlation coefficient ϕ relates the angular closeness of the sources and the form of the array manifold, and it assumes the following form

$$\phi = \frac{1}{K}e^{j(K_1-1)\delta\omega}\cdot\frac{1}{\sin(\delta\omega)} \cdot\left(\sin(K_1\delta\omega)+\sin(K_2\delta\omega)e^{j(M+\Delta-1)\delta\omega}\right) \tag{5}$$

The absolute minima of $\mathrm{E}\left\{(1-\cos\alpha)^2\right\}$ are obtained[3] for $\phi = 0$, *i.e.* by choosing an array manifold in which the steering vectors associated to the sources are orthogonal. Despite of this fact, good signal subspace estimation does not results in a minimization of the estimation variance. This is due to the fact that the definition of the angular distance between subspaces relies on the *maximum* angle formed by vectors belonging to the subspaces, while the estimation variance also depends on the local slope of the array manifold measured at the intersection with the signal–subspace; sensible rotation of the estimated subspace is allowed, leaving the estimation variance quite unaffected.

This fact is also deduced by looking at the expression of the mean squared estimation error, reported in [1],

$$\mathrm{E}\left\{\left(\hat{\omega}_l-\omega_l\right)^2\right\} = \frac{\sigma^2}{2N}\frac{\sum_{k=1}^{L}\frac{\lambda_k}{(\sigma^2-\lambda_k)^2}\left|\mathbf{a}^{\mathrm{H}}(\omega_l)\mathbf{s}_k\right|^2}{\sum_{k=1}^{K-L}\left|\mathbf{d}^{\mathrm{H}}(\omega_l)\mathbf{g}_k\right|^2} \tag{6}$$

[1] In the sequel, also the parameter ω will be, improperly, referred to as DOA.

[2] Strictly speaking, the matrix $\mathbf{A}$ is parameterized by the vector $\vec{\omega} = [\omega_1\,;\cdots,\,\omega_L]^{\mathrm{T}}$, and it should be denoted as $\mathbf{A}(\vec{\omega})$. To avoid a too cumbersome notation, in the sequel we will omit the dependence on $\vec{\omega}$, writing simply $\mathbf{A}$.

[3] This is due to the increasing monotonic behaviour of (3) in the interval $|\phi| \le 1$ or, equivalently, $1-A \le x \le 1$.

where

$$\mathbf{d}(\omega) = \frac{\partial}{\partial \omega}\mathbf{a}(\omega)$$

is the vector of the derivatives of the components of the steering vector, the vectors $\mathbf{s}_k$ are the eigenvectors of the true covariance matrix spanning the signal–subspace, λ_k are the associated eigenvalues, and the vectors $\mathbf{g}_k$ are the eigenvectors spanning the noise–subspace, *i.e.*

$$\mathbf{R}_{\mathbf{x}} = \mathrm{E}\{\mathbf{x}\cdot\mathbf{x}^{\mathrm{H}}\} = \sum_{k=1}^{L}\lambda_k\cdot\mathbf{s}_k\cdot\mathbf{s}_k^{\mathrm{H}} + \sigma^2\sum_{k=1}^{K-L}\mathbf{g}_k\cdot\mathbf{g}_k^{\mathrm{H}}$$

For two uncorrelated sources the expression of the eigenvalues is

$$\lambda_{1,2} = \frac{K}{2}\left((P_1+P_2)\pm\sqrt{(P_1-P_2)^2+4P_1P_2|\phi|^2}\right)+\sigma^2$$

and (non-normalized) eigenvectors $\mathbf{s}_k$ are

$$\mathbf{s}_{1,2} = \mathbf{a}(\omega_1) + \frac{\lambda_{1,2}-KP_1}{KP_1\phi}\cdot\mathbf{a}(\omega_2) \qquad (7)$$

By carefully looking at (6), we see that the numerator depends on ϕ in the same way of (3), but the dependence of the denominator is not still clear. To better investigate the dependence on ϕ , (6) has to be put in a more suitable form.

To this end, let us consider two sources having the same power, *i.e.* $P_1 = P_2 = P$. After some algebraic manipulations, we obtain

$$\mathrm{E}\left\{(\hat{\omega}_l-\omega_l)^2\right\} = \frac{\mathrm{NSR}/K}{2N} \cdot \frac{(1-|\phi|^2)+\mathrm{NSR}/K}{(1-|\phi|^2)\mathcal{I}-\mathcal{J}^2-\frac{1}{4}\left|\frac{\partial}{\partial\,\delta\omega}\phi\right|^2-\mathrm{Im}\left\{\mathcal{J}\phi\frac{\partial}{\partial\,\delta\omega}\overline{\phi}\right\}} \qquad (8)$$

where NSR $= 1/\mathrm{SNR}$, overbar denotes complex conjugation, and

$$\begin{aligned} K\cdot\mathcal{I} &= 1+4+9+\cdots+(K_1-1)^2 \\ &\quad +(K_1+\Delta-1)^2+\cdots+(M-1)^2 \\ K\cdot\mathcal{J} &= 1+2+3+\cdots+(K_1-1) \\ &\quad +(K_1+\Delta-1)+\cdots+(M-1) \end{aligned}$$

We can see that now $\phi=0$ maximizes the numerator, while the denominator depends on the derivative of ϕ w.r.t. the DOA spacing $\delta\omega$. It is shown that the subspaces distance criterion cannot be used to determine a splitting distance which minimizes the estimation error .

To this purpose, we follow here an alternative approach based on the minimization of the variance of the prediction error, as indicated in [3] where the particular case of a subarray formed by one sensor only has been addressed in a linear prediction framework. A similar approach has been also employed in [5] in speech compression applications.

Specifically, denoting by

$$\mathbf{x}_1 = [x_1\,;\cdots,\,x_{K_1}]^{\mathrm{T}} \quad ; \quad \mathbf{x}_2 = [x_{K_1+\Delta}\,;\cdots,\,x_M]^{\mathrm{T}}$$

the measurements drawn from the first subarray and the second subarray, respectively, the linear prediction problem is solved by determining the matrix $\mathbf{P}$ which minimizes the the sum of the variances of the prediction errors

$$\mathbf{e} = \mathbf{x}_2 - \mathbf{P}\cdot\mathbf{x}_1 \qquad (9)$$

The minimum value is readily found as

$$\sigma_e^2 \stackrel{def}{=} \mathrm{E}\{\mathbf{e}^{\mathrm{H}}\mathbf{e}\} = \mathrm{Trace}\left\{\mathbf{R}_2 - \mathbf{R}_{21}^{\mathrm{H}}\mathbf{R}_1^{-1}\mathbf{R}_{21}\right\} \qquad (10)$$

where $\mathbf{R}_1 = \mathrm{E}\{\mathbf{x}_1\cdot\mathbf{x}_1^{\mathrm{H}}\}$ is the covariance matrix of the measurements of the first subarray, $\mathbf{R}_2 = \mathrm{E}\{\mathbf{x}_2\cdot\mathbf{x}_2^{\mathrm{H}}\}$ is the covariance matrix of the measurements of the second subarray and $\mathbf{R}_{21} = \mathrm{E}\{\mathbf{x}_2\cdot\mathbf{x}_1^{\mathrm{H}}\}$ is the cross–covariance matrix between the two subarrays. The absolute minumum of σ_e^2 is found by sistematically varying Δ.

3 Conclusion

To show how the results of the previous section can be used in practical applications, simulation results are reported in fig.1, where the mean squared errors relative to the Root–MUSIC estimation of the DOA $\omega_1 = -\omega_2 = 3$ degrees, when a subarray configuration with $K_1 = K_2 = 5$ is considered, is plotted vs. the subarrays splitting distance Δ. For comparison purposes, the results relative to the full array with $M = K_1+K_2+\Delta-1$ sensors are also shown. The array spacing is half a wavelength $2d = \lambda$, three values of SNR $= -5,\ 7,\ 15$ dB are considered, the number of snapshots is $N = 100$ and averaging over 100 MonteCarlo runs have been carried out.

We see that optimum performances of the subarrays configuration are obtained for $\Delta \simeq 15$ for all the SNR values and the given source spacing $\delta\omega = 6$ degrees, corresponding to performance of a full array of $M = 24$ sensors, while only $K = 10$ sensors are used in the subarrays configuration.

In figs. 2 and 3, we report the minimum error prediction variance (10) and the subspace distance (3) as functions of the subarrays splitting distance Δ, respectively. We see that the optimum splitting distance is correctly indicated by the minimum prediction error criterion, while the subspace distance shows a kind of a *opposition–of–phase* behaviour in indicating the optimum splitting distancet. This is due to the lack of dependence on the derivative of ϕ w.r.t. the DOA spacing $\delta\omega$, which is inherently taken into account in the minimum prediction error criterion.

This fact does suggests some opportunities of practical use of the ideas here presented. For instance, this computational saving technique can be used in tracking pairs of sources after a full array discovery stage by applying the following two step procedure:

- fixed the subarrays aperture K_1 and K_2, search over Δ for the minimum prediction error. This implies inversion of a reduced order covariance matrix $\mathbf{R}_1$, which can be suitably carried out using Levinson recursion in the case of a uniformly linearly spaced subarray $\mathbf{x}_1$;
- apply the root-MUSIC techniques (or other techniques based on subspaces decompositions) to the KxK covariance matrix of the subarrays configuration.

The drawback is constituted by the increased ambiguity of the subarrays configuration, which is clearly show in fig.1 for Δ large enough. In essence, this technique can be defined as a MUSIC *interpherometric* method.

This technique can be used also in multiple invariance ESPRIT context. This is matter of current investigation, along with a more detailed theoretical analysis.

References

[1] P. Stoica, A. Nehorai, "MUSIC, maximum likelihood, and Cramer–Rao bound", *IEEE Trans. Acous. Sp. Sig. Proc.*, vol. 37, no. 5, May 1989.

[2] H. Wang, M. Kaveh, "On the performance of signal–subspace processing—Part I: Narrow–band systems", *IEEE Trans. Acous. Sp. Sig. Proc.*, vol. 34, no. 5, October 1986.

[3] Y. Yoganandam, V.U. Reddy, C. Uma, "Modified linear prediction method for directions of arrival estimation of narrow–band plane waves", *IEEE Trans. Ant. Prop.*, vol. 37, no. 4, April 1989.

[4] R.O. Schimdt, "Multiple emitter location and signal parameter estimation", *Proc. RADC Spectral Estimation Workshop*, Rome (NY), 1979.

[5] C. Becchetti, G. Jacovitti, "Improved spectral analisys of near periodic signals with long–term prediction", *Quinzième Colloque sur le Traitement du Signal et des Images*, Juan–Les–Pins, September 1995.

[6] M. Kaveh, H. Wang, "Treshold properties of narrow–band signal–subspace array processing methods", in **Advances in Spectrum Analysis and Array Processing**, vol.II, S. Haykin ed., Prentice–Hall, 1991.

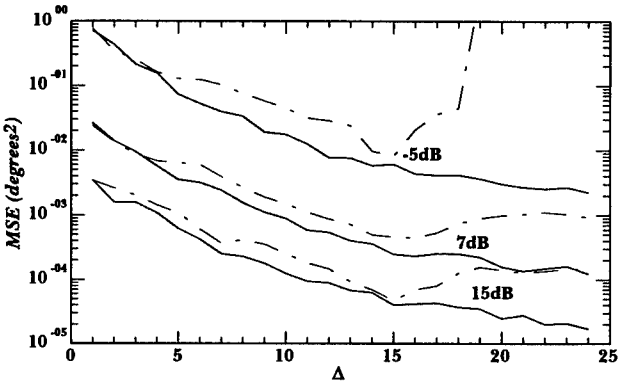


Figure 1: ***Comparison of performance of MUSIC applied to a full array (continuous curves) and to a subarrays configuration (dashed curves) as a function of the splitting distance Δ and for various SNR values.***

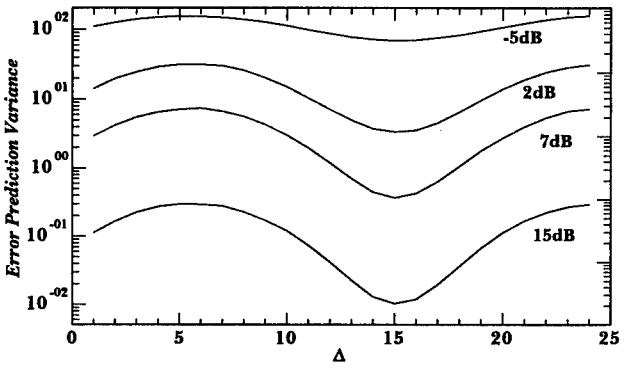


Figure 2: ***Error prediction variance vs. the splitting distance.***

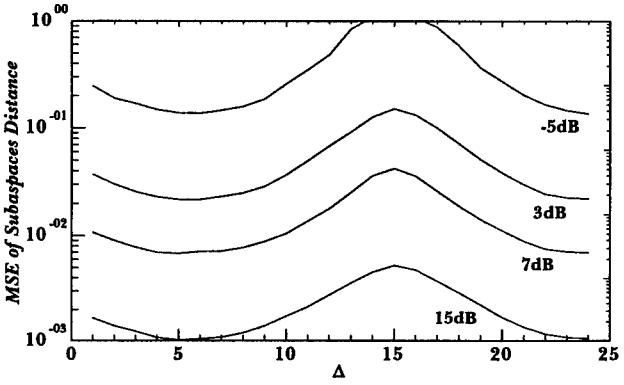


Figure 3: ***MSE of the distance between the true and the estimated signal subspaces vs. the splitting distance.***

Increasing Beamsteering Directions Using Polyphase Decomposition

Daniël W.E. Schobben and Piet C.W. Sommen
Department of Electrical Engineering, Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
D.W.E.Schobben@ele.tue.nl

Abstract

In many implementations of digital delay and sum beamforming, a sample rate much higher than the Nyquist rate is used. This allows for many synchronous beamsteering directions. Severe demands are made upon the analogue to digital converters however. Several methods have been proposed for reducing the sample rate required. These methods incorporate the delays that are needed for beamforming in time domain [3],[4] or in frequency domain [5]. A more efficient method for implementing a time-domain delay and sum beamformer using polyphase decomposition is presented in this paper. This method results in significant computational savings when the desired angular resolution is high compared to the number of sensors used and the number of simultaneously formed beams.

1. Introduction

Conventional continuous-time beamformers delay all sensor outputs so that propagation delays are cancelled and the sensor outputs can be combined coherently. In a discrete-time beamformer, these delays are performed digitally. Using discrete time delays only allows for delaying over an integer multiple of the sampling time period. Therefore, the number of synchronous beam-pointing directions is small for low sample rates, resulting in a poor angular resolution. To illustrate this, it is shown in Section 2 that a linear array sampled at v times the Nyquist rate can only be steered to $1 + 2v$ synchronous angles. A signal arriving from a non-synchronous direction can be received by steering the beam to the most nearby synchronous angle or by rounding the delays needed for beamforming to the delays available. Both methods introduce severe distortion and poor spatial discrimination for small v. In Section 3 the concept of interpolation beamforming is discussed. This technique uses interpolation, so that the sampling rate is increased artificially. In this way, delays are obtained which are a fraction of the unit delay [2]. In Section 4 an efficient method is presented for the implementation of the interpolation beamformer using polyphase decomposition. The resulting complexity is discussed in Section 5 and a numerical example shown in Section 6.

2. Linear Sensor Array Beamforming

Although the method to be presented can be applied to all array geometries, an example of the steering capabilities of a linear array is discussed. For a linear array the anticipated propagation delay of a flat wave plane from the first sensor to the i^{th} sensor equals

$$\tau_{\max} - \tau_i = \frac{id \sin\theta}{c}, \qquad (1)$$

where d is the sensor inter-distance, θ is the direction of arrival (DOA) relative to broadside (the direction perpendicular to the line-array), c is the wave plane propagation velocity and $\tau_{\max}$ is the propagation delay of the wave plane between the first and the last sensor. The wave plane incident to the linear array is depicted in Figure 1. The sensor outputs

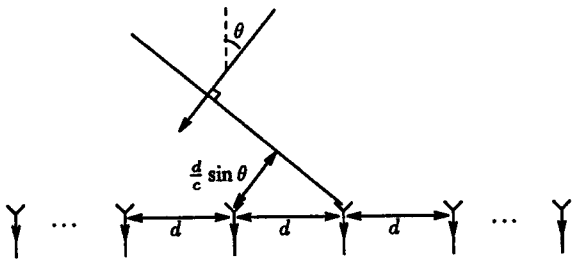


Figure 1. Wave front incident to a linear array

are sampled at the rate $f_s = \frac{1}{T_s}$ and consequently the beam can only be steered to the angles which yield a difference in propagation delay of uT_s seconds between neighboring sensors, with u an integer. This delay is cancelled using

0-8186-7576-4/96 $5.00

time-delays $uT_s = \frac{d \sin\theta}{c}$. The beam can thus be steered to

$$\theta_u = \sin^{-1}\left(\frac{cuT_s}{d}\right). \qquad (2)$$

A common choice for d that prevents spatial aliasing is $d = \frac{\lambda_0}{2}$, λ_0 being the minimum wavelength of the signal to be received. For this sensor inter-distance and a sample rate equaling $f_s = 2vf_0$, it follows from (2) that $|u| \leq v$. There exist $1 + 2\lfloor v \rfloor$ different u that obey this equation, so the beam can be steered to $1 + 2\lfloor v \rfloor$ different angles. For example, when the sample rate for equals $4f_0$ ($v = 2$), it follows from (2) that the beam can only be steered to $0, \pm 30^o$ and $\pm\ 90^o$. A higher angular resolution can be achieved by interpolating the sampled data. When $v = 2$ and more than 5 synchronous directions are desired, the data can be interpolated by a factor M. This is depicted in Figure 2 for $M = 4$. The solid lines indicate the sampled

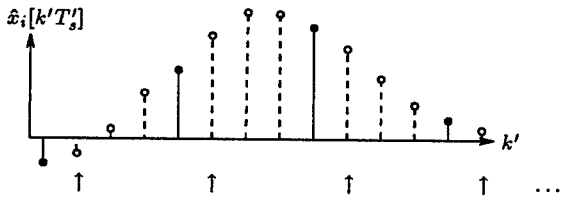


Figure 2. Data Interpolation

data and the dashed lines indicate the interpolated data. A delay of $\frac{3T_s}{4}$ for example can now be achieved by selecting the interpolated samples indicated with an $\uparrow$ in Figure 2. Interpolating with a factor $M = 4$ now allows for beam-steering to $0, \pm 7.18^o$, $\pm 14.5^o$, $\pm 22.0^o$, $\pm 30.0^o$, $\pm 38.7^o$, $\pm 48.6^o$, $\pm 60.0^o$, $\pm 90.0^o$. Clearly, only one of every M interpolated samples is used for beamforming.

3. Interpolation Beamforming

The interpolation process for a single beam is depicted in Figure 3. First the sensor data is sampled at a rate equal to or exceeding the Nyquist sampling rate. The i^{th} sampled sensor output $x_i[kT_s]$ is zero padded to obtain $\tilde{x}_i[k'T_s']$

$$\tilde{x}_i[k'T_s'] = \begin{cases} x_i\left[\frac{k'}{M}T_s\right], & \text{for } k' = 0, \pm M, \pm 2M, \ldots \\ 0, & \text{otherwise} \end{cases}, \qquad (3)$$

with $T_s' = \frac{T_s}{M}$. Then the $\tilde{x}_i$ are filtered with the interpolation filter H to obtain the $\hat{x}_i$

$$\hat{x}_i[k'T_s'] = \sum_{l=0}^{L-1} h[lT_s']\tilde{x}_i[(k'-l)T_s']. \qquad (4)$$

This filter is a Finite Impulse Response filter with impulse response $h[k'T_s']$. The $\hat{x}_i$ are delayed over p_iT_s' seconds, to compensate for the anticipated propagation delays, where the p_iT_s' are equal to the τ_i. Then down-sampling is used to obtain

$$x_i'[kT_s] = \hat{x}_i[(kM - p_i)T_s']. \qquad (5)$$

The beamformer output $y[kT_s]$ is obtained by summing the shaded x_i'. Shading means multiplying the x_i' with weights to enhance the angular discrimination. To simplify notation, the shading is not mentioned explicitly in figures and equations. Multiple beams can be formed from the interpolated sensor outputs $\hat{x}_i$ without performing additional multiplications.

Pridham and Mucci [3] argued that the scheme of Figure 3 is equal to the scheme in Figure 4 for the case that only one beam is formed. This can be seen by interchanging the filter H and the delays p_iT_s' and placing the filter H and the down-sampling in Figure 3 after the summation. This is allowed when all filters are identical, linear and time invariant. Furthermore, the filter and the down-sampling may be combined to reduce complexity. It will be shown in Section 5 that the complexity of the technique proposed in the following section is lower than that of the scheme in Figure 4 for high angular precision beamformers. Note that forming multiple beams is not possible without performing additional multiplications with this scheme.

4. Polyphase Decomposition

In this section an efficient implementation of the interpolation beamformer is presented. First, consider the data processing in Figure 3 for i^{th} sensor only. This is depicted in Figure 5(a). Here the delay p_iT_s' is interchanged with interpolation filter H. This is allowed since H is linear and time invariant. In Figure 5(b) filter H is decomposed into the filters $H_0, H_1 \ldots H_{M-1}$ using polyphase decomposition [1]. The impulse responses of the H_j can be calculated from the impulse response of H according to

$$h_j[k'T_s'] = \begin{cases} h[(k'+j)T_s'], & \text{for } k' \in \kappa \\ 0, & \text{otherwise} \end{cases}, \qquad (6)$$

for $j = 0, 1, \ldots, M-1$ and $\kappa = \{0, M, \ldots, M\left\lfloor\frac{L-j-1}{M}\right\rfloor\}$. Down-sampling these filter outputs is equivalent to down-sampling the data and then filtering with H_j', as depicted in Figure 5(c). The impulse responses of the H_j' are given by

$$h_j'[kT_s] = h_j[kMT_s'], \qquad (7)$$

for $k = 0, 1, \ldots, \left\lfloor\frac{L-j-1}{M}\right\rfloor$. Up-sampling with a factor M, delaying over p_iT_s' and down-sampling with a factor M is equal to delaying over $\frac{p_i}{M}T_s$ if p_i is an integer multiple of

M, and equal to zero otherwise. Therefore Figure 5(c) can be interpreted as choosing the filter H'_q and delaying over rT_s, where

$$q = M\left\lceil \frac{p_i}{M} \right\rceil - p_i \,, \qquad r = \frac{q+p_i}{M} = \left\lceil \frac{p_i}{M} \right\rceil . \tag{8}$$

The resulting scheme is depicted Figure 5(d). The combination of the delay of rT_s and the sub-filter H'_q represents an approximation of the desired delay $p_i T'_s$. The sub-filter must be of sufficient length to guarantee a good approximation of the desired delay. In contradiction to this demand, a long sub-filter requires many multiplications per second, and introduces a long beamformer delay.

5. Comparison of Computational Complexity

As a measure of complexity, the number of multiplications per seconds of the interpolating filter is considered. The filter H is assumed to be of length $L = wM$ throughout this section, with w integer. Although this is not necessary, it gives more insight in the calculation of the complexity. The proposed beamformer is compared with the beamformers of Figure 3 and 4.

In the scheme of Figure 3, N filters of length L are calculated at a rate f'_s, with N the number of sensors. Using $f'_s = Mf_s$ and taking advantage of the sparse input data of the interpolation filters, this yields a complexity of LNf_s multiplications per second. This complexity is independent of the number of beams to be calculated, and is therefore efficient when a large number of beams is required. Furthermore, assuming that the filter H is a linear phase filter, the number of multiplications per second can be reduced by approximately a factor 2, yielding a complexity of $\frac{LNf_s}{2}$ multiplications per second.

Filter H in Figure 4 is calculated for each beam at a rate f_s, since only one of every M samples is needed. The number of multiplications per second equals $LN_B f_s$, with N_B the number of beams to be calculated. Again, assuming that the interpolation filter is a linear phase filter, the resulting number of multiplications per second equals $\frac{LN_B f_s}{2}$.

The scheme in Figure 4 is more efficient than the scheme in Figure 3 if and only if the number of beams to be calculated is smaller than the number of sensors ($N_B < N$).

In the proposed beamformer, sub-filters H'_q are of length w. For each beam, N sub-filters are calculated at a rate f_s. The total number of multiplications per second equals $wNN_B f_s = \frac{LNN_B f_s}{M}$. In general, it is not possible to exploit the linear phase property of the interpolation filter to reduce the complexity further. When $M < 2N_B$ and $N < N_B$ interpolating all sensor data as in Figure 3 is most efficient. Combining all interpolation filters as in Figure 4 is most efficient when $M < 2N$ and $N > N_B$. Thus the proposed method outperforms its alternatives when a high angular resolution is required ($M > 2N$ and $M > 2N_B$). This scheme has a gain in computational complexity over its alternatives of $\frac{M}{2N}$ and $\frac{M}{2N_B}$ respectively. The quality of the delays formed depends on the sub-filter length $w = \frac{L}{M}$. When both the filter length L and the interpolation factor M are increased with the same factor, the number of beam-pointing directions further increases while the computational complexity does not increase for the proposed method. Consequently, the angles for which beams can be formed can now be chosen with arbitrarily precision while maintaining the same amount of multiplications per second. In practice, the filter length L is limited however since $\frac{L}{2}$ tab weights are stored into a finite amount of memory. For the two alternative methods the computational complexity does however increase proportionally with the filter length L.

6. Example of Computational Complexity

Next, an example is given to show that the conditions for the proposed method to be more efficient than its alternatives are easily met. Consider sub-filters of length $w = 10$, $N = 7$ sensors, $M = 20$ (41 different synchronous angles) and $N_B = 5$ (five beams are formed). The main lobes of the unshaded beam-patterns corresponding to the resulting synchronous beam-pointing directions are depicted in Figure 6 for θ in between 0° and 90°. For negative angles, the figure is symmetrical. The figure shows that there indeed is a need for a high M to exploit the best possible angular discrimination. However, when M is chosen much larger, the angular discrimination no longer improves, as the successive beams merely overlap. In practice, M will be in between 15 and 40 for a 7 sensor array which is sampled at the Nyquist rate, and the proposed method outperforms the alternatives discussed. The complexity for the proposed beamformer equals $wNN_B f_s = 350 f_s$ multiplications per second for this example. The alternate schemes of Figure 3 and 4 require $\frac{LNf_s}{2} = 700$ and $\frac{LN_B f_s}{2} = 500 f_s$ multiplications per second respectively. A significant efficiency gain is thus obtained.

7. Conclusions and Future Research

A new method using polyphase decomposition was proposed for reduced complexity interpolation delay and sum beamforming. Significant computational savings are reported for beamformers with a high angular discrimination.

In future research the polyphase equivalent scheme will be used to study relations between interpolation beamforming and other broadband array processing techniques. The use of an adaptive algorithms to track moving sources us-

ing delay and sum beamformers will also be considered in future research.

References

[1] A.N. Akansu and R.A. Haddad. *Multiresolution Signal Decomposition.* London : Academic Press, 1992.

[2] T.I. Laakson et al. Splitting the unit delay. *IEEE Signal Proc. Mag.*, 13(1):30–60, Jan. 1996.

[3] R.G. Pridham and R.A. Mucci. A novel approach to digital beamforming. *J. Acoust. Soc. Am.*, 63:425–434, Feb. 1978.

[4] R.W. Schafer and L.R Rabiner. A digital signal processing approach to interpolation. *Proc. IEEE*, 61:692–720, Jun. 1973.

[5] M.E. Weber and R. Heisler. A frequency-domain algorithm for wideband coherent signal processing. *J. Acoust. Soc. Am.*, 76:1132–1144, Oct. 1984.

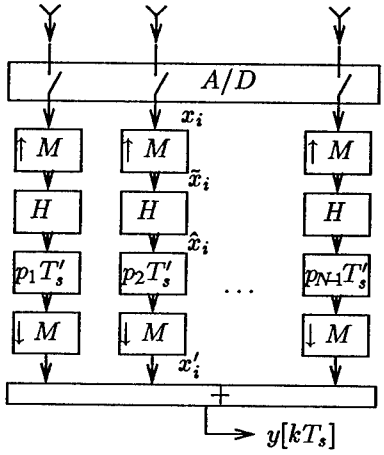


Figure 3. Interpolation beamformer

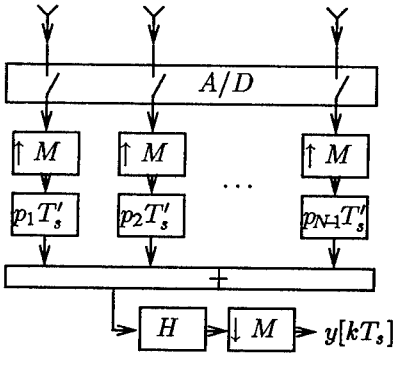


Figure 4. Beamformer equivalent

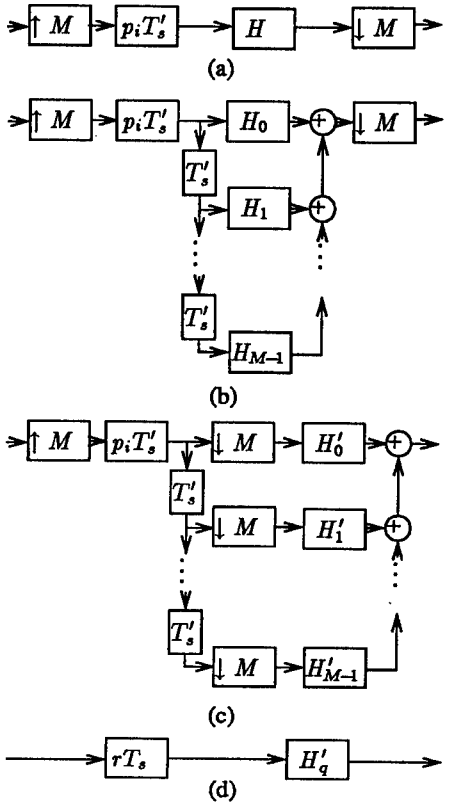


Figure 5. Interpolation filter (a), polyphase decomposed filter (b), polyphase equivalent (c) and compacted polyphase equivalent (d).

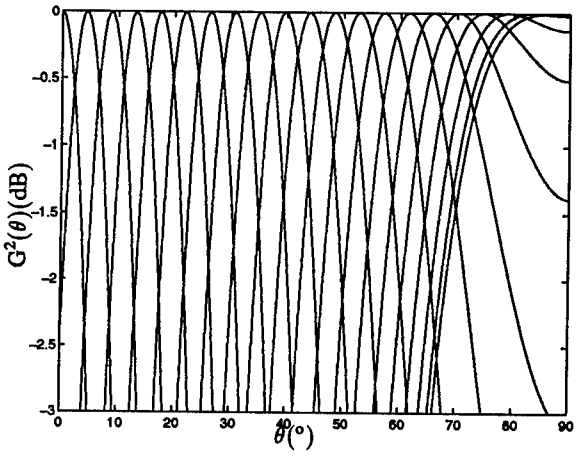


Figure 6. Beampointing patterns for all synchronous directions, M = 20 and N = 7

Plenary

Blind Fractionally-Spaced Equalization of Digital Cable TV

J. Treichler, C. Johnson, Jr.

BLIND FRACTIONALLY-SPACED EQUALIZATION OF DIGITAL CABLE TV

John R. Treichler
Applied Signal Technology
400 W. California Ave.
Sunnyvale, CA 94086 USA
(jrt@appsig.com)

C.Richard Johnson, Jr.
School of Electrical Engineering
390 Frank H.T. Rhodes Hall
Cornell University
Ithaca, NY 14853 USA
(johnson@anise.ee.cornell.edu)

Abstract

This paper describes the design requirements for a QAM demodulator chip recently developed to be part of a settop convertor for digital cable television. The chip demodulates 64- and 256-QAM signals at a maximum bit rate of 44 Mb/s and uses blind acquisition techniques so that no training or pilot signals need be sent by the transmitter.

1 Introduction

The desire to send many bits of data per Hertz of transmission bandwidth has caused the development of sophisticated communications systems using *quadrature amplitude modulation (QAM)*. First introduced for voiceband modems [1] the technology was then applied to microwave radio relay systems [2]. Its success in those applications has led to great interest in its use for other communication situations in which economic or regulatory considerations limit the available transmission bandwidth. An important example of such an application is the wireless and cable distribution of digital television [3]. This paper describes how digital transmission is used for cable television distribution and how the characteristics of a cable system affect the design of a suitable demodulator.

[0]CRJ's research on blind equalization is currently supported in part by NSF Grant MIP-9509011 and Applied Signal Technology.

2 Background

Figure 1 shows the block diagram of a digital communications system. The input data is applied to the modulator and transmitter, which convert the data stream into a bandlimited analog waveform and frequency-translate it into the frequency band appropriate for transmission. As the signal propagates to the receiver it is delayed, attenuated, and sometimes distorted in a frequency-selective manner. These effects, on which we will elaborate shortly, are modeled in the block diagram as the *propagation channel*. The receiver accepts the channel output, plus noise and interference inadvertantly present at the receiver input, and attempts to recover the input data sequence.

In the particular case of digital cable television transmission, Figure 1 gives way to the system shown in Figure 2. Compressed video, audio, telephony, and even other data services are combined into a composite data stream and modulated onto a carrier wave. Many of these, plus, possibly, older analog television signals, are summed together for transmission and distribution. Older systems do the transmission on coaxial cables only. Newer systems use both fiber optic transmission and coaxial cable (as shown in Figure 2), while the newest promise to send the signals directly to the customer premises on fiber. Once in the customer premises the signal is commonly split and distributed to many devices, including VCRs, television sets, and, in the future, cable modems.

More detail on the "headend" is shown in Figure 3. The video and audio for a particular tele-

0-8186-7576-4/96 $5.00 © 1996 IEEE

vision source are digitized and compressed. The resulting output data rate depends on the type of compression used and the desired fidelity of the received image and sound. Quality comparable to high-SNR NTSC transmission can be obtained using MPEG-2 compression, yielding an average bit rate of about 6 Mb/s. High definition television (HDTV), with its larger screen size and greater resolution, requires about 25 Mb/s. Since the modulation anticipated (more on this shortly) can carry a raw data rate of 30 to 40 Mb/s, this permits several digitized video/audio sources to be multiplexed together on a single "digital signal". This multiplexing can be done deterministically, that is, by giving each of the sources a fixed bit rate allocation, or it can be done dynamically, allowing the number of sources, their quality, and the types of sources to be managed by the headend. For example, one HDTV and two normal resolution TV sources might share the bandwidth, or, conversely, a sports program with a high degree of motion might be allowed to use some of the bits allocated to a video signal with a more static image.

Once the 30 to 40 Mb/s stream of multiplexed and possibly encrypted signals is put together, forward error correction bits are added and the composite stream is modulated onto a carrier. These modulated signals have a bandwidth less than 6 MHz so that they can be frequency-division multiplexed (FDM) onto the cable transmission medium. By adhering to the 6 MHz spacing, compatibility is maintained with older analog TV transmission systems. Note also that the 30 to 40 Mb/s carried on the modulated signal need not be used completely (or even partially) for television. Such a "data pipe" is usually refered to as a "cable modem", and can be used for a variety of services, including high-speed Internet connectivity from a server to a computer at the home or office.

In comparing Figure 1 with Figure 2 we see that the *propagation channel* includes all of the cable distribution equipment from the modulator output to the demodulator input in the destination device, and therefore includes all up- and downconverters, bandpass filters, combiners, trunk amplifiers, coaxial cable runs, and splitters. How these devices and equipment disrupt the signal's transmission can be understood after a discussion of the method by which the digital data is prepared for transmission over the analog medium.

Modern bandwidth-efficient transmission of digital data is based on the concept of sending pulses [4]. The input data is partitioned into sets of N bits and those bits are then used to determine the phase angle and peak amplitude of the pulse. The pulse shape itself is chosen to ensure a bandlimited signal spectrum. The receiver is designed to determine the amplitude and phase of each incoming pulse, determine which of the 2^N possibilities has been sent, and then report out the corresponding N bits. If the pulses are transmitted at the *symbol* or *baud rate* of f_B symbol per second, then the transmisssion system can carry $N \cdot f_B$ bits per second.

The effect of the cable transmission plant is to disperse the transmitted pulses in time. Its effect on a QAM signal is often assessed by looking at the signal's *constellation*. This is an overlay of many received symbol measurements. In the absence of noise, interference, and dispersion, and with perfect estimation of the signal's amplitude, carrier, and timing, the received measurements from a 64-QAM signal should look as they do in Figure 4(a). The presence of dispersion alone is sufficient to produce the degradation seen in Figure 4(b). In the absence of additive noise and receiver imperfections, the displacement between an actual received constellation point and the transmitted point shown in Figure 4(a) is a combination of the channel dispersion's effect on the particular pulse being considered and the intersymbol interference (ISI) induced by the channel on the adjacent pulses. Some of the received symbols are displaced sufficiently that the nearest neighbor decision rule makes errors. Because of the potential for frequent errors from this source, the demodulator requires an adaptive equalizer of some type to compensate for the effects of the cable plant's dispersion.

In further comparing Figure 1 with Figure 2 we see that the noise and interference includes all noise introduced by the active components of the system plus the interference produced inside the system and received from outside it. The noise is usually controlled by careful design and maintenance of the system. The interference, usually refered to as *ingress*, is combatted by minimizing any intermodulation distortion within the system and by ensuring good maintenance of the system to prevent strong externally generated signals, such as from radio or broadcast television, from entering the distribution system.

3 The Demodulator's Requirements

In light of this background, the requirements for the demodulator can be enumerated:

- The modulator/demodulator pair must reliably carry as much data as possible over currently available cable systems. The use of QAM on a 6 MHz channel limits the Baud rate to about 5 MHz. The noise floor present in a well-engineered conventional cable system limits the QAM constellation size to about 256. For poorer systems, a constellation size of 64 might be used to add some "system margin" at the expense of 25% of the available transmission rate. Thus, the demodulator is required to handle up to 256 QAM with a maximum baud rate of 5.5 MHz or so.

- The demodulator must operate in a non-cooperative manner, that is, it should not need any special training or synchronization from the transmitting modulator. Further, the user should be able to change channels rapidly ("channel surf") without subtantial re-acquisition delays being introduced by the demodulator.

- The demodulator needs to be cheap and to operate with other cheap components.

- Finally, the demodulator must handle the signal impairments to which it may be subjected. These include both the interference (intermodulation distortion and "ingress") and the signal dispersion introduced by the distribution plant itself. Conventional demodulator design cannot inexpensively deal with large amounts of interference and therefore these are traditionally handled by vigilant system maintenance. Channel dispersion, however, is a fundamental characteristic of a distribution plant and the demodulator must compensate for it. In order to reach a suitable specification for it, however, we must first determine the degree of dispersion present in cable TV systems.

4 Characterizing The Cable Television Propagation Channel

While modeling of the cable propagation channel can and has been done analytically, the approach taken here is to measure it in real cable television systems. We first describe the method employed and then the results.

In practical circumstances the propagation characteristics of the channel between a transmitter and receiver are not known *a priori*. Further, a one-time calibration of a channel's characteristics is not useful since channels are known to vary with time owing to influences from environmental and manmade factors. To deal with this time variation it is useful to have channel modeling techniques which can use "signals of opportunity" to probe the channel to be analyzed. The method used here, first described by Gooch and Harp [6], uses a demodulator to obtain symbol estimates from a PSK or QAM signal of opportunity and then uses these symbols along with the received signal itself as inputs to a channel modeller. This scheme is shown in Figure 5. The key to this technique's success is the use of a *blind equalizer* in the demodulator to "open the eye" enough for the demodulator to initially acquire the signal. Once acquisition has occured, the demodulator begins to use its own symbol decisions as the *desired input* to an LMS-directed equalizer update algorithm. (See Wolff, Treichler, and Gooch [10] for an early description of such a demodulator.) These symbol decisions are, of course, the same regenerated symbols needed as one of the inputs to the modeling stage.

Gooch and Harp [6] used a LMS-directed FIR adaptive filter to estimate the pulse response of the propagation channel. The filter's input is the stream of regenerated symbols, interpolated with alternate zeros to create a *fractionally-sampled* input rate of $2f_B$, where f_B is the symbol or Baud rate of the received signal. The *reference* or *desired* input to the adaptive modeller is a version of the input signal delayed to compensate for the processing delay of the demodulator. The LMS algorithm is used to adapt the coefficients of the complex-valued filter pulse response. The convergent solution is well known to closely approximate the least-squares fit between the actual channel and the model. The error signal $e(k)$ contains unmodeled components, misadjustment noise, and receiver noise. In passing it should be noted that this error signal can be

spectrum analyzed to reveal the presence and characteristics of additive signal impairments such as cochannel interference [6]. Use of this approach to identify *ingress* into a cable system will be discussed in Section 6.

An example of the result of this modeling procedure is shown in Figure 6. The power spectrum of a 64-QAM, 5.1 MBaud signal appears in Figure 6(a). Adjacent to it is the power transfer function of the estimated channel. This was obtained by first developing an FIR model of the channel pulse response, as described above, and then computing the log magnitude square of the FFT of that complex-valued pulse response shown in Figure 7. Note the close correspondence of the channel shaping between the received spectrum and that of the model.

By inspecting the log magnitude of the estimated pulse response in Figure 7, we can see that the channel does not conform to a simple two- or three-ray specular model but in fact the received signal is the combination of many delayed and scaled versions of the transmitted symbol stream. A more detailed examination of many such channel estimates indicates that the dispersion can be broken into two classes, "microreflections" and "macroreflections". The macroreflections have large amplitude compared with the transmitted signal and have relative delays on the order of microseconds, indicating a strong reflection from the end of a long improperly terminated stub. (Recall that the round-trip delay on a coaxial cable is about 18 microseconds/mile.) The microreflections are multitudinous but small in amplitude, stemming from a large number of lower level reflections on short cable sections within the system. The macroreflections must be found by the maintence crews and removed, since building the demodulator to compensate for them is uneconomical. The microreflections, however, are a fact of life even in a well-designed, well-maintained system and must be accommodated by the demodulator. Examination of a large number of the channel models of the type seen in Figure 7 shows that a reasonable estimate of the maximum delays seen for a cable system's microreflections is 2 to 3 microseconds. A database (to be resident in the National Science Foundation's Signal Processing Information Base (SPIB) at Rice University and linked to http://www.ee.cornell.edu/faculty/RJohnson.html) includes a representative sample of the received signals used to draw these conclusions.

Given this estimate for the maximum delay spread of 3 microseconds for the cable propagation channel, how long does the demodulator's equalizer need to be? This question has been recently addressed in [9], which discusses the recent technical result that a fractionally-spaced equalizer need be no longer than the maximum expected delay spread of the channel. In light of this result the length of the fractionally spaced equalizer should be at least 16 symbols long (so the data rate of 5.1 Mbaud < 16 symbols / 3 microseconds).

5 The Demodulator Design

Many different approaches have been used to design a demodulator for digital signals. An indication of the choices available in this design process are shown in Figure 8. In general the demodulator must (1) bandpass filter the incoming signal, (2) adjust the input signal amplitude, (3) estimate and remove any carrier component, (4) equalize the channel's dispersive effects, (5) "slice" the input signal to obtain pulse amplitude and phase measurements, (6) decide which pulse amplitude and phase was actually transmitted, and (7) convert that decision into the associated bit pattern. Demodulators for digital cable transmission incorporate forward error correction as well.

Even though it is only one component of the demodulator, the adaptive equalizer's design takes on special importance for three reasons: (1) its performance is crucial to the goal of maximizing the transmission rate through the dispersive channel, (2) it is the most complicated of all the demodulator's components, and (3) it consumes a large fraction of the computation needed to implement the complete demodulator. Amplifying on this third point, it is not unusual for the adaptive equalizer to consume more than 80% of the multiply/add cycles needed to demodulate a 256-QAM signal. Given that, it becomes an important design consideration to limit the length of the equalizer to only that required to handle adequately the range of propagation channels expected to be encountered.

With an eye toward minimizing and simplying the computation needs, early demodulators used so-called *T-spaced* equalizers [4]. After filtering, gain control, and carrier removal, the input signal was sampled once per symbol (pulse). The timing of this sampling clock was adjusted so that the sam-

ples were taken at the "top dead center" of the received pulses. These samples then entered the equalizer's tapped delay line filter. A linear combination of them was fed on to the measurement and comparison stages. Error measurements made in the decision circuit were fed back to the equalizer's adaptation algorithm to optimize the choice of filter weighting coefficients.

While theoretically reasonable (i.e., "one sample per pulse") and computationally desirable, practical design of high speed modems has gravitated away from T-spaced equalizers and toward the use of *fractionally-spaced* equalizers (FSE), so called because the equalizer taps are closer together in time than the symbol interval T. Equivalently, and perhaps more intuitively, this means that the input to the equalizer is sampled faster than the symbol rate f_B. The output rate is still at the symbol rate, making the FSE a decimating or even resampling filter.

If the temporal spread of the equalizers are held to the same value, then the FSE obviously consumes more computation than a T-spaced design. Why then use them? The answer is that even though they require more computation, they simplify the rest of the demodulator's design and allow it to work at virtually theoretical levels. The principal reason for this is that even though the pulses arrive at rate f_B, the actual bandwidth of the signal is somewhat larger, typically 10 to 40% higher. As a result, sampling the conditioned input at the rate of f_B Hz is not enough to satisfy the Nyquist theorem. While not important if all parameters of the signal were known, the fact that the input signal must be processed to extract timing and carrier information means that sampling at f_B is not fast enough. There are also some curious signal cancellation effects that arise when the signal components alias into a band of only f_B Hertz.

The actual input rate to the FSE is usually governed by a variety of hardware considerations. The rate must be high enough to satisfy the sampling theorem, but lowering the rate reduces the computational requirements. The most common choice is to sample the conditioned input signal at exactly twice the symbol rate f_B, making the filter tap spacing equal to $\frac{T}{2}$, half of the symbol spacing. The resulting equalizer thus decimates its input by a factor of 2, producing one output for every two input samples. It is not uncommon to operate at a fractional rate either. The demodulator chip to be described shortly uses a $\frac{5T}{6}$ design, in which a digital timing recovery circuit and resampler supplies complex-valued samples into the equalizer at a rate only 20% higher than the symbol rate.

In response to these requirements described in Sections 3 and 4, the QAM demodulator chip pictured in [9] was developed. The demodulator chip fits into a settop convertor design of the general type shown in Figure 9. A conventional TV tuner is used to extract the selected RF channel and translate it to the standard 45 MHz IF. This analog signal is then bandpass sampled and the resulting 8-bit samples are applied to the demodulator chip. The chip first measures the power of the input and feeds back a control signal to the amplifier which preceeds the A/D. This loop constitutes an *automatic gain control (AGC)*. The signal is then quadrature downconverted to produce a complex-valued sample stream. The image rejection filtering is performed asynchronously to the input clock in such a way that the filter output rate is synchronous to the QAM symbol rate. This "asynchronous resampling" is controlled by a circuit which extracts a tone at the symbol rate and feeds information back to the filter. The resulting rate-synchronous sample stream is applied to a fractionally spaced adaptive equalizer. Its output, decimated to exactly one complex sample per symbol, is applied to the digital carrier tracking loop, which removes residual carrier frequency and phase, produces "soft decisions", and quantizes the soft decisions to produce 8-bit symbol outputs. The initial prototype chip consumes about 3 watts and executes the equivalent of 700 million multiplications per second.

In order to let the viewer select any TV channel at will, the demodulator must be able to acquire all of its tracking parameters, including the equalizer, without aid from the transmitter. To do this the adaptive equalizer uses the Constant Modulus Algorithm [7, 8] to initially "open the eye" and then automatically switches over to decision direction once carrier phase acquisition is complete. Decision feedback is not employed owing to the pipelined nature of the chip's VLSI design.

6 Additional Uses of the QAM Demodulator Chip

While developed originally for use in digital cable settop converters, the demodulator chip will

be useful for at least three other applications as well. The first and second are for demodulation of digital TV signals which are broadcast over radio frequency (RF) channels instead of being sent through a coaxial or fiber cable medium. High definition television is to be transmitted in the US over the same 6 MHz-wide VHF and UHF channels over which analog television is now sent. Although vestigal sideband (VSB) transmission is currently planned, the ubiquity of QAM will probably win out. Once it does, the QAM demodulator chip can be used directly. The other broadcast medium is Multipoint Microwave Distribution Systems (MMDS), also called "wireless cable", in which analog and digital television signals of the same structure as used for cable transmission are sent instead over a broadcast signal in the 2.5 GHz microwave band. Both of these scenarios have substantially different propagation characteristics than cable-transmitted signals usually do, implying that the adaptive equalization used in the demodulator must be robust and that the equalizer's length must match the delay spreads of 2 to 3 microseconds often seen in the broadcast environment.

The third application of the demodulator is in test equipment used for maintaining the cable system itself. By using the demodulator chip as a part of the block diagram shown in Figure 5 it is possible to build a handheld piece of equipment capable of noninvasively testing cable signals and characterizing any problems encountered. Such a piece of test equipment is shown in Figure 10. It can tune to any RF channel, measure the signal quality, and test for the presence of macroreflections and ingress. As an example, consider Figure 11, which shows not only the signal constellation and spectrum, but also the channel model and ingress spectrum for an actual cable TV signal. The plots indicate that the quality degradation encountered stems not from a macroreflection but in fact from ingress from a local FM radio station.

7 Conclusion

This paper has described a recently developed blind QAM demodulator chip designed to be part of a settop converter for digital cable television. The design considerations effecting the blind equalizer component, such as length and update algorithm, have been stressed. Field operating data and some data-based channel models for QAM transmission across cable are being made available (as described in [9]) as a stimulant to further research on blind equalization useful in this application.

References

[1] L.-F. Wei, "Trellis-coded Modulation with Multidimensional Constellations," *IEEE Trans. on Info. Thy.*, vol. IT-33, pp. 483-501, July 1987.

[2] K. Feher, editor, Special Issue on Digital Radio, *IEEE Trans. on Comm.*, vol. COM-27, No. 12, Dec. 1979.

[3] K. Laudel, "Performance Results of an Low-cost Alternative Equalizer Architecture for 64/256-QAM Demodulation in a CATV Receiver", *NCTA Tech. Papers*, May 1994 and May 1995.

[4] J. A. C. Bingham, **The Theory and Practice of Modem Design**, Wiley Interscience, 1988.

[5] "Efficient QAM Equalizer/Demodulator with Non-integer Sampling," U.S. Patent granted 1995 (number not yet assigned).

[6] R.P. Gooch and J.C. Harp, "Blind Channel Identification Using the Constant Modulus Adaptive Algorithm", *Proc. Int. Conf. on Comm.*, June 1988.

[7] J.R. Treichler and B.G. Agee, "A New Approach to Multipath Correction of Constant Modulus Signals", *IEEE Trans. on Acoustics, Speech, and Sig. Proc.*, vol. ASSP-31, pp. 459-472, April 1983.

[8] D.N. Godard, "Self-recovering Equalization and Carrier Tracking in Two-dimensional Data Communications Systems", *IEEE Trans. on Communications*, vol. COM-28, pp. 1867-1875, Nov. 1980.

[9] J.R. Treichler, I. Fijalkow, and C.R. Johnson, Jr., "Fractionally Spaced Equalizers: How Long Should They Really Be?, *IEEE Sig. Proc. Mag.*, vol. 13, May 1996.

[10] V. Wolff, R. Gooch, and J. R. Treichler, "Specification and Development of an Equalizer/Demodulator for Wideband Digital Microwave Radio Signals," *WESCON Conf. Record*, paper 6.3, 1987.

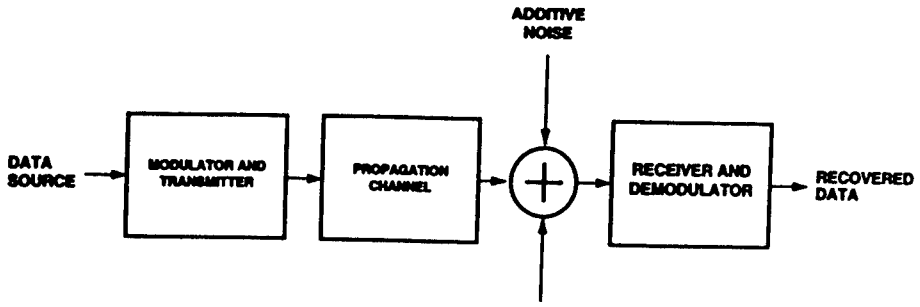


Figure 1. A Simple Model of a Digital Communications System

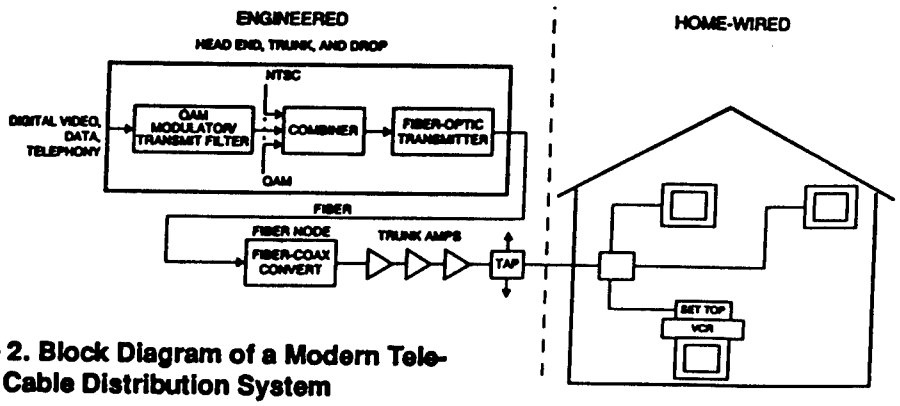


Figure 2. Block Diagram of a Modern Television Cable Distribution System

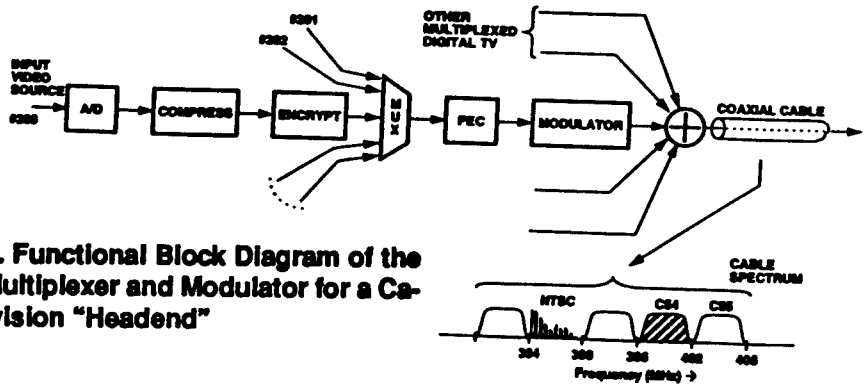


Figure 3. Functional Block Diagram of the Digital Multiplexer and Modulator for a Cable Television "Headend"

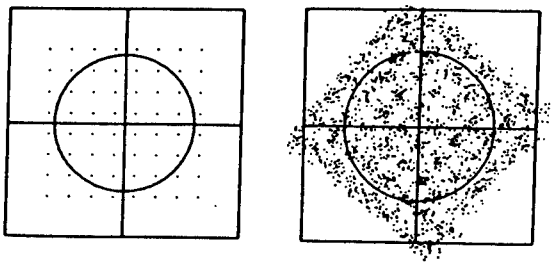


Figure 4. The Impact of a Simple Linear Channel on a 64-QAM Signal

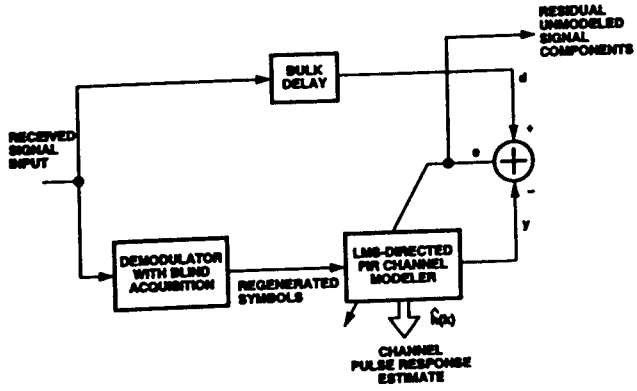


Figure 5. A Technique for Developing FIR Models for Propagation Channels Carrying PSK or QAM Signals

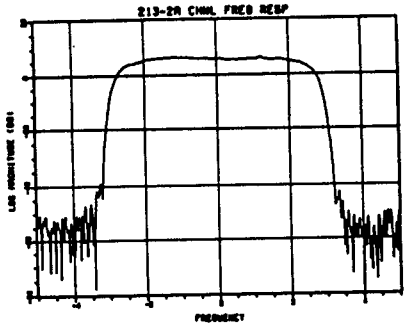


Figure 6. An Modeling Example: (a) Received power spectrum of a 64-QAM, 5.1 MBaud Cable Modem Signal, (b) The power transfer function of the estimated propagation channel

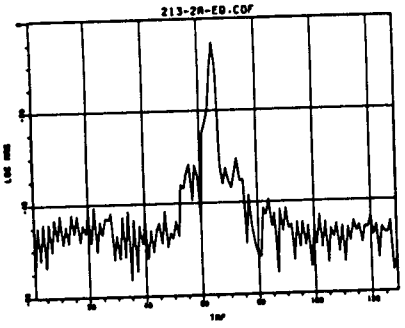


Figure 7. The Channel Estimate Impulse Response Magnitude for the 64-QAM Signal

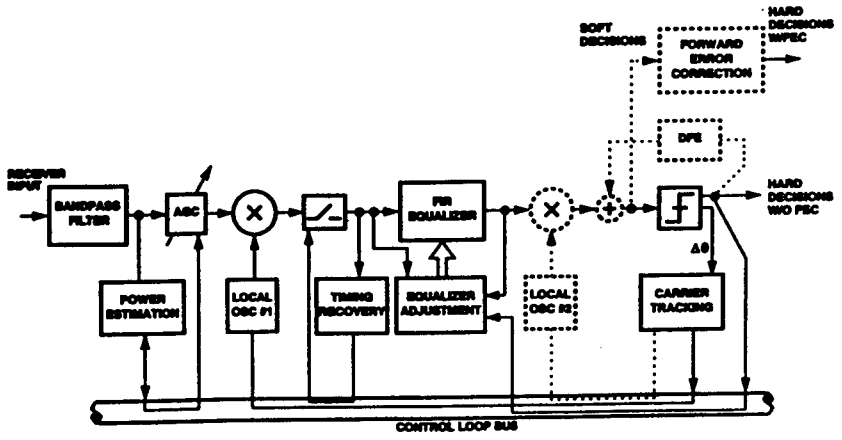


Figure 8. The Block Diagram of a Generic Equalized Data Demodulator

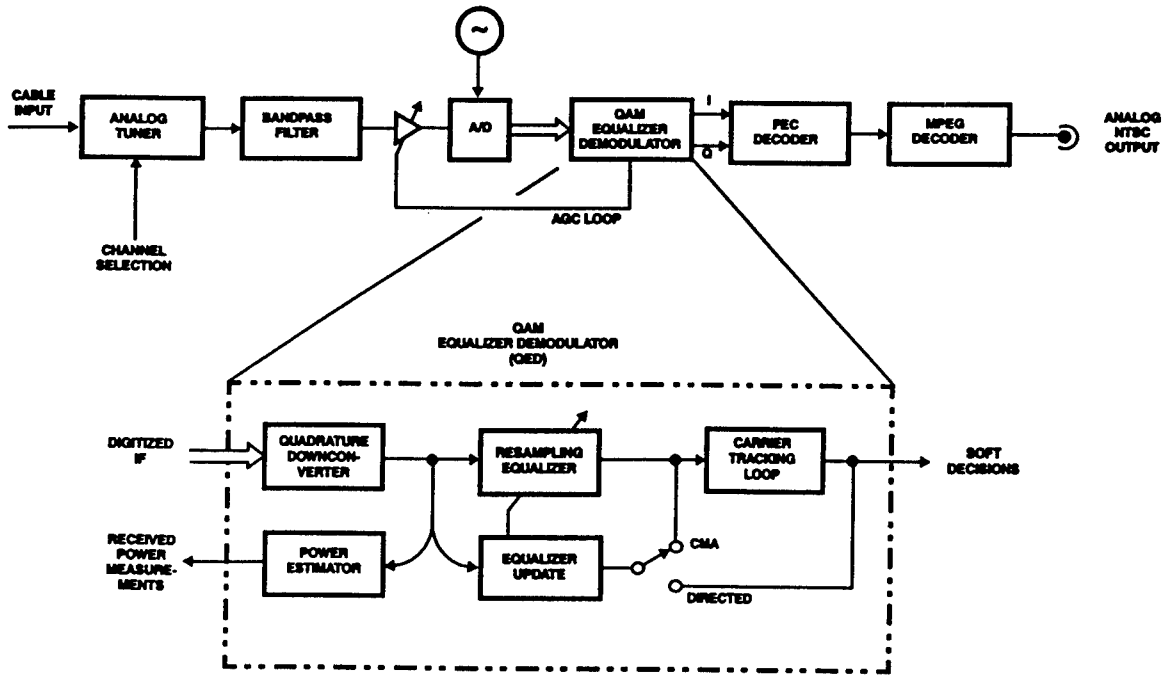


Figure 9. Block Diagram of a Generic Digital Settop Convertor Using the "QED" Chip

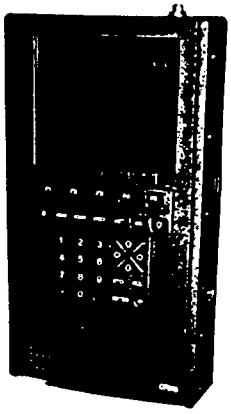


Figure 10. A Piece of Digital Television Test Equipment Which Uses Both the VES4613 "QED" Chip and Gooch-Harp Algorithm for Propagation Channel Estimation

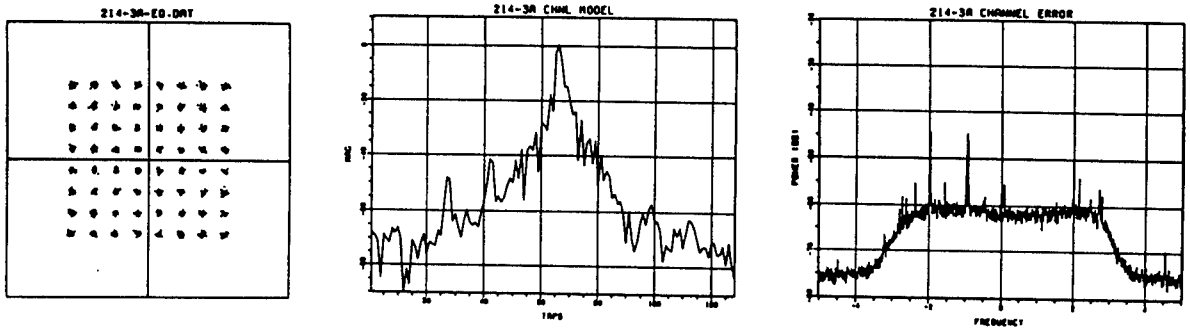


Figure 11. Using the Gooch-Harp Technique to Reveal Ingress in an Actual Digital Cable TV System

MP-1: Signal Processing for Communications I

AN OUTER-PRODUCT DECOMPOSITION ALGORITHM FOR MULTICHANNEL BLIND IDENTIFICATION

Zhi Ding
Department of Electrical Engineering
Auburn University
Auburn, AL 36839-5201
Email: ding@eng.auburn.edu

Abstract

Blind channel identification and equalization have attracted a great deal of attention recently due to their potential application in mobile communications and digital HDTV systems. In this paper, we present a new algorithm based on channel parameter outer-product decomposition. This new algorithm can be viewed as a generalization of a recently proposed linear prediction algorithm. It produces more accurate channel estimates and is more robust to over-modeling errors in channel order estimate.

1 Introduction

In popular data communication systems such as the digital mobile systems and digital HDTV systems, data signals are often transmitted through unknown channels which may introduce severe linear distortion. In order to improve the system performance, it is important for the receiver to remove channel distortions through equalization or sequence estimation. Because the available channel input training sequence may be too short or even non-existent for channel identification, blind channel identification can play useful roles in these systems.

Blind channel identification relies solely on the received channel output signal and some *a priori* statistical knowledge of the original channel input signal.

A linear prediction based approach was first presented by by Slock [5] and was later generalized and refined by Meriam *et al* [6]. Unlike many of the subspace methods that tend to be very unreliable when the channel order is over-estimated, the linear prediction approach is found to be rather robust. However, as will become clear in this paper, the linear prediction algorithm (LPA) does not fully exploit all the available second order statistical information of the channel output.

In order to derive a more robust algorithm for channels with weak precursor impulse responses, the focus of this paper is to derive the estimate based on the full outer-product decomposition of the channel parameter vector. Our results will show that based on a complete outer-product decomposition, channel identification can be significantly improved.

2 Problem Formulation

A multi-user QAM data communication system can be captured by a baseband representation. If the N user channels are all linear and causal with impulse response $\{h_u(t),\ u = 1, 2, \ldots N\}$, the received output signal can be written as

$$x(t) = \sum_{u=1}^{N} \sum_{k=-\infty}^{\infty} s_{k,u} h_u(t - kT - t_u) + w(t), \quad s_{k,u} \in \mathcal{A}_u, \tag{2.1}$$

where T is the symbol baud period and $\mathcal{A}_u$ is the input signal set of user u. The noise $w(t)$ is stationary, white, and independent of channel input sequences $s_{k,u}$, but not necessarily Gaussian. Note that $h_u(t)$ is the "composite" channel impulse response that includes transmitter and receiver filters as well as the physical channel response.

It is known that channel identification based on second order statistics is possible only for oversampled channel output. Let the sampling interval be $\Delta = T/p$ where p is an integer. The oversampled discrete signals and responses are

$$x_i \triangleq x(i\Delta), \quad h_u[i] \triangleq h_u(i\Delta) \quad \text{and} \quad w_i \triangleq w(i\Delta). \tag{2.2}$$

Suppose $\{h_u(t)\}$ has finite support $[0, T_h)$, which spans $m_0 + 1$ integer periods. By defining the following notations

$$\begin{aligned}
\mathbf{x}[k] &\triangleq [x_{kp}\ \ x_{kp+1}\ \ \cdots\ \ x_{kp+p-1} \\
&\qquad x_{(k-1)p}\ \ x_{(k-1)p+1}\ \ \cdots\ \ x_{kp-Mp+1}]' \\
\mathbf{s}_k &\triangleq [s_{k,1}\ \ s_{k,2}\ \ \cdots\ \ s_{k,N}\ ]; \\
\mathbf{s}[k] &\triangleq [s_k\ \ s_{k-1}\ \ \cdots\ \ s_{k-m_0-M+1}]' \\
\mathbf{w}[k] &\triangleq [w_{kp}\ \ w_{kp+1}\ \ \cdots\ \ w_{kp-Mp+1}]' \\
\mathbf{h}_u[i] &\triangleq [\ h_u[ip]\ \ h_u[ip+1]\ \ \cdots\ \ h_u[ip+p-1]\ ]', \\
\mathbf{H}_i &\triangleq [\mathbf{h}_1[i]\ \mathbf{h}_2[i]\ \ldots\ \mathbf{h}_N[i]],
\end{aligned}$$

0-8186-7576-4/96 $5.00

we can form a $Mp \times (m_0 + M)$ block Toeplitz matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_{m_0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_{m_0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_{m_0} \end{bmatrix}. \tag{2.3}$$

Clearly, m_0 is the order of the N dynamic FIR channels. With these notations, the sampled channel output signal vector can be written as

$$\mathbf{x}[k] = \mathbf{H}\mathbf{s}[k] + \mathbf{w}[k]. \tag{2.4}$$

Consequently, the channel output covariance matrix becomes

$$R_{m_0} = E\{\mathbf{x}[k]\mathbf{x}[k]^H\} = \sigma_s^2 \mathbf{H}\mathbf{H}^H + \sigma_w^2 I \tag{2.5}$$

assuming that the channel input signal is white with zero mean and $R_s = E\{\mathbf{s}[k]\mathbf{s}[k]^H\} = \sigma_s^2 I$ while the noise is spatially white with zero mean with $R_w = E\{\mathbf{w}[k]\mathbf{w}[k]^H\} = \sigma_w^2 I$.

Our objective is to identify the channel $\mathbf{H}$ from the second order statistics of the channel output signal $\mathbf{x}[k]$ given in R_{m_0} under the identifiability condition [1] that both $\mathbf{H}$ and R_s are full-rank. The use of second order statistics for blind channel identification was first exploited by Tong, Xu, and Kailath [1]. The basic concept hinges on the signal and noise subspace separation through singular value decomposition (SVD) of the auto-covariance matrix R_{m_0}.

The sub-channel matching (SCM) method presented in [3] and the subspace method of [2] can both be posed as a minimum eigen-vector problem under proper channel length constraints. The special block Toeplitz structure is utilized in both algorithms. When the channel length is over-estimated, common zeros must be factorized out from the sub-channel estimates. As a result, both algorithms are very sensitive to channel length mis-matching.

In [5] and [6], a linear prediction algorithm (LPA) was presented for channel estimation. It is shown to be more robust to over-estimated channel length. However, as we will show later in this paper, the LPA only uses part of the overall information because the channel estimate is based on the first p columns of the estimated channel parameter vector outer-product matrix. As a more robust and accurate channel estimation algorithm, the outer-product decomposition algorithm we propose will exploit second order statistics more effectively.

3 Algorithm Development

We will form an outer-product of the channel parameter matrix

$$\mathbf{h} \triangleq [\, \mathbf{H}_0' \;\; \mathbf{H}_1' \;\; \dots \;\; \mathbf{H}_{m_0}' \,]'. \tag{3.1}$$

Let

$$X[k] \triangleq [x_{kp} \;\; x_{kp+1} \;\; \dots \;\; x_{kp+p-1}]' = \sum_{i=0}^{m_0} \mathbf{H}_i s_{k-i}'. \tag{3.2}$$

For notational convenience, define

$$R(n) \triangleq E\{X[k]X[k-n]^H\} = \sigma_s^2 \sum_{i=n}^{m_0} \mathbf{H}_i \mathbf{H}_{i-n}^H. \tag{3.3}$$

The channel output covariance covariance matrix can be written as

$$R_{m_0} \triangleq E\{\mathbf{x}[k]\mathbf{x}^H[k]\} = \sigma_s^2 \mathbf{H}\mathbf{H}^H + \sigma_w^2 I. \tag{3.4}$$

Denote

$$H_a \triangleq \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_{m_0} & \mathbf{0} \dots & \mathbf{0} \\ \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{0} & \mathbf{0} \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{H}_{m_0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \dots & \mathbf{0} \end{bmatrix}. \tag{3.5}$$

If we define $p \times p$ block matrices as

$$D_{i,j} \triangleq \sum_{k=i-1}^{m_0} \mathbf{H}_k \mathbf{H}_{k+j-i}^H, \quad 1 \le i, j \le m_0 + 1, \tag{3.6}$$

it can be verified that

$$H_a H_a^H = \begin{bmatrix} D_{1,1} & D_{1,2} & \dots & D_{1,m_0+1} \\ D_{2,1} & D_{2,2} & \dots & D_{2,m_0+1} \\ \vdots & \vdots & \ddots & \vdots \\ D_{m_0+1,1} & D_{m_0+1,2} & \dots & D_{m_0+1,m_0+1} \end{bmatrix} \tag{3.7}$$

This matrix is an $(m_0 + 1)p \times (m_0 + 1)p$ Hermitian matrix. Now define a new matrix as

$$D_{m_0+1} = \begin{bmatrix} D_{2,2} & \dots & D_{2,m_0+1} & \mathbf{0} \\ D_{3,2} & \dots & D_{3,m_0+1} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ D_{m_0+1,2} & \dots & D_{m_0+1,m_0+1} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{3.8}$$

We can form another Hermitian matrix from

$$\Delta_{m_0+1} \triangleq H_a H_a^H - D_{m_0+1} = \mathbf{h}\mathbf{h}^H. \tag{3.9}$$

Clearly, matrix Δ_{m_0+1} forms the outer-product of the channel parameter matrix $\mathbf{h}$. Its singular value decomposition will generate $\mathbf{h}Q$, in which Q is an $N \times N$ unitary matrix. This ambiguity is intrinsic to the multi-user blind identification problem and cannot be resolved unless additional information is available. If a multi-channel equalizer is built according to the estimate $\mathbf{h}Q$, the N receiver outputs will be memoryless combinations of the N channel inputs and will need to be separated.

In order to estimate the Δ_{m_0+1} matrix, first construct

$$R_c = \begin{bmatrix} R(0)-\sigma_w^2 I & R(1) & \cdots & R(m_0) \\ R(1) & R(2) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ R(m_0) & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} = \sigma_s^2 H_a \mathbf{H}^H. \tag{3.10}$$

In addition, it can also be easily shown that

$$R_{m_0} - \sigma_w^2 I = \sigma_s^2 \mathbf{H}\mathbf{H}^H. \tag{3.11}$$

In order to estimate the product $H_a H_a^H$, it is important to note that when $\mathbf{H}$ has full column rank, $\mathbf{H}\mathbf{H}^H$ is invertible and $\mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{H} = I$. Then

$$R_c(R_{m_0} - \sigma_w^2 I)^{-1} R_c^H = \sigma_s^2 H_a H_a^H \tag{3.12}$$

where σ_s^2 is known.

If there is only a single user, the channel impulse response vector can be estimated from the rank one outer-product matrix, through eigen-decomposition, QR decomposition, or simply post-multiplying a random column vector. We thus name the method "outer-product decomposition algorithm" (OPDA).

Notice that OPDA requires two singular value (or eigenvalue) decompositions in its implementation. Its complexity is therefore similar to the linear prediction algorithm (LPA) presented by Meriam *et al.* [6], the TXK method [1], and the sub-channel matching method [3]. However, LPA estimates the channel only from the first p columns of the outer-product matrix. If the channel impulse response has weak precursor samples, then LPA is likely to be highly inaccurate since noise and numerical error will likely dominate the first few columns of $\widehat{\Delta D}_{m_0+1}$. Therefore, OPDA is expected to provide more robust performance than LPA.

4 Simulation Results

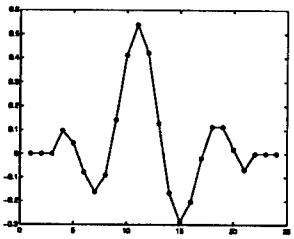
We now present simulation results to illustrate the channel identification performance of the proposed OPDA. Our experiments are based on a single user with a multi-path channel model. We consider a raised-cosine pulse $P(t)$ limited in $6T$ with roll-off factor 0.10 and a two ray multi-path channel

$$c(t) = \delta(t) - 0.7\delta(t - T/3).$$

The impulse response

$$h(t) = c(t) * h(t) = P(t) - 0.7P(t - T/3)$$

is shown in Figure 4. The data input signal is i.i.d. BPSK and the oversampling factor is $p = 3$. In all our simulations, M is chosen to be twice as long as $P(t)$.



In the first set of simulation tests, we compare the two methods OPDA and LPA based on 100 and 200 bauds of channel output samples. The channel order is unknown and is estimated using the MDL criterion. The normalized mean square error (MSE) of the channel estimate under different channel SNR levels is shown in Figure 1.

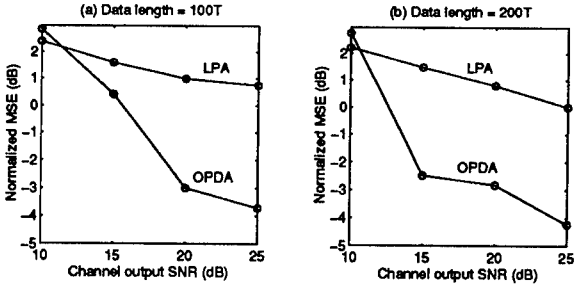


Figure 1: Normalized MSE of channel estimate given different SNR levels.

For several different data lengths, the resulting normalized MSE is shown in Figure 2. Once again, the results show that OPDA and LPA are equally ineffective when SNR is low. But when the SNR is higher, OPDA out-performs LPA significantly.

We also tested the comparative robustness of the two algorithms when channel mismatching is present. Fixing SNR=20dB, we manually varied the channel length estimate from $2T$ to $10T$. Notice that the true channel length is $6T$. The results clearly show that while LPA is less sensitive to errors in channel order estimate, its performance is generally much worse compared with that of OPDA. When the channel order estimate deviates modestly from the true channel order, OPDA generates a much smaller normalized MSE.

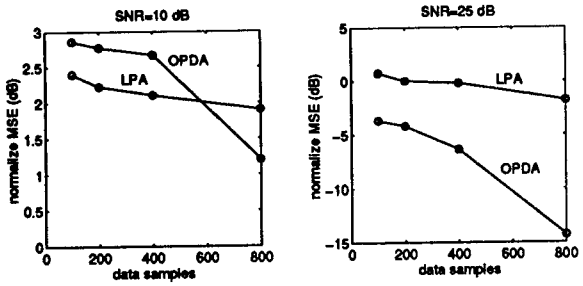


Figure 2: Normalized MSE of channel estimate given different data lengths.

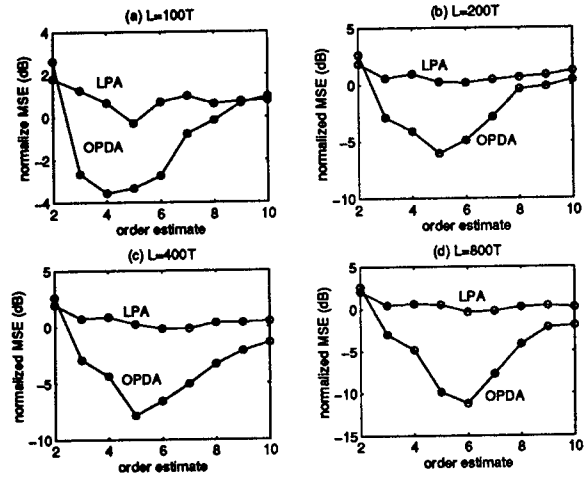


Figure 3: Normalized MSE of channel estimate given channel length mismatch.

Finally, we compare a group of typical impulse responses estimated from 50 independent trials of the OPDA and LPA under 20dB SNR and data length of $L = 400T$. Assuming the channel length is correctly estimated, The estimated impulse responses are shown in Figure 4.

5 Conclusions

We present a new robust and accurate blind channel identification algorithm OPDA based on outer-product decomposition. This new algorithm can be viewed as a generalized method of the recently proposed linear prediction algorithm (LPA). The new OPDA is capable of generating much more superior identification results.

Acknowledgement: This work is supported by the NSF Grant MIP-9210100 by the US Army Research Office.

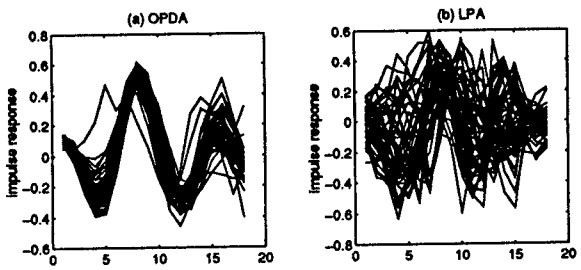


Figure 4: 50 independent channel estimates.

References

[1] L. Tong, *et al.*. "A new approach to blind identification and equalization of multi-path channels," *IEEE Trans. on Info. Theory*, IT-40:340-349, March 1994.

[2] E. Moulines, *et al.*, "Subspace Methods for the Blind Identification of Multichannel FIR Filters", *Proc. 1994 IEEE ICASSP*, pp. IV:573-576, Adelaide, 1994.

[3] H. Liu, G. Xu, and L. Tong, "A Deterministic Approach To Blind Identification of Multi-Channel FIR Systems", *Proc. 1994 IEEE ICASSP*, pp. IV:581-584, Adelaide, May 1994.

[4] Y. Li and Z. Ding, "A New Non-Parametric Cepstral Method for Blind Channel Identification from Cyclostationary Statistics", *Proc. 1993 MILCOM*, pp. 648-652, Boston, Oct. 1993.

[5] D. Slock, "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction", *Proc. 1994 IEEE ICASSP*, pp. IV:585-588, Adelaide, May 1994.

[6] K. A. Meriam, P. Duhamel, D. Gesbert, L. Loubaton, S. Mayrargue, E. Moulines and D. Slock, "Prediction Error Methods for Time-Domain Blind Identification of Multichannel FIR Filters," *Proc. 1995 IEEE ICASSP*, pp. 1968-1971, Detroit, MI, May 1995.

[7] M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-33:387-392, April 1985.

Decoupled Blind Symbol Estimation Using an Antenna Array

A. Ranheim*

P. Pelin†

Department of Applied Electronics, Chalmers University of Technology
S-412 96 Göteborg, Sweden

Abstract

The problem of separating superimposed digitally modulated signals using an array of antennas is considered. The proposed method exploits the finite alphabet structure to demodulate one signal at the time, resulting in a computationally efficient solution. The resulting signal estimates are shown to be exact in the noise-free case. In noisy scenarios, the performance is comparable with that of the recently proposed iterative least squares approach, which demodulates all signals simultaneously at a higher computational cost.

1 Introduction

Array processing techniques can be used to discriminate between spatially separated co-channel signals, and can consequently increase the capacity in wireless communication systems. This paper discusses how to reliably demodulate one or more desired signals of interest (SOI) from the output of an array, in the presence of other co-channel signals and noise. Traditional approaches exploit the spatial structure of the array, and as such depend on high-resolution estimates of the DOA's (Direction Of Arrival) of the incoming signals. Since modern wireless communication systems are characterized by a highly variable propagation environment, this spatial structure is poorly defined [3]. On the other hand, these methods make no assumptions about the signals themselves, and are thus not exploiting the structural information present in the signals. Various *blind copy algorithms* have been proposed to alleviate this problem [1],[10]. The referenced techniques require synchronized signals and must assume flat frequency fading. Generalizations are considered in [2, 6, 9] and [10]. This paper proposes a new approach, based on decoupling the estimation problem (i.e.treating one signal at a time). This leads to an algorithm with similar or better performance for a typical scenario, and furthermore reduces the computational cost involved in the estimation procedure significantly. These claims are supported by simulation results and a complexity count. Consistency and uniqueness issues are also addressed.

*e-mail:*ranheim@ae.chalmers.se*, phone: +46-31-772 1813, fax: +46-31-772 1782

†e-mail:*pelin@ae.chalmers.se*

2 Signal Model

With d syncronized signals arriving at an m element antenna array, the complex output vector after matched filtering and symbol-rate sampling, can be expressed by the following familiar equation

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{v}(n) \tag{1}$$

where $\mathbf{A}$ is the collection of total array response vectors (spatial signatures), scaled by the signal amplitudes

$$\mathbf{A} = [p_1\mathbf{a}_1 \dots p_d\mathbf{a}_d]. \tag{2}$$

$\mathbf{s}(n) = [b_1(n) \dots b_d(n)]^T$, $b_i(n) = \pm 1$ (BPSK), and $\mathbf{v}(n)$ is spatially and temporally white noise. For simplicity we consider BPSK signals, but extensions to arbitrary linear modulation formats is straightforward. A block formulation is obtained by taking N snapshots, yielding

$$\mathbf{X}(N) = \mathbf{A}\mathbf{S}(N) + \mathbf{V}(N) \tag{3}$$

whre $\mathbf{X}(N) = [\mathbf{x}(1) \dots \mathbf{x}(N)]$, $\mathbf{S}(N) = [\mathbf{s}(1) \dots \mathbf{s}(N)]$, and $\mathbf{V}(N) = [\mathbf{v}(1) \dots \mathbf{v}(N)]$. The spatial structure of the data is represented by $\mathbf{A}$, while the matrix $\mathbf{S}$ represents the temporal strucure.

By defining one signal (at a time) to be the signal of interest (SOI), (3) can be rewritten in

0-8186-7576-4/96 $5.00 © 1996 IEEE

the following way

$$\begin{aligned}\mathbf{X}(N) &= \mathbf{a}_1\mathbf{s}_1 + \sum_{l=2}^{d}\mathbf{a}_l\mathbf{s}_l + \mathbf{V}(N) \\ &= \mathbf{a}_1\mathbf{s}_1 + \mathbf{J}(N) \end{aligned} \quad (4)$$

where the first signal is taken to be the SOI, without loss of generality. The term $\mathbf{J}(N)$ thus corresponds to interfering signals plus noise.

3 Decoupled Symbol Estimation: Algorithm

Since it is desired to estimate the signals with little or no spatial knowledge, the idea is to iteratively estimate $\mathbf{a}$ and $\mathbf{s}$, based on the formulation in (4).

3.1 Algorithm

Given an initial estimate of $\mathbf{A} = \hat{\mathbf{A}} = [\hat{\mathbf{a}}_1 \ldots \hat{\mathbf{a}}_d]$, the following weighted least-squares criterion function is minimized

$$\min_{\mathbf{a},\mathbf{s}}(\mathbf{X}-\mathbf{as})^*\mathbf{W}(\mathbf{X}-\mathbf{as}) = \min_{\mathbf{a},\mathbf{s}}\|\mathbf{W}^{-\frac{1}{2}}(\mathbf{X}-\mathbf{as})\|^2 \quad (5)$$

Here, $\mathbf{W}$ should ideally be chosen as $\mathbf{R}_J^{-1}$ [5], which can be interpreted as a prewhitening of the data vector $\mathbf{x}(n)$. However, it can easily be shown using the matrix inversion lemma, that using the inverse of the sample estimate of the covariance of the array output produces asymptotically (for large N) equivalent signal estimates. Equation (5) can thus be reformulated as follows

$$\min_{\mathbf{b},\mathbf{s}}\|\mathbf{Z}-\mathbf{bs}\|^2 \quad (6)$$

with $\mathbf{Z} = \hat{\mathbf{R}}_x^{-\frac{1}{2}}\mathbf{X}$, $\mathbf{b} = \hat{\mathbf{R}}_x^{-\frac{1}{2}}\mathbf{a}$, and $\hat{\mathbf{R}}_x = \frac{1}{N}\mathbf{X}\mathbf{X}^*$ The solution to (6) w.r.t. $\mathbf{s}$ is

$$\hat{\mathbf{s}} = (\hat{\mathbf{b}}^*\hat{\mathbf{b}})^{-1}\hat{\mathbf{b}}^*\mathbf{Z} = \frac{1}{\|\hat{\mathbf{b}}\|^2}\hat{\mathbf{b}}^*\mathbf{Z} \quad (7)$$

Exploiting the finite-alphabet property, this solution is projected onto its closest discrete values in the signal space (± 1). The (modified) steering vector $\hat{\mathbf{b}}$ is then updated by minimizing (6) w.r.t. $\mathbf{b}$. The solution is

$$\hat{\mathbf{b}} = \mathbf{Z}\hat{\mathbf{s}}^*(\hat{\mathbf{s}}\hat{\mathbf{s}}^*)^{-1} = \frac{\mathbf{Z}\hat{\mathbf{s}}^*}{N} \quad (8)$$

Note that (8) is simply a *temporally* matched filter to the current signal estimate, whereas (7) represents a *spatially* mathed filter. The process is repeated until $\mathbf{s}$ converges, after which the algorithm continues with the next signal.

3.2 Consistency and Uniqueness

A relevant question is whether or not the algorithm is able to "capture" the transmitted signals. Since the iterative scheme corresponds to a relaxed optimization procedure, it is a simple matter to show that it is guaranteed to converge to a local minimum. Whether or not this corresponds to a "true" minimum depends in general on the initial estimate. However, even if it does, it is a non-trivial question if the global minimum yields a consistent estimate of the transmitted waveform. Clearly, this is possible only for high enough signal-to-noise ratio (SNR), so we will analyze the quality of the global minimum assuming that the noise variance tends to zero.

Substituting the solution for $\mathbf{b}$ in (8) into (6), gives the following minimization problem

$$\min_{\mathbf{s}}\|\mathbf{Z}-\frac{1}{N}\mathbf{Z}\mathbf{s}^*\mathbf{s}\|^2 = \min_{\mathbf{s}}\|\mathbf{Z}-\mathbf{Z}\frac{\mathbf{s}^*\mathbf{s}}{\mathbf{s}\mathbf{s}^*}\|^2 \quad (9)$$

Reformulating in terms of projection matrices

$$\min_{\mathbf{s}}\|\mathbf{Z}(\mathbf{I}-\mathbf{P}_{\mathbf{s}^\bullet})\|^2 \Leftrightarrow \max_{\mathbf{s}}\|\mathbf{Z}\mathbf{P}_{\mathbf{s}^\bullet}\|^2 \Leftrightarrow \max_{\mathbf{s}}\|\mathbf{Z}\mathbf{s}^*\|^2 \quad (10)$$

where the last equality follows since for BPSK signals $P_{s^\bullet} = \mathbf{s}^*(\mathbf{s}\mathbf{s}^*)^{-1}\mathbf{s} = \mathbf{s}^*\mathbf{s}/N$. Furthermore, by using $\mathbf{Z} = \hat{\mathbf{R}}^{-\frac{1}{2}}\mathbf{X}$, the following can easily be derived

$$\max_{\mathbf{s}}\|\mathbf{Z}\mathbf{s}^*\|^2 = \max_{\mathbf{s}}\{\mathbf{s}\mathbf{P}_{\mathbf{X}^\bullet}\mathbf{s}^*\} \quad (11)$$

Using Schwartz inequality,

$$\mathbf{s}\mathbf{P}_{\mathbf{X}^\bullet}\mathbf{s}^* \le \|\mathbf{s}\|^2\|\mathbf{P}_{\mathbf{X}^\bullet}\|_2^2 = N \quad (12)$$

with equality when $\mathcal{R}(\mathbf{s}^*) \subseteq \mathcal{R}(\mathbf{P}_{\mathbf{X}^\bullet})$, i.e. $\mathbf{s}_1^* = \mathbf{X}^*\mathbf{t}'$ for some column vector $\mathbf{t}'$. In the noise-free case we have $\mathbf{X}_{m|N} = \mathbf{A}\mathbf{S}_0$, giving

$$\mathbf{s}_1^* = (\mathbf{A}\mathbf{S}_0)^*\mathbf{t}' = \mathbf{S}_0^*\mathbf{t}\ . \quad (13)$$

Thus, the signal estimate converges to a linear combination of the d transmitted signals. Under suitable "persistence of excitation" conditions, $\mathbf{t}$ must contain a ± 1 in one position and zeros otherwise, implying that $\mathbf{s}_1$ is one of the true transmitted signals with a possible sign change. Specifically, since all signal vectors $\mathbf{s}_1 \ldots \mathbf{s}_d$ are treated likewise, we can write for the d signal estimates

$$[\mathbf{s}_1^* \ldots \mathbf{s}_d^*] = [\mathbf{s}_{01}^* \ldots \mathbf{s}_{0d}^*]\mathbf{T}\ . \quad (14)$$

In [10] it is shown that the above can hold only if $\mathbf{T}$ is a diagonal matrix with ± 1 entries, or a permutation matrix, or a product of the two; provided that the columns of $\mathbf{S}_0$ include all the

2^d possible distinct d-vectors, with ± 1 elements. The latter is a mild condition, which is satisfied in most cases of practical interest. We conclude that the global minimizer of the decoupled criterion function converges to any of the d transmitted signals as the noise variance tends to zero.

4 Performance and Complexity

4.1 Performance

Figure 1 below shows the results of a simulation comparing the performance of the proposed algorithm with that of ILSP [8]. A total of $d = 3$ signals are impinging on a 4-element uniform linear array (ULA), and the BER vs. SNR is evaluated. A 5 bit training sequence was used to get an initial estimate of the steering vectors $\hat{\mathbf{A}}$. The results clearly show that an improved performance has been obtained in this scenario. (the BER of the signal with DOA=106^0 was ~ 0.25 for both algorithms, regardless of SNR).

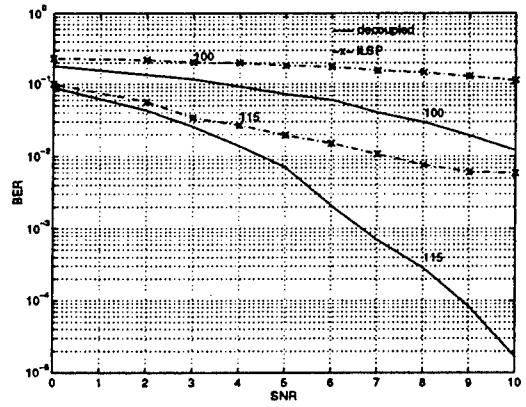


Figure 1. *Performance of decoupled WLS approach and ILSP algorithm. Simulated data.*

The algorithm was also tested on a real dataset collected at the University of Texas at Austin, and compared to ILSP. Two closely spaced signal transmitted bursts of 198 symbols/burst (still BPSK), and 4 antenna elements was used at the receiver. Different noise realizations were then generated and added to the data in order to evalute the BER vs. SNR performance (for the strongest signal only). The results are given in figure (2), and demonstrates that the two algorithms performs similarly in this scenario.

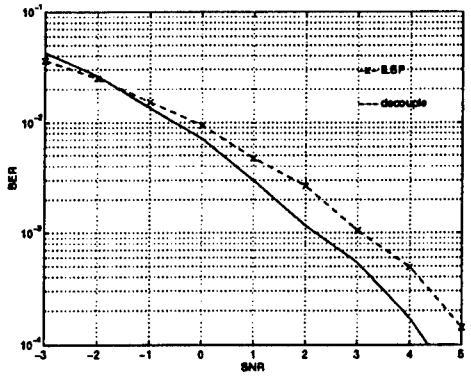


Figure 2. *Performance of decoupled WLS approach and ILSP algorithm. Real data.*

In general, one can say that the decoupled algorithm outperforms ILSP for large burst lengths and a small array. The explanation for this is that the approximation $\hat{\mathbf{R}}_x^{-1} \propto \mathbf{R}_J^{-1}$, used in section 3.1 improves with larger N and smaller m. On the other hand, in scenarios with $m \gg d$ and N relatively small; it is our experience that ILSP gives a slight improvement compared to our proposed method.

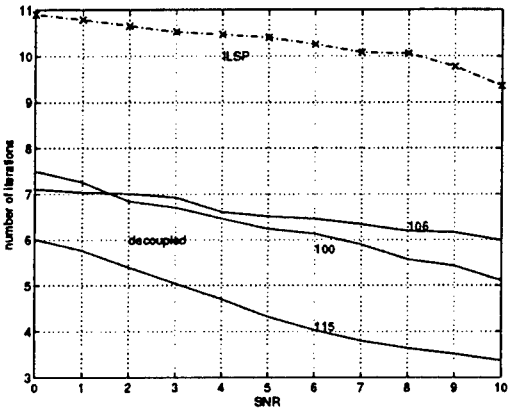


Figure 3. *Complexity of decoupled WLS approach and ILSP algorithm*

4.2 Complexity

Since the proposed algorithm is based on the same iterative approach as ILSP, it is interesting to compare the complexity of the two methods. Before the iterative estimation of $\mathbf{s}$ and $\mathbf{b}$ begin, $\hat{\mathbf{R}}_x^{-\frac{1}{2}}$ and the product $\hat{\mathbf{R}}_x^{-\frac{1}{2}}\mathbf{X}$ must be computed. This requres $\mathcal{O}(m^2) + m^2N$ flops. Note

that this computation is only carried out once for a given block of data $\mathbf{X}(m|N)$. Looking at eqn.(7), it is sufficient to compute the product of the $\hat{\mathbf{b}}^*(1|m)$-vector with the modified data-matrix $\mathbf{Z}(m|N)$, requiring $2mN$ flops [4]. Similarly, to update the $\hat{\mathbf{b}}$ estimate, the same kind of product is computed. This gives a total of $4mN$ flops per iteration and signal. A similar count for ILSP results in $2Nmd + 2d^2(N - \frac{d}{3}) + md^2$ flops to solve for $\hat{\mathbf{A}}$ and $2Nmd + 2d^2(m - \frac{d}{3}) + Nd^2$ flops to solve for $\hat{\mathbf{S}}$ (both per iteration). Consequently, the proposed algorithm results in a significant reduction in computational complexity as compared to the ILSP algorithm.

In order to get a fair comparison of complexity, one should also look at the convergence properties of the two algorithms. The number of iterations required for convergence is compared in figure 3 (same scenario as in 4.1). Even if the total number of iterations for the proposed algorithm (add the three solid lines) exceed that of ILSP in this case, it does not offset the large difference in complexity in terms of flops count.

As an illustration, the following typical numbers were obtained using a flops count in MATLAB for the scenario above at $SNR = 5dB$: ILSP requires (w. ~ 10.4 iterations on the average, see figure 3)

$$(10.4iter)(32650flops/iter) \simeq 339000flops.$$

The decoupled algorithm requires a total of

$$(6552flops/iter)(6.2+6.5+4.2iter) \simeq 110700flops$$

in addition to the inversion of $\hat{\mathbf{R}}_x$ (only once!).

5 Conclusion

The simulation results indicate that the proposed algorithm has similar or improved performance compared to ILSP, and that this improvement is accompanied by a significant reduction in computational complexity. This is particularly notable if not all signals are of interest. The method can be extended in a straightforward manner in order to include the case of non-synchronized users and time-dispersive channels, using e.g. conventional synchronization and equalization techniques [7]. Simulations performed by the authors (not included here due to space limitations), have confirmed this claim.[1]

[1] Acknowledgements. The financial support from the Swedish National Board for Industrial and Technical Development is gratefully acknowledged. The authors are also grateful to Professor G. Xu at the University of Texas at Austin, for providing the data used in figure 2.

References

[1] S. Daas A. Swindlehurst and J. Yang. "Analysis of Decision Directed Beamformer". *Submitted to IEEE Trans. on Signal Processing*, October 1993.

[2] S. Talwar A. van der Veen and A.Paulraj. "Blind Estimation of Multiple Digital Signals Transmitted over FIR Channels". *IEEE Signal Processing Letters*, Vol.2(NO.5), May 1995.

[3] A.F. Naguib G. Raleigh, S.N. Diggavi and A. Paulraj. "Characterization of Fast Fading Vector Channels for Multi-Antenna Comunication Systems". In *Proc. Asilomar Conf. Sig. Syst. Comput.*, Pacific Grove, CA, 1994.

[4] G.H.Golub and C.F. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, MD, 1984.

[5] Steven M. Kay. *Fundamentals of Statistical Signal Processing-Estimation Theory*. Prentice-Hall, Englewood Cliffs, New Jersey, 1993.

[6] H. Liu and G. Xu. "A Deterministic Approach to Blind Symbol Estimation". In *Proc. 28^{th} Asilomar Conference on Signals, Systems and Computers*, 1994.

[7] John G. Proakis. *Digital Communications*. McGraw-Hill, third edition edition, 1995.

[8] M. Viberg S. Talwar and A. Paulraj. "Estimating Multiple Co-Channel Digital Signals Using an Antenna Array". *IEEE Trans. on AP*, Vol.41:1687–1694, Dec 1993.

[9] D.T.M. Slock. "Blind Joint Equalization of Multiple Syncronous Mobile Users Using Oversampling and/or Multiple Antennas". In *Proc. 28^{th} Asilomar Conference on Signals, Systems and Computers*, 1994.

[10] Shilpa Talwar. *"Blind Space-Time Algorithms for Wireless Communication System"*. PhD thesis, Stanford University, 1996.

Impulsive Noise Modeling with Stable Distributions in Fading Environments *

Jacek Ilow
Surface Radar Section
Defence Research Establishment Ottawa
Ottawa, Ont K1A 0Z4, Canada
jilow@rdrsrv.rdr.dreo.dnd.ca

Dimitris Hatzinakos
Dept. of Electrical and Computer Eng.
University of Toronto,
Toronto, Ont. M5S-1A4, Canada
dimitris@comm.toronto.edu

Abstract

The objective of this paper is to introduce a statistical and physically based mechanism giving rise to α-stable noise models. We show that the additive interference which is present in many environments can be modeled as symmetric α-stable by assuming: (i) independent signaling (effects) from a large number of interferers of the same type (modulation); (ii) Poisson distribution of interferers in space; and (iii) inverse power attenuation of signal strength with distance. Our approach to α-stable noise modeling is based on the LePage series representation [5] as opposed to the influence function approaches taken in [1],[8]. The formulas derived are used to predict noise statistics in environments with lognormal shadowing and Rayleigh fading. The LePage series framework allows us to investigate practical constraints in the system model adopted, such as the finite number of interferers and nonhomogeneous Poisson fields of interferers.

1 Introduction

The characterization of the corrupting noise distribution is an important requirement for most system design problems because it leads to the development of noise suppression methods. The most widely used noise model is the Gaussian random process. However, in some natural environments, the Gaussian noise model may not be appropriate. This is evident from a higher probability of large amplitude values than is consistent with Gaussian distributions. A number of models have been proposed for such impulsive phenomena, either by fitting experimental data or based on physical grounds. Recently, it has been suggested that among all the heavy-tailed distributions, the family of stable distributions provides the most accurate model for impulsive noise [1],[8]. In communications, stable noise models have been verified experimentally in various underwater communications and radar applications [1], [9]. Stable distributions share defining characteristics with Gaussian distributions, such as the stability property and the generalized central limit theorem (GCLT), and, in fact, include Gaussian distributions as a limiting case. Because stable distributions, except for the Gaussian case, have infinite variance, at first sight, it appears that stable noise models do not have the wide applicability enjoyed by second-order processes. However, in this paper we present a realistic physical mechanism giving rise to stable noise. We do this by considering the nature of noise sources, their distributions in time and space, and propagation conditions.

2 System Model

In our system model, a detector is located at the center of a plane where there is a large number ($N \to \infty$) of transmitters using the same power and modulation. The distances between the detector and interfering terminals are denoted as r_i, where $r_1 \leq r_2 \leq, \cdots, \leq r_N$. In general, after the correlation detection of passband interference, the interfering signal is represented as an n-dimensional vector given by

$$\mathbf{Y} = \lim_{N \to \infty} \sum_{i=1}^{N} a(r_i)\mathbf{X}_i, \qquad (1)$$

where $a(r)$ represents the signal attenuation over distance r, and $\mathbf{X}_i = [X_{i,1}, \ldots, X_{i,n}]$ is a random vector with n coordinates $X_{i,j}, j = 1, \ldots, n$ which are real

*This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

random variables (RVs). The jth coordinate of $\mathbf{X}_i$ is the correlation of $x_i(t)$ with the function $\varphi_j(t)$ [1]. In this paper, we are concerned with characterizing the distribution of $\mathbf{Y}$, and in order to do this, we assume that $\mathbf{X}_i$ are spherically symmetric[2] (SS) RVs. Because all interfering terminals use the same modulation scheme and transmit at the same power, it is reasonable to assume that the random vectors $\mathbf{X}_i$ are i.i.d. Moreover, the distribution of $\mathbf{X}_i$ is independent of r_i. To explain the noise modeling in (1), it is useful to consider a system with on-off frequency shift keying (FSK) and noncoherent detection. In this system, $\varphi_1(t) = \cos(2\pi f_0 t)$ and $\varphi_2(t) = \sin(2\pi f_0 t)$, and the projection of $x_i(t)$ onto $\{\varphi_1(t), \varphi_2(t)\}$ results in $\mathbf{X}_i = [\cos(\Theta_i), \sin(\Theta_i)]$, where Θ_i is uniformly distributed in $(0, 2\pi]$. This means that $\mathbf{X}_i$ is circularly symmetric (CS), a bivariate case of a SS vector. With respect to the terminal positions, we assume that terminals form a Poisson point process with the expected number of terminals per unit area/volume given by λ [6].

3 Stable Interference Models in Environments with Deterministic Propagation Laws

We assume initially that the signal amplitude loss function over distance r is given by the following deterministic propagation law

$$a(r) = \frac{K}{r^m}, \tag{2}$$

where the constant K depends on the transmitted power. The attenuation factor m can vary from slightly more than 1 for hallways within buildings to larger than 3 for dense urban environments and office buildings. Combining (1) and (2), the noise equation is

$$\mathbf{Y} = \sum_{i=1}^{\infty} \frac{K}{r_i^m} \mathbf{X}_i. \tag{3}$$

In the Appendix A, we sketch the proof of the following:

Theorem 1 *If the RVs $\mathbf{X}_i$ are i.i.d. and SS and the interferers/scatterers form a Poisson field, then the characteristic function of the interference vector $\mathbf{Y}$ in (3) is SS α-stable, i.e.,*

$$\phi_{\mathbf{Y}}(\mathbf{t}) = \exp(-\gamma ||\mathbf{t}||^{\alpha}), \tag{4}$$

where $\alpha = \frac{2}{m}$ and $\alpha = \frac{3}{m}$ for interferers distributed in the plane and volume, respectively. The parameter γ, called dispersion, is given as

$$\gamma = -\lambda \mathcal{P} K^{\alpha} \int_0^{\infty} \frac{\Phi_0'(x)}{x^{\alpha}} dx, \tag{5}$$

or equivalently as

$$\gamma = \lambda \mathcal{P} K^{\alpha} C_{\alpha}^{-1} E \mid X_{i,j} \mid^{\alpha}, \tag{6}$$

where $\Phi_0(x) = \Phi_{\mathbf{X}}(||\mathbf{t}||)$ is a generating characteristic function of the SS RVs $\mathbf{X}_i$ [3] ; $'$ denotes differentiation, and $C_{\alpha} = \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}$. The constant $\mathcal{P} = \pi$ for interferers in the plane, and $\mathcal{P} = \frac{4}{3}\pi$ for scatterers in the volume.

For $\mathbf{X}_i = [\cos(\Theta_i), \sin(\Theta_i)]$ (non-coherent on-off FSK), with Θ_i uniformly distributed in $(0, 2\pi]$, we have $\Phi_0(x) = J_0(x)$, where $J_{\nu}(\cdot)$ is a νth order Bessel function of the first kind [4]. This model for $\mathbf{X}_i$ is assumed in many radar applications. Because $J_0'(x) = -J_1(x)$, the formula 6.561.17 from [2] can be used in (5), to calculate that the dispersion of the SS stable RV $\mathbf{Y}$ in (3) with the deterministic power propagation law as in (2) is given for $0.5 < \alpha < 2$ by

$$\gamma_{\text{determ}} = \lambda \mathcal{P} K^{\alpha} \frac{\Gamma(1 - \alpha/2)}{2^{\alpha}\Gamma(1 + \alpha/2)}. \tag{7}$$

In this equation, the admissible range of the path loss exponent is $1 < m < 4$ for interferers distributed in the plane, and $\frac{3}{2} < m < 6$ for scatterers distributed in the volume.

4 Stable Interference Models in Lognormal Shadowing and Rayleigh Fading Environments

So far we have assumed that the received signal strength decreases with range raised to some exponent. However, experimental results show that this is only the average behavior of the signal. The received signal at fixed range is not constant because of different terrain characteristics and statistical fluctuations in propagation conditions. Typically, the following random effects should be included in a study: (i) the random link attenuation due resulting from lognormal shadowing and (ii) Rayleigh fading.

In the presence of lognormal shadowing, the pdf of the signal strength is of the form [3]

$$p(a(r)|\widehat{a(r)}) = \frac{1}{\sqrt{2\pi}\sigma a(r)} \exp[-\frac{1}{2\sigma^2} \ln^2(\frac{a(r)}{\widehat{a(r)}})], \tag{8}$$

[1] The projection of a continuous-time waveform transmitted by the i-th terminal $x_i(t)$ onto $\varphi_j(t)$, or equivalently the correlation of these two, is given as $X_{i,j} \triangleq \int_0^T \varphi_j(t) x_i(t) dt$, where T is a symbol interval.

[2] The random vector $\mathbf{X}$ is said to be SS if its characteristic function $\Phi_{\mathbf{X}}(\mathbf{t})$ depends only on the Euclidean (L_2) norm of $\mathbf{t}$, i.e., $\Phi_{\mathbf{X}}(\mathbf{t}) = \phi(||\mathbf{t}||)$, where $||\mathbf{t}|| = (\sum_{i=1}^n t_i^2)^{\frac{1}{2}}$.

[3] Here, $\Phi_0(x)$ is a function of the scalar variable $x = ||\mathbf{t}||$, which for a SS RV $\mathbf{X}_i$ does not depend on n.

where $\widehat{a(r)} = \frac{K}{r^m}$ is the median of $a(r)$ as given in (2) and $\sigma = 0.5\sigma_s$. The parameter σ_s is the standard deviation of the instantaneous power, and it depends on the environment. Values of σ_s on the order of 8 to 10 dB are reported in the literature [3]. So in order to include the lognormal shadowing effect in our model, we have to consider $a(r)$ in (1) to be a RV given as $a(r) = \frac{K}{r^m}\exp(\sigma G)$, where G is the standard Gaussian RV with zero mean ($G \sim \mathcal{N}(0,1)$). The interfering signal is then

$$\mathbf{Y} = K\sum_{i=1}^{\infty}\frac{1}{r_i^m}\exp(\sigma G_i)\mathbf{X}_i. \qquad (9)$$

We assume here that G_i are i.i.d. The hypothesis of independence between shadowing effects from different users is generally accepted [4]. Therefore, we can apply *Theorem 1* in (9) with $\mathbf{X}_i$ replaced by $\exp(\sigma G_i)\mathbf{X}_i$ [4]. Then, in environments with lognormal shadowing, $\mathbf{Y}$ is again α-stable with $\alpha = \frac{2}{m}$ and $\alpha = \frac{3}{m}$ for interferers distributed in the plane and volume, respectively. To calculate the dispersion, we use (6)

$$\begin{aligned}\gamma &= \lambda K^\alpha \mathcal{P} C_\alpha^{-1} E \mid \exp(\sigma G_i) X_{i,j} \mid^\alpha \\ &= \lambda K^\alpha \mathcal{P} C_\alpha^{-1} E \mid X_{i,j} \mid^\alpha E \mid \exp(\sigma G_i) \mid^\alpha \qquad (10)\\ &= \gamma_{\text{determ}} \exp(\tfrac{1}{2}\alpha^2\sigma^2),\end{aligned}$$

where γ_{determ} is a dispersion of the corresponding system with the deterministic power propagation law. The last equation in (10) follows from the first moment relation for lognormal RVs.

If $a(r)$ is Rayleigh distributed, for a given r, $a(r)$ can be represented [3] as $a(r) = \frac{K}{r^m}\sqrt{\frac{2}{\pi}}\mathcal{R}$, where the RV $\mathcal{R} = \sqrt{G_I^2 + G_Q^2}$ is Rayleigh distributed with $G_I, G_Q \sim \mathcal{N}(0,1)$. Then, we have to substitute $\sqrt{\frac{2}{\pi}}\mathcal{R}_i\mathbf{X}_i$ for $\mathbf{X}_i$ in Theorem 1, and $\mathbf{Y}$ is α-stable with the same characteristic exponent as in the deterministic power propagation scenario. The dispersion is calculated in the same fashion as in (10). Because $E|\mathcal{R}|^\alpha = 2^{\frac{\alpha}{2}}\Gamma(1+\frac{\alpha}{2})$, the dispersion is

$$\gamma = \gamma_{\text{determ}}(\tfrac{4}{\pi})^{\frac{\alpha}{2}}\Gamma(1+\tfrac{\alpha}{2}). \qquad (11)$$

The dispersion factors $\frac{\gamma}{\gamma_{\text{determ}}}$ for lognormal shadowing, Rayleigh fading and combined shadowing and fading are shown in Fig. 1 as a function of α. The curves are plotted with the shadowing standard deviation $\sigma_s = 10$ dB. We see that in all cases examined, the dispersion factors are increasing functions of α.

[4] The RVs $\exp(\sigma G_i)\mathbf{X}_i$ are spherically symmetric (SS) because a product of a univariate RV and a SS RV is SS. Also, they are independent because $\{G_i\}$ and $\{\mathbf{X}_i\}$ are assumed to be independent sequences of mutually independent RVs.

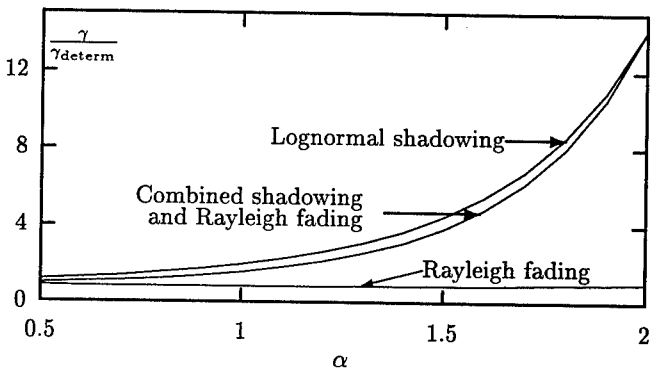


Figure 1. Dispersion factor for Rayleigh fading and lognormal shadowing ($\sigma_s = 10$ dB).

5 Practical Considerations

In Section 2, we have made two idealized assumptions: (i) we assumed an infinite number of interferers; and (ii) we assumed that an interfering signal was present for the entire duration of the matched-filtering interval. In [3], we show that these assumptions do not constrain our analysis. Moreover, we demonstrate that the α-stable model applies when interferers form non-homogeneous Poisson fields. The last result is achieved by mapping the processes in the plane (volume) into homogeneous processes on the line. This is because the LePage series representation applies only to Poisson processes with a constant rate. For example, if the non-homogeneous point processes in the plane has the rate function $\lambda(r) = \lambda_0 r^{\beta-2}$ and $\beta < 2m$, then r_i^β represents the homogeneous Poisson process with the rate $\lambda = \lambda_0\frac{2\pi}{\beta}$. With this result, we can proceed as in Section 2 and arrive at the stable model with $\alpha = \frac{\beta}{m}$ and $\gamma = -\lambda_0\frac{2\pi}{\beta}K^\alpha\int_0^\infty\frac{\Phi_0'(x)}{x^\alpha}dx$. Also, if we assume that the interferers are Poisson distributed only in a sector of the plane with an angle ϕ_s and that their density is λ, then we can map such a process to a homogeneous Poisson point process in the whole plane with the rate $\lambda^* = \lambda\frac{\int_0^{\phi_s}d\phi}{2\pi} = \lambda\frac{\phi_s}{2\pi}$. The latter scenario is applicable to directional antennas as opposed to omnidirectional discussed so far.

6 Concluding Remarks

In this paper, we have characterized interference for multiple access communication systems in which interferers are assumed to be Poisson-distributed in the plane. The same development applies to radar clutter.

Assuming the average inverse power attenuation of signal strength over distance, interference in the system is shown to be an SS α-stable noise. This model specifies system noise with two parameters: the characteristic exponent α and the dispersion γ. The formulas derived in the paper allow us to predict noise statistics in environments with lognormal shadowing and Rayleigh fading.

The hypothesis of α-stable noise is partially confirmed by the impulsive character of clutter and multiple access interference. But in the end, it must be resolved against experimental data. Alpha-stable noise model verification in radar applications is currently underway and the results will be announced shortly.

Appendix A

Our proof of Theorem 1 is based on the generalized LePage series representation of SS α-stable distributions:

Theorem 2 *Let $\{\tau_i\}$ denote the "arrival times" of a Poisson process*[5] *with rate λ and let $\{\mathbf{X}_i\}$ be SS i.i.d. vectors in $\mathbb{R}^n$ satisfying $E|\mathbf{X}_i|^\alpha < \infty$, or equivalently $E|X_{i,j}|^\alpha < \infty$. Then*

$$\mathbf{Y} = \sum_{i=1}^{\infty} \tau^{-\frac{1}{\alpha}} \mathbf{X}_i \tag{12}$$

converges a.s. to a SS α-stable random vector $\mathbf{Y}$ with the characteristic function (ch.f.)

$$\phi_{\mathbf{Y}}(\mathbf{t}) = \exp(-\gamma ||\mathbf{t}||^\alpha). \tag{13}$$

The dispersion parameter γ is given as

$$\gamma = -\lambda \int_0^\infty \frac{\Phi_0'(x)}{x^\alpha} dx. \tag{14}$$

The detailed proof of Theorem 2 can be found in [3]. To link the multivariate version of the LePage series in Theorem 2 with the noise equation in (3), we need to map a Poisson point process in the plane (volume) onto the homogeneous Poisson process on the line. To achieve this, we use the following proposition which results from the mapping theorem of Poisson point processes [6]:

Proposition 1 *For a homogeneous Poisson point process in the plane (volume) with the rate λ, assuming that points are at distances r_i ($r_1 < r_2 < \cdots$) from the origin, $\Gamma_i = r_i^2$ ($\Gamma_i = r_i^3$) represents Poisson arrival times on the line with the constant arrival rate $\pi\lambda$ ($\frac{4}{3}\pi\lambda$).*

[5] In this paper, we use the term *arrival times* or *occurrence times* of a Poisson process to mean a Poisson process on the line, where time is just a hypothetical variable.

Now, for interferers distributed in the plane, we rewrite $\mathbf{Y}$ in (3) as

$$\mathbf{Y} = K \sum_{i=1}^{\infty} \frac{1}{(r_i^2)^{\frac{m}{2}}} \mathbf{X}_i. \tag{15}$$

From *Proposition 1*, $\Gamma_i = r_i^2$ represents Poisson "occurrence" times on the line with the arrival rate $\pi\lambda$, and based on *Theorem 2* [6], $\mathbf{Y}$ is SS α-stable with the characteristic exponent $\alpha = \frac{2}{m}$ and dispersion $\gamma = -\lambda\pi K^\alpha \int_0^\infty \frac{\Phi_0'(x)}{x^\alpha} dx$. The multiplicative constant K changes the dispersion of α-stable RV by K^α [7]. Similar proof follows for interferers distributed in th volume. The equivalence of Eqs. (5) and (6) follows from the integral formula ([2], 3.823)

$$\int_0^\infty \frac{1-\cos(zt)}{t^{\alpha+1}} dt = |z|^\alpha \frac{\Gamma(1-\alpha)\cos(\frac{\pi}{2}\alpha)}{\alpha} \tag{16}$$

by replacing the constant z with RV $X_{i,j}$ and taking expectation of both sides.

References

[1] C.L.Nikias and M.Shao. *Signal Processing with Alpha-Stable Distributions and Applications.* John Wieley & Sons, Inc., New York, 1995.

[2] I. Gradshteyn and I. Ryzhik. *Table of Integrals, Series, and Products.* Academic Press Inc., San Diago, 1994.

[3] J. Ilow. *Signal Processing in Alpha-Stable Noise Environments: Noise Modeling, Detection and Estimation.* Ph.D. thesis, University of Toronto, Toronto, 1996.

[4] J. Ilow, D. Hatzinakos and A.N. Venetsanopoulos. Performance bounds evaluation of FH SS radio networks with interference modeled as a mixture of Gaussian and alpha-stable noise. *PIMRC'95, Toronto*, pages 153–157, Sept. 1995.

[5] A. Janicki and A. Weron. *Simulation and Chaotic Behavior of α-Stable Stochastic Processes.* Marcel Dekker, Inc., New York, 1994.

[6] D. J. Kingman. *Poisson Processes.* Ford Science Publications, Oxford, 1993.

[7] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes.* Chapman & Hall, New York, 1994.

[8] E. Sousa. Performance of a spread spectrum packet radio network link in a Poisson field of interferences. *IEEE Trans. Inform. Theory*, 38, No.6:1743–1754, Nov 1992.

[9] G. Tsihrintzis and C. Nikias. Fast estimation of the parameters of alpha-stable impulsive interference using asymptotic extreme value theory. *Proc. ICASSP'95, Detroit, MI.*, pages 1840–1843, 1995.

[6] Here we use Γ_i rather than τ_i as given in *Theorem 2* to emphasize that the Poisson process on the line results from our mapping of squared distances.

Robustness of Fractionally-Spaced Equalization by CMA to Lack of Channel Disparity and Noise

A. Touzni, I. Fijalkow
ENSEA / ETIS
6 av. du Ponceau,
95014 Cergy-Pontoise Cedex, France
touzni, fijalkow@ensea.fr

J.R. Treichler
Applied Signal Technology Inc.
Sunnyvale, CA, USA

Abstract

We study the Fractionally-Spaced Equalization by CMA (FSE-CMA) robustness to channel noise and lack of disparity. When there is lack of disparity, we will show that, whereas other recent technics as linear prediction or subspace like methods fail, FSE-CMA can still equalize. In particular for long enough equalizer FSE-CMA exhibits a "smoothing effect" which leads to an interesting trade-off between achieving zero-forcing equalization and noise enhancement.

1. Introduction

Constant Modulus Algorithm (**CMA**)[1], is one of the most commonly used blind algorithm to suppress InterSymbol Interference (**ISI**) in digital transmission systems. It is called **FSE-CMA** (Fractionally-Spaced Equalization by CMA) when used in a **channel diversity** scheme generated by either oversampling the received data or multivariate data observed behind a sensors array. In a previous work ([2]), it has been shown that the FSE-CMA criterion minimization achieves perfect equalization (in the noise-free context) under the so-called **Zero-forcing** conditions (no common zero in the multichannel transfer function, i.e., $c_0(z)=1$ in Figure 1 and a long enough equalizer)([2], [3]). Moreover, in the contrary of the second-order statistics based methods ([4], [3], [5], [6]...), FSE-CMA still performs reasonable equalization even when there is lack of **channel disparity** (i.e., $c_0(z) \neq 1$) (see the noise-free preliminary study [2], for instance). Furthermore we have shown, in a previous study, that under ZF conditions FSE-CMA exhibits some robustness to channel noise ([7]).

In this contribution we are motivated by the desire to evaluate the FSE-CMA global performance criterion in realistic noisy conditions. So, we address the effect of additive white noise and lack of channel disparity on the FSE-CMA criterion, in terms of the input-output remaining mean square error. This will also allow to define an equalizability bound that will permit to compare the optimal FSE-CMA performance to other recent Fractionally-Spaced Equalization technics.

2. FSE under lack of disparity

Under lack of disparity, we consider the Fractionally-Spaced model driven by a zero-mean i.i.d. sequence $s(n)$ and corrupted by an L-dimensional additive noise $\vec{w}(n) = (\vec{w}_1(n), \cdots, \vec{w}_L(n))^T$ (Figure 1).

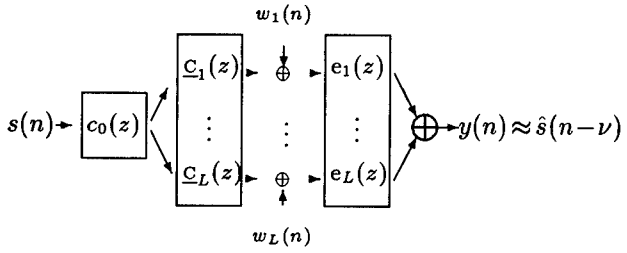


Figure 1. **FSE Scheme Under Lack of Disparity**

The linear equalization problem consists on choosing the "best" L-variate Finite Impulse Response (**FIR**) equalizer transfer function $\vec{e}(z)$, of degree N, such as $y(n) \approx s(n-\nu)$, with ν is an arbitrary delay. (each $e_k(z)$ writes as $e_k(z) = \sum_{p=0}^{N} e_{k,p} z^{-p}$). The channel is described by $c_0(z)$ a possibly non-minimum phase scalar transfer function of degree Z_0 and $\underline{c}(z)$ an L-variate FIR non-reducible vector transfer function (i.e., there is no common zero to all components $\underline{c}_k(z)$ of degree $(Q-Z_0)$).

This problem formulation is turned on choosing the NL long equalizer impulse response $\vec{e}$, such as:

$$y(n) = \left(\vec{e}^T \underline{\mathbf{C}} \mathbf{C}_0\right) S(n) + \vec{e}^T \vec{W}(n) \approx \hat{s}(n-\nu) \qquad (1)$$

where $S(n)$ contains the last $(N+Q)$ input symbols in the past of $s(n)$ (with $N-1>Q$) and $\vec{W}(n)$ is the multivariate noise vector $(\vec{w}(n),\cdots,\vec{w}(n-N+1))^\top$. Hence, $\mathbf{C}_0$ denotes the $(Q-Z_0+N)\times(N+Q)$ channel convolution matrix associated to $c_0(z)$ and $\underline{\mathbf{C}}$ denotes the $(NL)\times(Q-Z_0+N)$ channel convolution matrix associated to the multichannel minimum phase transfer function $\underline{\vec{c}}(z)$. Note that, $\underline{\mathbf{C}}$ and $\mathbf{C}_0$ are respectively full column-rank and full row-rank Sylvester matrices ([2]).

3. Smoothed FSE-CMA criterion

Lemma 1 *The FSE-CMA normalized criterion ([1]) is:*

$$J_\gamma(\vec{e})=E\left[(r_2-y^2(n))^2\right]/E\left[s^2\right]^2 \qquad (2)$$

with $r_2=E[s^4]/E[s^2]$, and $E[.]$ stands for the mean expectation operator. Assuming independence between the noise and source signals and a temporally/spatially white gaussian noise, we get:

$$J_\gamma(\vec{e})=J_0(\vec{e})+\gamma\,\|\vec{e}\|^2\left\{2\left(3\|\,\mathbf{C}_0^\top\underline{\mathbf{C}}^\top\vec{e}\,\|^2-\rho\right)+3\gamma\,\|\vec{e}\|^2\right\} \quad (3)$$

where $J_0(\vec{e})$ is the noise-free cost-function, γ is the noise to signal power ratio writes as , $\gamma=E[w^2]/E[s^2]$ and $\rho=E[s^4]/E[s^2]^2$ is the input signal kurtosis.

Proof: *Using (1), the proof is deduced by a straightforward calculus (see [2]). Note that the expression can be easily extended for a non gaussian noise since we know the fourth-order moment of the noise contribution.* □

3.1. Further results under lack of disparity

The channel/equalizer impulse response setting zeroing the noise-free cost-function $J_0(\vec{e})$ writes as:

$$h_\nu=\mathbf{C}_0^\top\underline{\mathbf{C}}^\top\vec{e}\;\Leftrightarrow\;\underline{\vec{c}}(z)^\top\vec{e}(z)=z^{-\nu}/c_0(z) \qquad (4)$$

which is not possible with a FIR equalizers. Because of lack of disparity, the best achievable h may be far from any optimal setting $h_\nu=(0\cdots010\cdots0)^\top$. More precisely, the only achievable impulse responses $h=\mathbf{C}_0^\top\underline{e}$ live in the subspace spanned by the columns of $\mathbf{C}_0^\top$. In particular, the closest to h_ν achievable h is given by the orthogonal projection of h_ν on the range of $\mathbf{C}_0^\top$, $h=\mathbf{C}_0^\top(\mathbf{C}_0\mathbf{C}_0^\top)^{-1}\mathbf{C}_0h_\nu$, We set $\mathbf{\Pi}_0=\mathbf{C}_0^\top(\mathbf{C}_0\mathbf{C}_0^\top)^{-1}\mathbf{C}_0$. In fact, for a given achievable $h=\mathbf{C}_0^\top\underline{e}$ there exists a unique $\underline{e}$ such as $\underline{e}=\underline{\mathbf{C}}\vec{e}$ and $NL-(N+Q-Z_0)$ possible settings for $\vec{e}$. In this case, the cost-function extrema of $J_0(\vec{e}(\underline{e}))$, satisfy $\mathbf{C}_0\Delta(\mathbf{C}_0^\top\underline{e})\mathbf{C}_0^\top\underline{e}=0$, with $\Delta(\mathbf{C}_0^\top\underline{e})=(3\|\mathbf{C}_0^\top\underline{e}\|^2-\rho)I-(3-\rho)\,\mathrm{diag}(\mathbf{C}_0^\top\underline{e}\,\underline{e}^\top\mathbf{C}_0)$, where $diag(A)$ stands for the matrix extracted from A with the same diagonal entries and 0 elsewhere. They can be classified as ([2]):

- one maximum ($\underline{e}=0$),
- global minima (when $\mathbf{C}_0^\top\underline{e}=h_\nu$ is achievable with $\underline{e}\neq0$) or saddle points ($\Delta(\mathbf{C}_0^\top\underline{e})\,\mathbf{C}_0^\top\underline{e}=0$ and $\mathbf{C}_0^\top\underline{e}\neq h_\nu$),
- local minima ($\mathbf{C}_0^\top\Delta(\mathbf{C}_0^\top\underline{e})\,\mathbf{C}_0^\top\underline{e}=0$ and $\underline{e}$ does not belong to the previous categories).

Note that, a potential global minimum imply that the corresponding $\underline{e}$ is expressed as $(\mathbf{C}_0\mathbf{C}_0^\top)^{-1}\mathbf{C}_0h_\nu$, i.e., $h=\mathbf{\Pi}_0h_\nu$. Since $\mathbf{C}_0$ can not be square (it is a $(N+Q-Z_0)\times(N+Q)$ matrix), there should exist no global minima such as $h=h_\nu$. However, when N becomes "large", $\mathbf{C}_0$ tends to become square so that $\mathbf{\Pi}_0$ becomes close to the identity matrix. Of course, as in the non-fractional case, undesired settings may exist. However, the larger N is, the closer the corresponding channel / equalizer is becoming to some h_ν.

3.2. Perturbation in noisy case

From lemma 1, one can see in noisy context, that $J_0(\vec{e})$ is regularized by an additional deterministic factor $\Phi_\gamma(\vec{e}(\underline{e}))$ driven by γ. This leads to a balance between the minimization of criterions: J_0 and Φ_γ. The result is a "smoothing effect" expressed through twofolds constraints: (**i**) a minimization of $\|\,\mathbf{C}_0^\top\underline{\mathbf{C}}^\top\vec{e}\,\|^2$ that leads to get an impulse response $\|h\|$ as small as possible; (**ii**) a minimization on $\|\vec{e}\|$ which tends to forbid the equalizer norm to be too high, reducing consequently the noise enhancement $\gamma\|\vec{e}\|^2$ (see (1)).

Thus, the minima of $J_\gamma(\vec{e})$ realize a desirable balance between the noise-free good equalization settings and the noise enhancement due to the equalizer norm. To solve the minimization problem, we propose a two-steps minimization procedure. First, we minimize $J_\gamma(\vec{e})$ over the subspace of vectors $\vec{e}$ such as $\underline{e}=\underline{\mathbf{C}}^\top\vec{e}$ for a given $\underline{e}$. The resulting value of $\vec{e}$ is a function of $\underline{e}$, denoted $\vec{e}(\underline{e})$. Then, we minimize $J_\gamma(\vec{e}(\underline{e}))$ over the subspace of $(N+Q-Z_0)$-long vectors $\underline{e}$. Invoking equation (3), the two steps minimization can be expressed as

$$\min_{\underline{e}}J_\gamma(\vec{e}(\underline{e}))=\min_{\underline{e}}\{J_0(\vec{e}(\underline{e}))+\gamma\min_{\vec{e},\,\underline{e}=\underline{\mathbf{C}}^\top\vec{e}}\{\Phi_\gamma(\vec{e}(\underline{e}))\}\} \quad (5)$$

The procedure is simplified because $J(\vec{e})$ is a function of $\underline{e}$ only, so that the first step consists of the smoothing cost-function $\Phi_\gamma(\underline{e})$ minimization only.

The first step minimization of the quadric cost-function $\Phi_\gamma(\vec{e}(\underline{e}))$ (for a given $\underline{e}$), under the linear constraint $\underline{e}=\underline{\mathbf{C}}^\top\vec{e}$, can be performed using Lagrange multiplier technic and leads to the zero-order approximation (for a SNR large enough):

$$\vec{e}(\underline{e})=\underline{\mathbf{C}}(\underline{\mathbf{C}}^\top\underline{\mathbf{C}})^{-1}\underline{e}+o(1)$$

The second step consists of minimizing:

$$J_\gamma(\vec{e}(\underline{e}))=J_0(\vec{e}(\underline{e}))+\gamma\,\|\vec{e}(\underline{e})\|^2(3\|h(\underline{e})\|^2-\rho)+o(\gamma) \quad (6)$$

where the noise-free cost-function $J_0(\vec{e}(\underline{e})) = E[(r_2 - (\underline{e}^T \mathbf{C}_0 S(n))^2)^2]$. Taking the derivative of (6) with respect to $\underline{e}$, the extrema of (3) are solution of the equation:

$$4\,\mathbf{C}_0\Delta(\mathbf{C}_0^T\underline{e})\mathbf{C}_0^T\underline{e} \;+\; 2\gamma\left[3\,\|\underline{\mathbf{C}}(\underline{\mathbf{C}}^T\underline{\mathbf{C}})^{-1}\underline{e}\|^2\,\mathbf{C}_0\mathbf{C}_0^T\,\underline{e} + (3\|\mathbf{C}_0^T\underline{e}\|^2-\rho)(\underline{\mathbf{C}}^T\underline{\mathbf{C}})^{-1}\underline{e}\right]+o(\gamma)=0 \quad (7)$$

Our task is to provide a close-form solution to the equation (7). Since we don't know how to explicit the noise-free minima expression, we consider herein the perturbation of $\underline{e}=\underline{e}_\nu = (\mathbf{C}_0\mathbf{C}_0^T)^{-1}\mathbf{C}_0 h_\nu$, corresponding to $h = h_\nu$. We know that for a "large enough" equalizer length it will be a good approximation in the noise-free case.

4. Close-form extrema

In order to get some insight in the noisy case, we assume that the approximation error is "smaller" than the perturbation due to the noise. This should hold for "large" values of N and "not too small" values of γ. In the same time, γ must be small enough to allow a first order approximation in terms of γ. The validity of this assumption is checked by simulations in the sequel.

Proposal 1 *For a small enough γ, we assume the global channel-equalizer setting $\underline{e}_\gamma$ to be a first order perturbation of $\underline{e}_\nu = (\mathbf{C}_0\mathbf{C}_0^T)^{-1}\mathbf{C}_0 h_\nu$ in terms of γ as,*

$$\underline{e}_\gamma = \underline{e}_\nu + \gamma\bar{\underline{e}}_\nu + o(\gamma) \quad (8)$$

Then, $\bar{\underline{e}}_\nu$ satisfies,

$$\bar{\underline{e}}_\nu \approx -\frac{3}{2}\,\underline{e}_\nu^T(\underline{\mathbf{C}}^T\underline{\mathbf{C}})^{-1}\underline{e}_\nu\;(\mathbf{C}_0\Psi_\nu\mathbf{C}_0^T)^{-1}\,\mathbf{C}_0\mathbf{C}_0^T\,\underline{e}_\nu - \frac{(3-\rho)}{2}\,(\mathbf{C}_0\Psi_\nu\mathbf{C}_0^T)^{-1}(\underline{\mathbf{C}}^T\underline{\mathbf{C}})^{-1}\underline{e}_\nu \quad (9)$$

The corresponding channel / equalizer settings can be viewed as a perturbation of h_ν,

$$h_\gamma = \mathbf{C}_0^T\underline{e}_\gamma \approx h_\nu - \frac{\gamma}{2}\mathbf{C}_0^T(\mathbf{C}_0\Psi_\nu\mathbf{C}_0^T)^{-1}\mathbf{C}_0\left[3\underline{e}_\nu^T(\underline{\mathbf{C}}^T\underline{\mathbf{C}})^{-1}\underline{e}_\nu h_\nu + (3-\rho)\mathbf{C}_0^H(\underline{\mathbf{C}}^T\underline{\mathbf{C}})^{-1}\mathbf{C}_0^{HT}h_\nu\right]+o(\gamma) \quad (10)$$

where Ψ_ν is a $(N+Q)\times(N+Q)$ diagonal matrix with entry $(3-\rho)$ when $i\neq\nu+1$ and 2ρ when $i=\nu+1$, and $\mathbf{C}_0^H = \mathbf{C}_0^T(\mathbf{C}_0\mathbf{C}_0^T)^{-1}$.

Note that, for a large value of N, the symbol $\approx$ in (9) stands for the approximation of $\mathbf{C}_0^T\underline{e}_\nu \approx h_\nu$. If we assume in addition that the input is constant modulus, $\rho=1$ and $\Psi_\nu = 2I$, the global impulse response minima are of the form:

$$h_\gamma \approx h_\nu - \frac{\gamma}{2}\left[3\,(h_\nu^T\mathbf{C}_0^H(\underline{\mathbf{C}}^T\underline{\mathbf{C}})^{-1}\mathbf{C}_0^{HT}h_\nu)\;h_\nu + \mathbf{C}_0^H(\underline{\mathbf{C}}^T\underline{\mathbf{C}})^{-1}\mathbf{C}_0^{HT}\,h_\nu\right] + o(\gamma) \quad (11)$$

This result is similar to the expression of h_γ when ZF is exactly achievable (i.e., when $(\mathbf{C}^T\mathbf{C})^{-1}$ is replaced by $\mathbf{C}_0^H(\underline{\mathbf{C}}^T\underline{\mathbf{C}})^{-1}\mathbf{C}_0^{HT}$. We notice once again that FSE-CMA criterion has very specific properties for constant modulus input signals.

Proof of Proposal 1: *Introducing assumption (8) in the equation (7), the proof consists on evaluating $\bar{\underline{e}}_\nu$; i.e., the first order solution (in terms of γ) of the equation (7). Since $\mathbf{C}_0^T\underline{e}_\nu \approx h_\nu$ for a large enough N, we obtain easily $\bar{\underline{e}}_\nu$ as a solution of the linear system: $2\,\mathbf{C}_0\Psi_\nu\mathbf{C}_0^T\bar{\underline{e}}_\nu = -3\underline{e}_\nu^T(\underline{\mathbf{C}}^T\underline{\mathbf{C}})^{-1}\underline{e}_\nu\mathbf{C}_0\mathbf{C}_0^T\underline{e}_\nu + (3-\rho)\left(\underline{\mathbf{C}}^T\underline{\mathbf{C}}\right)^{-1}\underline{e}_\nu + o(1)$. Where $\mathbf{C}_0\Psi_\nu\mathbf{C}_0^T$ is invertible if the input signal is not gaussian (i.e., $\rho\neq 3$).*□

Simulations: A 2-dimensional multichannel vector $\vec{c}(z)$ is defined by the zeros of each transfer function as $c_1(z) = (-1.4, -0.4)$ and $c_2(z) = (1.1, -0.4)$. The observation number is set to N=8. Figure 2 displays the impulse response taps of h_γ (obtained by running the algorithm to minimize the criterion) versus SNR. Note that h_γ is very close to a canonical vector for a large enough N and SNR. In Figure 3, we display the analytical impulse response introduced in Proposal 1. We can see that both curves are very close.

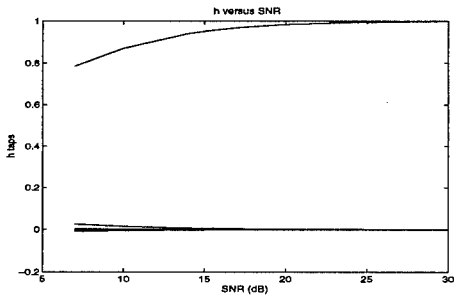


Figure 2: **h versus SNR**

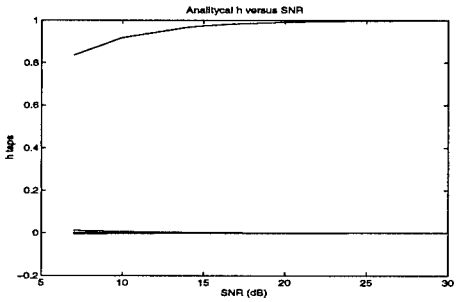


Figure 3: **analytical h versus SNR**

5. Mean square input/output error

In this subsection we are motivated in evaluating the mean equalizability performance in terms of normalized input/output mean square error (MSE) defined as $E[(y(n)-s(n-\nu))^2]/E[s^2]$.

Proposal 2 *The MSE is as a sum of residual ISI and noise enhancement (driven by γ). For a large enough SNR the FSE-CMA MSE can be approximated by:*

$$\underbrace{\|(I-\mathbf{\Pi}_0)h_\nu\|^2}_{Zero-Forcing}+\gamma\underbrace{h_\nu^\top\mathbf{C}_0^H(\underline{\mathbf{C}}^\top\underline{\mathbf{C}})^{-1}\mathbf{C}_0^{H\top}h_\nu}_{Noise\ Enhancement}+o(\gamma) \quad (12)$$

The equalizability bound is expressed as the sum of an irreducible error due to the pseudo-inversion of $c_0(z)$ and a linear error proportional to γ. Note that for a long enough equalizer, $\mathbf{\Pi}_0 \approx \mathbf{I}$, then the MSE is mostly due to noise enhancement.

Proof of Proposal 2: *The MSE writes as MSE* $= \|h-h_\nu\|^2+\gamma\,\|\vec{e}\|^2$. *Introducing the parametrization* $h=\mathbf{C}_0^\top\underline{e}$ *and the assumption (8), we get* $\vec{e}_\gamma=\vec{e}(\underline{e}_\nu)=\underline{\mathbf{C}}(\underline{\mathbf{C}}^\top\underline{\mathbf{C}})^{-1}(\mathbf{C}_0\mathbf{C}_0^\top)^{-1}\mathbf{C}_0h_\nu+o(1)=\underline{\mathbf{C}}(\underline{\mathbf{C}}^\top\underline{\mathbf{C}})^{-1}\underline{e}_\nu+o(1)$, *and* $h_\gamma=\mathbf{C}_0^\top\underline{e}_\gamma=\mathbf{C}_0^\top\underline{e}_\nu+o(1)$ *which yields immediately to (12)* □.

An interesting point is to notice that the first-order FSE-CMA MSE is the same as the MMSE deduced by minimization of $E[(y(n)-s(n-\nu))^2]/E[s^2]$, even if the channel equalizer global impulse response minima differ between criterions.

Simulations: We use a 2-dimensional multichannel vector $\underline{c}(z)$ is defined by the zeros of each transfer function as $(-1.4, 1.1)$ and we take $c_0(z)=z+0.4$. The observations number is set to $N=8$ (Figure 4) and $N=2$ (Figure 5). Both curves show the accuracy between the experimental and the analytical FSE-CMA MSE (12). In Figure 4, N is long enough to have to have $\|(I-\mathbf{\Pi}_0)h_\nu\|^2\approx 0$. In Figure 5 the analytical curve $(-)$ is the sum of the experimental irreducible zero-forcing $(.-)$ and the linear Noise-Enhancement error $(--)$.

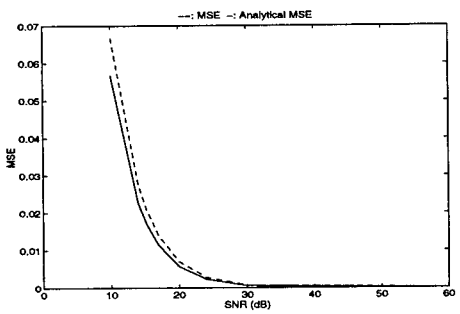


Figure 4: **FSE-CMA MSE (N=8)**

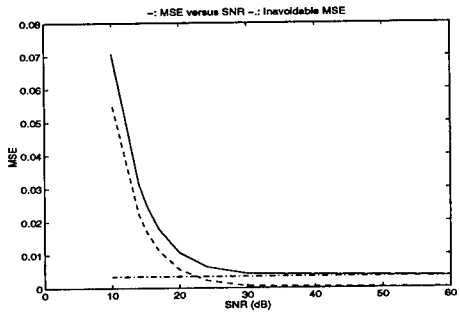


Figure 5: **FSE-CMA MSE (N=2)**

6. Conclusion

Under the realistic assumptions of lack of channel disparity and additive channel noise, we have established in this contribution an analytical close-form of the FSE-CMA global impulse response. Whereas other recent second-order methods fail, we have shown that FSE-CMA realizes an interesting trade-off between noise-enhancement and achievable equalizability. In order to evaluate the equalizability performance a close-form expression of the mean input/output steady-state square error has been derived. For large SNR value and large N, FSE-CMA performances are very similar to the best achievable performances of a blind linear equalizer.

References

[1] J.R. Treichler, and B.G. Agee, *A new approach to Multipath correction of constant modulus signals*, IEEE Tr. on ASSP, vol. 31, n. 2, pp. 459-472, 1983.

[2] I. Fijalkow, C.E. Manlove and C.R. Johnson Jr., *Adaptive Fractionally Spaced Blind Equalization*, submitted to IEEE Tr. on SP, january 1995. A partial version appears in Proc. ICASSP'95.

[3] Z. Ding, Y. Li, *On channel identification based on second-order cyclic spectra*, IEEE Tr. on SP, pp.1260-1264, May 1994.

[4] L. Tong, G. Xu, and T. Kailath, *Blind identification and equalization based on second-order statistics: a time domain approach*, IEEE Trans. on IT, vol. 40, n. 2, pp. 340-349, 1994.

[5] E. Moulines, et al., *Subspace methods for the blind identification of multichannel FIR filters*, IEEE Tr. on SP, 1995.

[6] D.T.M. Slock, *Blind fractionally-spaced equalization, perfect reconstruction filter banks and multichannel linear prediction*, in Proc. ICASSP'94, 1994.

[7] A. Touzni, I. Fijalkow, J.R. Treichler, *Fractionally-spaced CMA under channel noise*, in Proc. ICASSP'96, 1996.

AN IMPROVED CYCLIC ADAPTATIVE BEAMFORMING (CAB) ALGORITHM FOR A MULTIPLE AGILE BEAM MOBILE COMMUNICATION SYSTEM

Yan Meng[1], *Kon Max Wong*[1] *and Ken Lazaris-Brunner*[2]

[1]Communications Research Lab, McMaster University
Hamilton, Ont., Canada, L8S 4K1, Email:mengy@mcmaster.ca
[2]Com Dev Ltd., Cambridge, Ont., Canada, N1R 7H6

ABSTRACT

In this paper, a blind adaptive beamforming algorithm is presented which improves the performance of CAB [4]. Noting that the weighting vectors of CAB are not in general proportional to signal steering vectors in the case of multiple signals, a singular vector rotation technique is used to iteratively estimate the steering vectors. Using the estimated steering vectors as the constraint matrix in the LCMV algorithm [2], better interference suppression is achieved. Computer simulations are conducted to demonstrate that the performance of the proposed algorithm is superior to that of CAB for the scenario of multiple co-channel users transmitting at the same frequency.

1. INTRODUCTION

The use of multiple high gain agile beams from a multiple element array antenna with on-board digital beamforming [1] is being considered in the next generation of mobile satellite communication systems (MSCS). The main advantages of the system are that it offers a flexible solution for channel allocation and it can actively suppress co-channel interference. Active interference suppression can be achieved by using on-board adaptation. The Linear Constrained Minimum Variance (LCMV) algorithm seems to be the most suitable adaptive beamforming method for the multiple agile beam MSCS [8]. The LCMV method requires the locations of mobile users in order to steer the high gain beam towards the desired users and place the null at specific co-channel interferences. Mobile user localization can be established by on-board processing using high-resolution techniques, which, however, can be very computationally intensive and calibration of the array is necessary. On the other hand, blind adaptive beamforming methods exploiting the cyclostationarity [3] of communication signals attract attention because of its advantages of no requirement for mobile localization and no need for array calibration.

The cyclic adaptive beamforming (CAB) algorithm [4] being one of the blind adaptive beamforming methods has been proposed as a good candidate for spatial re-use of frequency spectrum. However, the performance of CAB deteriorates when multiple desired signals are present. Here, an improved CAB algorithm is proposed which can iteratively generate a better estimation of the steering vectors of multiple signals than CAB does. Using the estimated steering vectors as the constraint matrix in the LCMV algorithm, better interference suppression is achieved. Computer simulations are conducted to demonstrate the performance of the proposed algorithm.

2. BLIND ADAPTIVE BEAMFORMING ALGORITHM

The basic idea of CAB is to formulate the cyclic (conjugate) correlation of the array output $\boldsymbol{x}(n)$ and its frequency-shifted version $\boldsymbol{u}(n) = \boldsymbol{x}(n+n_0)e^{j2\pi\alpha n}$ (or $\boldsymbol{u}(n) = \boldsymbol{x}^*(n+n_0)e^{j2\pi\alpha n}$) at a particular cyclic frequency α of the desired signals so that the interference and noise which do not exhibit the same cyclic frequency can be eliminated. It has been proved [4] that the weighting vectors of CAB corresponding to individual desired signals are the left singular vectors of the cyclic (conjugate) correlation matrix of the array output $\boldsymbol{R}^{\alpha}_{xu}$, i.e.,

$$\boldsymbol{R}^{\alpha}_{xu} = \boldsymbol{W}_{CAB}\boldsymbol{\Lambda}_s\boldsymbol{V}^{\dagger} \tag{1}$$

where $\boldsymbol{W}_{CAB}$ is the left singular vector matrix (each column of $\boldsymbol{W}_{CAB}$ denotes the weighting vector of each desired signal), $\boldsymbol{\Lambda}_s$ is the singular value matrix and $\boldsymbol{V}$ is the right singular vector matrix.

The authors acknowledge Canadian Space Agency for funding the digital beamforming project.

CAB can asymptotically achieve optimal SINR when there is a single desired signal and the weighting vector $\boldsymbol{w}_{CAB}$ is proportional to the steering vector of the desired signal. However, the performance of CAB deteriorates when there are multiple signals, each having the same cyclic frequency. This is not surprising because in general the left singular vectors are not proportional to the individual signal steering vectors unless the multiple signals are well-separated in the sense that signals have spatial separations of more than one beamwidth and/or are of very uneven power. Therefore, CAB intends to work in the scenario of single desired user, i.e., co-channel users all have different cyclic frequencies. This results in extra bandwidth consumption.

In this section, an improved CAB algorithm is proposed which exploits the fact that the left singular vectors of $\boldsymbol{R}_{xu}^{\alpha}$ and the signal steering vectors span the same subspace (signal subspace) so that a singular vector rotation technique is used to iteratively estimate the steering vectors assuming that the signals are statistically independent of each other. The steering vector model includes both the individual elemental errors and the spatial properties of the signal, thus the procedure is valid for uncalibrated array. The improved CAB algorithm allows multiple desired users, or co-channel users operating at same frequency to achieve bandwidth saving.

It is well known that the matrix $\boldsymbol{R}_{xu}^{\alpha}$ can also be rewritten as its steering vector decomposition, i.e.,

$$\boldsymbol{R}_{xu}^{\alpha} = \boldsymbol{D}\boldsymbol{R}_{s}^{\alpha}\boldsymbol{D}^{\dagger} \tag{2}$$

where $\boldsymbol{D}$ is the matrix of signal steering vectors, $\boldsymbol{R}_{s}^{\alpha}$ is the cyclic (conjugate) correlation matrix of the signals. Therefore, we can see that

$$\text{Column space of } \boldsymbol{W}_{CAB} \equiv \text{Column space of } \boldsymbol{D} \tag{3}$$

It has been proved [5] that there exists a unitary matrix $\boldsymbol{Q}$ such that

$$\boldsymbol{D}\boldsymbol{R}_{s}^{\alpha\frac{1}{2}} = \boldsymbol{W}_{CAB}\boldsymbol{\Lambda}_{s}^{\frac{1}{2}}\boldsymbol{Q} \tag{4}$$

Since $\boldsymbol{D}$, $\boldsymbol{R}_{s}^{\alpha}$ and $\boldsymbol{Q}$ are unknown, then Eq.(4) does not have a unique solution, however, *a priori* information about the structure of a steering vector can be used to iteratively find the matrices $\boldsymbol{D}$, $\boldsymbol{R}_{s}^{\alpha}$ and $\boldsymbol{Q}$ that satisfy the above equality.

The detailed procedure of iteratively solving Eq.(4) can be found in [6,7].

3. OPTIMAL-CONSTRAINED LMS WEIGHTING VECTORS

Once $\boldsymbol{D}$, i.e., the steering vectors of co-channel users are resolved, the LCMV beamforming algorithm can be used to suppress interference. The principle of the LCMV beamforming is to constrain the beamformer so that signals from the directions of interested are passed with specified gain and phase. The weighting vector $\boldsymbol{w}_k$ is chosen to minimize the output variance (power) subject to the response constraints, i.e.,

$$\min_{\boldsymbol{w}_k} \quad \boldsymbol{w}_k^{\dagger}\boldsymbol{R}_x\boldsymbol{w}_k \tag{5}$$

$$\text{s.t.} \quad \boldsymbol{D}^{\dagger}\boldsymbol{w}_k = \boldsymbol{g} \tag{6}$$

where $\boldsymbol{R}_x = E\{\boldsymbol{x}(n)\boldsymbol{x}^{\dagger}(n)\}$ is the correlation matrix of the antenna array output and $\boldsymbol{g}$ is the response vector of the form such as

$$\boldsymbol{g} = [0 \cdots 0\ 1\ 0 \cdots 0]^{T} \tag{7}$$

where "1" in $\boldsymbol{g}$ occurs at the kth position for the response to the kth desired user and "0" are the response to the interferences, and ^{T} denotes transpose.

The optimal weighting is given as, by solving the minimization in Eqs.(5) and (6),

$$\boldsymbol{w}_{k\ opt} = \boldsymbol{R}_x^{-1}\boldsymbol{D}[\boldsymbol{D}^{\dagger}\boldsymbol{R}_x^{-1}\boldsymbol{D}]^{-1}\boldsymbol{g} \tag{8}$$

Since the correlation matrix $\boldsymbol{R}_x$ is unknown *a priori*, it has to be learned by an adaptive technique. In constrained gradient-descent optimization, the weighting vector is initialized at a vector satisfying the constraint in Eq.(6), and at each iteration the weighting vector is moved in the negative direction of the constrained gradient. Thus, the adaptation can be done as

$$\begin{aligned} \boldsymbol{w}_k^{(n+1)} &= (\boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}^{\dagger}\boldsymbol{D})^{-1}\boldsymbol{D}^{\dagger})[\boldsymbol{w}_k^{(n)} \\ &\quad -\mu\hat{\boldsymbol{R}}_x(n)\boldsymbol{w}_k^{(n)}] + \boldsymbol{w}_k^{(0)} \end{aligned} \tag{9}$$

$$\boldsymbol{w}_k^{(0)} = \boldsymbol{D}(\boldsymbol{D}^{\dagger}\boldsymbol{D})^{-1}\boldsymbol{g} \tag{10}$$

where μ is a scalar to control the step size of the adaptive process and is usually chosen as

$$0 < \mu \ll 1/\lambda_{max} \tag{11}$$

with λ_{max} being the maximum eigenvalue of the correlation matrix $\boldsymbol{R}_x$. $\hat{\boldsymbol{R}}_x(n)$ in Eq.(9) denotes an estimation for $\boldsymbol{R}_x$ at the nth iteration. An available and simple approximation for $\boldsymbol{R}_x$ at the nth iteration is the outer-product of array output $\boldsymbol{x}(n)\boldsymbol{x}^{\dagger}(n)$. Substitution of this estimation into Eq.(9) gives

$$\begin{aligned} \boldsymbol{w}_k^{(n+1)} &= (\boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}^{\dagger}\boldsymbol{D})^{-1}\boldsymbol{D}^{\dagger})[\boldsymbol{w}_k^{(n)} \\ &\quad -\mu\boldsymbol{x}(n)\boldsymbol{x}^{\dagger}(n)\boldsymbol{w}_k^{(n)}] + \boldsymbol{w}_k^{(0)} \end{aligned} \tag{12}$$

$$\boldsymbol{w}_k^{(0)} = \boldsymbol{D}(\boldsymbol{D}^{\dagger}\boldsymbol{D})^{-1}\boldsymbol{g} \tag{13}$$

An alternative estimation of $\hat{\boldsymbol{R}}_x(n)$ is given by

$$\hat{\boldsymbol{R}}_x(n+N) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}(n+i)\boldsymbol{x}^{\dagger}(n+i) \tag{14}$$

Substituting Eq.(14) into Eq.(9) yields an adaptation of weighting vectors using block data (block length is N), i.e.,

$$\boldsymbol{w}_k^{(n+N)} = (\boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}^\dagger\boldsymbol{D})^{-1}\boldsymbol{D}^\dagger)[\boldsymbol{w}_k^{(n)} - \frac{\mu}{N}\sum_{i=1}^{N}\boldsymbol{x}(n+i)\boldsymbol{x}^\dagger(n+i)\boldsymbol{w}_k^{(n)}] + \boldsymbol{w}_k^{(0)} \quad (15)$$

$$\boldsymbol{w}_k^{(0)} = \boldsymbol{D}(\boldsymbol{D}^\dagger\boldsymbol{D})^{-1}\boldsymbol{g} \quad (16)$$

In this way, computational load can be reduced and possibly better performance can be achieved [9].

We notice that the estimated signal steering vectors provide an initial weighting. The detailed derivation and the convergence of the adaptive procedure can be referred to [2].

4. COMPUTER SIMULATIONS

The performance of the improved CAB algorithm is demonstrated by two computer simulations using a 7-element uniform linear arary with half-wavelength spacing. Simulated data are generated incorporating array calibration errors where calibration phase error are uniformly distributed over $\pm\pi/8$ and gain error are uniformly distributed over $[0.8, 1.2]$. White Gaussian noise at each array element is added.

In the first example, three co-channel users of BPSK signals with identical normalized data rate 0.5, normalized frequency offset 0.2 and roll-off factor 0.5 incident upon the array from DOA of 15°, 0° and −30° with respect to (w.r.t.) the normal of the array. The relative power are 0 dB, 0 dB and 10 dB respectively. Low $SNR = -7$ dB is chosen to illustrate the performance in the presence of weak desired signals. Fig.1(a) shows the beam pattern resulting from Eq.(8) using the steering vectors estimated by Eq.(4) with the signal from 15° being considered as the desired signal. We observe that two deep nulls are placed at −30° and 0° to suppress the interferences, and 0 dB gain at 15° (the response of desired signal is chosen as 1). For comparison, the beam pattern resulting from Eq.(8) using the weighting vectors of CAB, $\boldsymbol{W}_{CAB}$, is plotted in Fig.1(b), it is apparent that the suppression of the signal from −30°, which is far apart from the other two signals, is adequate resulting a deep null at −30° while the other two weighting vectors do not correspond to the steering vectors of the signal from 15° and 0°, and result no null at 0°.

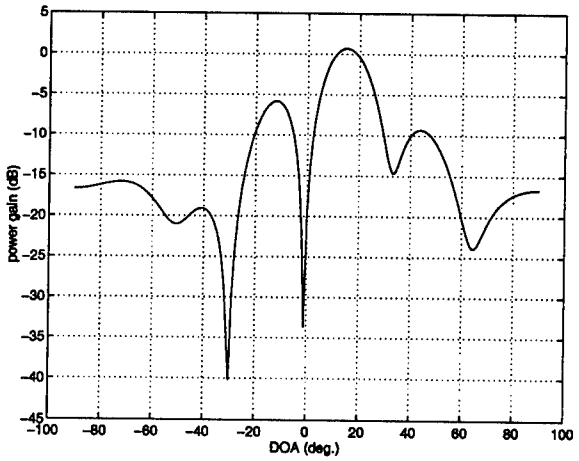


Fig.1(a) Beam pattern w.r.t. the signal from 15° using the estimated steering vectors

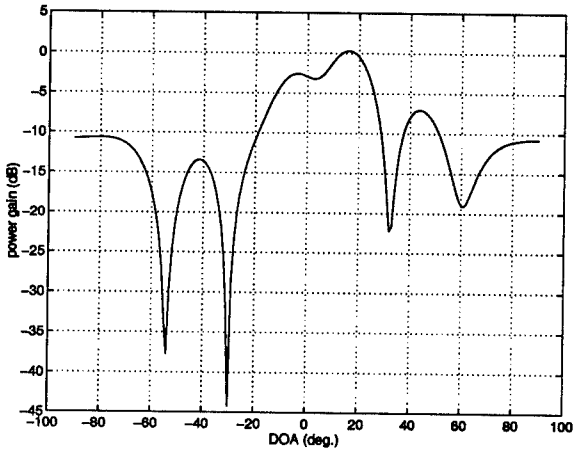


Fig.1(b) Beam pattern w.r.t. the signal from 15° using $\boldsymbol{W}_{CAB}$

In the second example, three co-channel users of BPSK signals incident upon the array from 15°, 0° and −10° w.r.t. the normal of the array with identical normalized data rate 0.5 and roll-off factor 0.5 but different normalized frequency offset 0.2, 0.2 and 0.3 respectively. The signal powers are again 0 dB, 0 dB and 10 dB respectively. The signals from 15° and 0° are considered as the desired signals. $SNR = 0$ dB. In the experiment, the estimated steering vectors of the two desired signals are obtained by iteratively solving Eq.(4). Then the block data adaptation given by Eqs.(15) and (16) with block length being equal to 5 samples is employed. The output SINR are plotted in Fig.2. We observe that the output SINR of the signals converge after 600 samples.

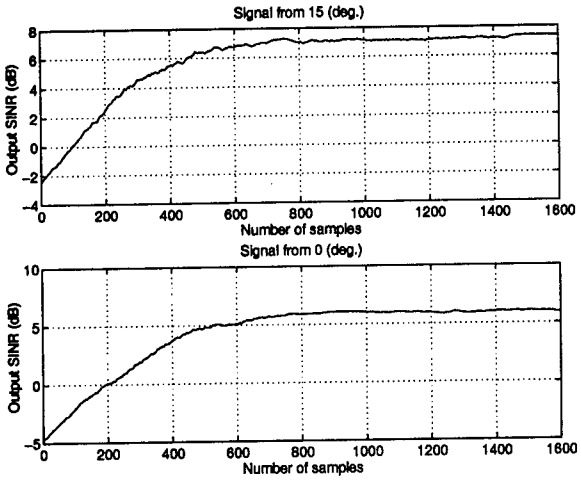


Fig.2 Output SINR of the two desired users

5. CONCLUSION

The improved CAB algorithm provides a solution of user allocation in an agile beam system to achieve more efficient frequency re-use by improving the system performance for multiple users working at the same frequency.

6. REFERENCES

[1] "Final report of study on digital beamforming networks," *ESA Contract Rep.*, Jul., 1990.

[2] O.L.Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp.926-935, Aug., 1972.

[3] B.G.Agee, S.V.Schell and W. A. Gardner, "Spectral self-coherence restoral: a new approach to blind adaptive signal extraction using antenna array," *Proc. IEEE*, vol. 78, No. 4, pp. 753-767, 1990.

[4] Q.Wu and K.M.Wong, "Blind adaptive beamforming for cyclostationary signals," *Proc. ICASSP*, 1994, Australia.

[5] G.Bienvenu and L.Kopp, "Optimality of High Resolution Array Processing Using the Eigensystem Approach," *IEEE Trans. ASSP*, vol. 31, No. 5, pp. 1235-1248, Oct. 1983.

[6] D.A.Wheeler and P.R.Atkins, "Estimation of source bearings at a randomly deformed array by signal space transformation," *Proc. ICASSP*, pp. 3345-3348, 1991, Canada.

[7] C.Y.Tseng, D.D.Feldman and L.J.Griffiths, "Steering vector estimation in uncalibrated arrays," *IEEE Trans. SP*, vol. 43, No. 6, pp. 1397-1412, Jun. 1995.

[8] "Feasibility study on digital beamforming for MSCS", *Contract Rep.*, Com Dev Canada, Dec., 1995.

[9] R.T.Compton,"Adaptive Antennas", *Prentice-Hall Inc.*, New Jersey, 1988.

On the Application of Time-Frequency Distributions in the Excision of Pulse Jamming in Spread Spectrum Communication Systems

Moeness G. Amin *, Alan R. Lindsey †, and Chenshu Wang*

*Department of Electrical and Computer Engineering
Villanova University
Villanova, Pa 19085

†Rome Lab/C3BB
525 Brooks Road
Rome, NY 13441

Abstract

In this paper, time-frequency distributions (TFD) are applied for interference excision in spread spectrum communication systems. The focus is on jammers consisting of pulses of constant envelop frequency modulated interference. The time-support and the instantaneous frequency (IF) information provided by the TFD are used to reduce the jammer effect on the receiver performance. This is achieved by applying an excision notch filter with a null placed at the interference IF. The filter is turned on and off in synchronous with the interference duty cycle. The bit error rates at different frequencies are given and compared with those obtained using the multiresolution analyses.

I. Introduction

Direct Sequence Spread Spectrum (DSSS) systems are widely used in communications in a variety of applications including suppression of a strong interfering signal due to jamming or multipath propagation and providing multiple simultaneous use of the same spectrum. These systems, however, are not jammer proof. In order to increase their jammer resistance, many existing DSSS systems are augmented with other forms of signal processing, which act on improving receiver characteristics and increasing the overall jammer resistance [1,2]. Linear excision filters are often used to mitigate interference. The filter coefficients can be generated using various estimation methods, including block high resolution and adaptive least mean squares techniques. Most of the existing interference excision algorithms, however, assume a stationary environment, or jammers with slowly-varying spectral characteristics. As such, receiver performance becomes unsatisfactory under highly nonstationary conditions and rapidly changing jamming environment. It is therefore desirable to devise excision methods which are based on jammer characteristics in the time-frequency domain, where the nonstationary characteristics of the jammer are revealed and accurate information on its power localization in both time and frequency is provided. In turn, one may be able to remove the nonstationary jammer with minimum distortion of the desired signal.

Two time-frequency based interference excision techniques have been recently proposed for improved receiver performance under nontraditional jammers. In the first approach, interference excision is achieved using time-frequency distributions. This approach was introduced by Amin [3] and detailed in [4,5,6]. In this case, the interference instantaneous frequency, obtained using appropriate time-frequency distributions, is used to form a time-varying linear phase excision filter. This filter has a notch which is in tune with the jammer IF. The second approach is based on multiresolution analysis[7], where the energy localization properties of the wavelet transform are employed to overcome the windowing effects associated with the short-time Fourier transform. For jammer excision, the wavelet transform is applied to the data and the coefficients of highest values, representing the jammer energy, are then removed. From the nature of these two techniques, it is clear that while the time-frequency distribution excision methods are most efficient for constant envelop frequency modulated signals, where the jammer energy is concenterated around its IF, the wavelet transform is primarily effective when the jammer energy is captured in one or few of the transform bins. The later requires the wavelet tiling of the time-frequency plane to be in close match with the jammer characteristics.

In this paper, the performance of the above two techniques under pulse jamming is investigated. The jammer is a train of sinusoidal or chirp pulses with fixed duty cycles. The time-frequency distribution using several kernels including Wigner, Choi-Willimas, the Cone shape, and others offer the means to detect the beginning and the end of each pulse [8]. Additionally, these kernels yield a good estimate of the jammer

0-8186-7576-4/96 $5.00

instantaneous frequency during the pulse period. As such, the excision filter can be designed with an appropriate notch and can be turned on and off according to the duty cycle of the jammer.

In Section 2, a brief review of TFD is presented with discussion on the interference excision systems based on the instantaneous frequency estimate. The wavelet transform, as it is applied to the underlying problem is discussed in Section 3, and Section 4 presents the results of the bit error rate simulations where the TFD-based excision and the wavelet transform excision techniques are compared.

2. TFD Interference Excision Systems

Time-frequency distributions (TFD) are uniquely characterized by a two dimensional function, which is referred to as a "kernel". The t-f kernel can be designed such that the corresponding TFD satisfies several desired properties. For a full discussion of the time-frequency distributions and kernel design methods, the reader is referred to reference [8]. Among the desired t-f properties is the capability to satisfy the instantaneous frequency condition. Generally, this property allows the TFD to encounter peaks at the derivative of the phase of each signal component, irrespective of their time-varying nature.

The time-frequency distribution C_f of the signal f(t) is defined as

$$C_f(t,\omega;\varphi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(t-u,\tau) f(u+\tau/2) f^*(u-\tau/2) e^{-j\omega\tau}\, du\, d\tau \qquad (1)$$

where "t" is the time index and "f" is the frequency index. The t-f kernel $\phi(t,\tau)$ is a function of the time and lag variables. The well known Wigner distribution is a special case of (1) with $\phi(t,\tau)=\delta(t)$. A closer look at equation (1) reveals the simple fact that the TFD is the Fourier transform (FT) of an estimated autocorrelation function. However, contrary to the common way of performing time-averaging, the dependency of $\phi(t,\tau)$ on τ allows the autocorrelation function estimation to be different for different lags.

In addition to the instantaneous frequency, there are other common desired properties which qualify a TFD for proper representations of signals in time and frequency. These properties include the time support and frequency support. Both properties are important for the cases of excision of pulsed and bandlimited jammers, since they, respectively, allow the TFD to be zero (shows no power) at all time instants and frequency bins where the signal is not present. The TFD should also satisfy the marginals properties in which the distribution of signal power over only the time variable or the frequency variable can be separately obtained from the joint TFD.

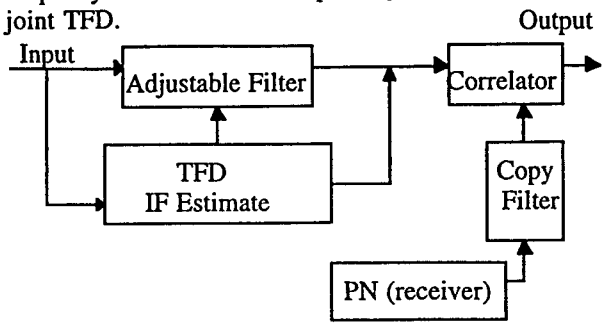


Fig.1 TFD Excision System

The interference excision system based on the TFD is shown in Fig.1. The IF is estimated using t-f kernels with desirable properties. Most importantly, the IF and time-support conditions must be satisfied. The IF is used to define a notch of a three coefficient zero-phase filter. This filter is applied to both the input data and the PN at the receiver. The output of both filters are then correlated and a decision rule is applied . In the simulation section, the Choi-Williams kernel is used for IF estimation.

3. Wavelet Domain Excision

Much research has been accomplished applying wavelet and multirate methods to communications [9.] and in particular, to the interference excision problem [7]. For this study, a standard discrete wavelet transform (DWT) is performed on the received spread spectrum binary phase shift keyed (SS-BPSK) signal (rectangular pulse shaping) and the resulting coefficients, representing the signal in the wavelet basis, are modified via an excision rule that zeroes out the highest 10% of the transform coefficients. The reader will take caution that this is only one of many excision rules available, and is not necessarily optimal for this application. It is an intuitively appealing rule in the sense that the DWT decomposes signals into dyadic subbands which localize narrowband interference. Assuming the jammer to signal energy ratio (JSR) is sufficiently high, this localization causes the coefficients in the frequency bin where the narrowband interference lies to be significantly greater than the rest of the transform coefficients. Except for high frequency interferers, excising the highest ten percent of the coefficicients is sufficient to remove the noise. Unfortunately, a significant portion of data is lost as well.

The communication system simulated in this work consisted of a BPSK signal with a pseudo-noise (PN) spreading code applied at the transmitter, additive white gaussian noise (AWGN,) and constant frequency and frequency modulated continuous wave (CW) jammers with energies. At the receiver, the DWT output was sent to the excision block followed by despreading of the

wideband signal and a correlator. A signum based decision rule provided the data estimates. Spreading codes of 128 and 32 chips per bit were employed 10% jammer duty cycle. Typically 128 bits were simulated at a time and an 11- or 13-level DWT (12 or 14 dyadic frequency slots) was performed on the entire spread sequence. At 128 chips per bit, this means a transform length of $128^2 = 2^{14}$, hence the choice for number of levels in the DWT. The sampling rate for the system was chosen to be 1 sample per chip - effectively limiting the highest frequency jammer to half the chip rate. However, the first-null bandwidth of SS-BPSK is the chip rate, and hence only jammers in the lower half of the spectrum are considered. This is not a crucial issue, since TFD methods are not frequency dependent and so they are unaffected by this limitation, and the tiling of the DWT is such that signals with frequencies in the upper half of the spectrum only worsen the performance.

4. Simulations

Fig.(2-a) compares the bit error rates in the case of pulsed sinusoid with (1/7.1) normalized frequency using 128 chips/bit for the TFD and DWT excision methods. In addition, the BER corresponding to no preprocessing is also shown. For the TFD method, we have included the BERs with exact IF as well as estimates of the IF using equation (1) with 128 and 8 bin FFT. It is perfectly clear that all TFD BER curves are significantly better than the DWT method. The 128-bin FFT outperforms the 8-bin FFT, due to bias caused in the IF estimate using fewer frequency bins. It is noteworthy that exact IF provides no errors up to 80 dB Jammer-to-signal ratio.

Fig.(2-b) shows the same set of curves as Fig.(2-a), except we now use 32 chips/bit. The relative behavior of the two time-frequency excision methods remain approximately the same. Overall, the reduction in the gain leads to an increase in the bit error across the JSR. The experiments conducted for Fig.1 were repeated using higher frequency (1/2.3). The corresponding BER curves are shown in Fig.(2-a,b). The superiority of the TFD methods remain invariant. Fig.3 shows the BER curves for the case of a pulsed chirp jammer, using 128 and 32 chips/bit. The performance of the TFD is slightly deteriorated from the case of fixed sinusoid. Still, the TFD has a remarkable performance which drastically improves over the DWT performance.

5. Conclusions

The interference excision system based on time frequency distributions shown in Fig. 1, outperforms the wavelet transform excision method for constant envelope pulsed interference of either constant or modulated frequency. Using exact IF information yields better results than IF estimates, but this is to be expected. In defense of multiresolution methods, however, it should be noted that the jammer types considered here are not conducive to MRA decompositions, and as a matter of fact, a regular FFT outperforms the wavelet in this scenario, especially for constant frequency jammers. Pulsed interference without IF information (bursts of uncorrelated energy) were not considered in this study, but it is suggested that the TFD methods would not perform as well in this case, and the performance of the wavelet excision scheme would improve.

Acknowledgment:
This work is supported by Rome Lab, contract number F30602-96-0077.

References

1. J. Ketchum and J. Proakis, " Adaptive algorithms for estimating and suppressing narrow band interference in PN spread spectrum systems", IEEE Transactions on Communications, May 1982.

2. R. Iltis, J. Ritcey ,and L. Milstein, "Interference rejection in FFH systems using least squares estimation techniques", IEEE Transact. on Communications, vol. 38, Dec. 1990.

3. M. G. Amin, "Interference excision in spread spectrum communication systems using time-frequency distributions," Technical Report, AFOSR, Rome lab, Sept. 1994.

4. S. Tyler and M. Amin, " Mitigating interference in direct sequence spread spectrum communication systems, " Rome Lab. Technical Journal, 1st Volume, June 1995.

5. M. Amin, G. Venkatesan and S. Tyler, " A new approach for spread spectrum using time-frequency distributions," Proceedings of the SPIE Conference on Advanced algorithms and Architectures for Signal Processing, San Diego, CA , July 1995.

6. S. Roberts and M. Amin, " Linear Vs. Bilinear time-frequency methods for interference mitigation in direct sequence spread spectrum communication systems," Proceedings of the Asilomar Conference Signals, Systems, and Computers, Pacific Grove, CA, November 1995.

7. M. Medley, G. Saulnier, and P. Das, "Applications of the wavelet transform in spread spectrum communications systems", SPIE, Wavelet Applic., Orlando, Florida, April 1994.

8. L. Cohen, Time-Frequency Analysis, Prentice Hall, Englewood Cliffs, New Jersey, 1995.

9. A.R. Lindsey and M.J. Medley, "Wavelet Transforms and Filter Banks in Communications," Proceedings of 1996 SPIE Conference - Wavelet Applications for Dual Use, Orlando, Florida, to appear April 8,1996.

Fig 2(a) BER (128 chips/bit,w0=2*pi/7.1)

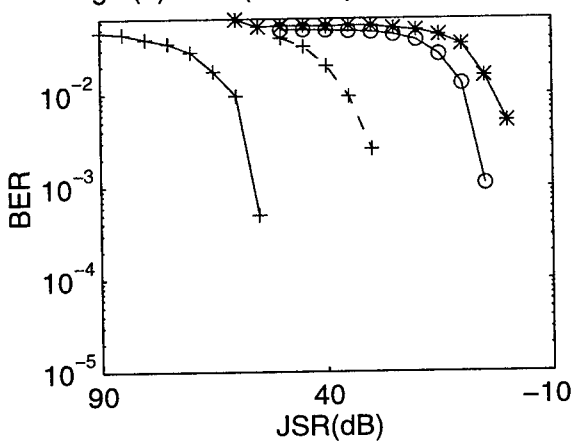


Fig 2(b) BER (32 chips/bit,w0=2*pi/7.1)

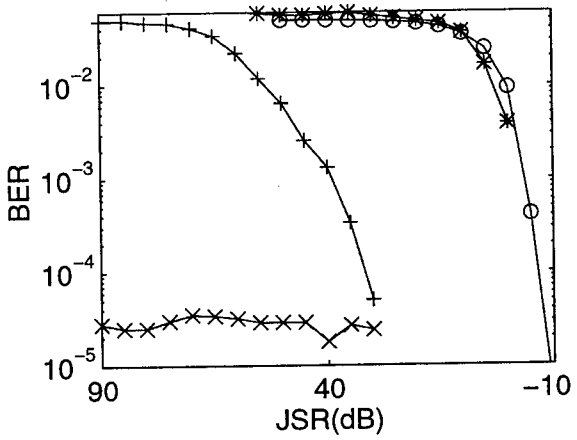


Fig 3(a) BER (128 chips/bit,w0=2*pi/2.3)

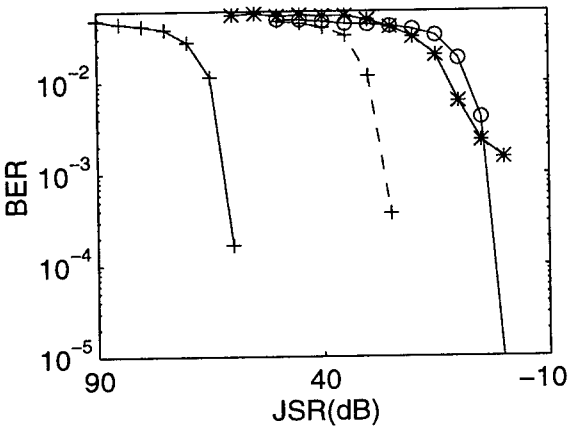


Fig 3(b) BER (32 chips/bit,w0=2*pi/2.3)

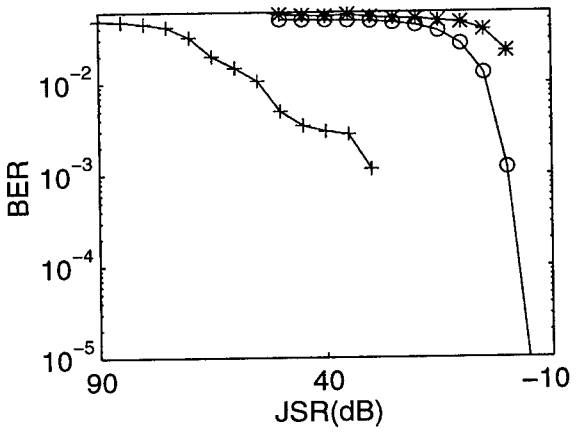


Fig 4(a) BER (128 chips/bit,chirp(0–pi))

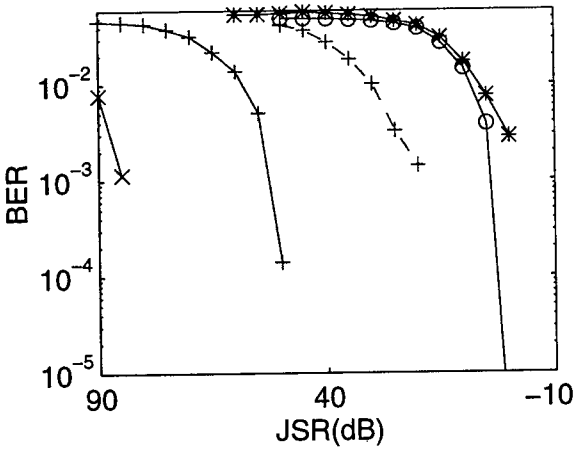
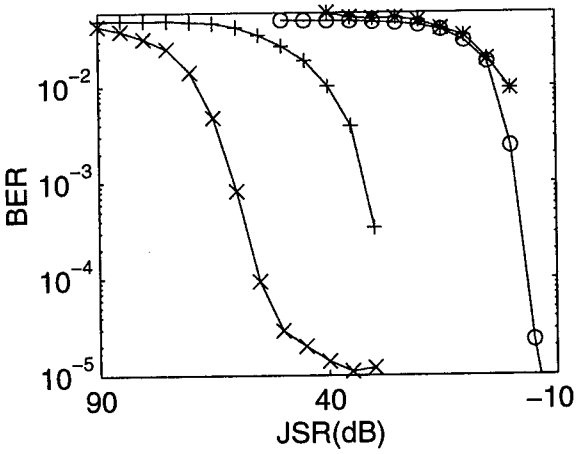


Fig 4(b) BER (32 chips/bit,chirp(0–pi))



In all figures: '*(-)' DWT , '+(--)' 8-point FFT (TFD) , '+(-)' 128-point FFT(TFD), 'x(-)' Exact IF(TFD) , 'o(-)' No Preprocessing

A reduced computation multichannel adaptive equalizer based on HMM

Sylvie Perreau
Alcatel CIT
France
email: perreau@sig.enst.fr

Langford B. White
DSTO
PO Box 1500 Salisbury 5108 Australia.
email:lbw@cd.dsto.defence.gov.au

Pierre Duhamel
Telecom Paris
France
email:duhamel@sig.enst.fr

Abstract

This paper describes a non-linear adaptive equalizer based on a sub-optimal HHM formulation leading to a small computational complexity. A similar approach was already proposed in the monochannel case in [4], and we show here that, in a multichannel context, large improvements are obtainable. It is well known that Maximum Likelihood methods are subject to local minima problems. Although of reduced importance in our previous approach (due to the on-line adaptation), the problem was still present. Since it is now well known that in the multichannel case, the blind equalization problem has a unique minimum, one can hope that the local minima problems can be solved in this context. However, a straightforward formulation of the previous algorithm in the multichannel case does not make it. Hence, we propose a new algorithm allowing Conditional Means estimates of the emitted symbols and blind identification of each impulse response of the channels, involving altogether a maximum likelihood formulation (by means of an approximated EM algorithm) and a criterion making use of the spatial diversity of the multichannel system. Simulations are provided, showing the identification of the impulse responses of the various channels, as well as the symbol estimation performances in terms of Bit Error Rate (BER). The improvements over the single channel case are highlighted.

1. Introduction

We consider here reception through multiple sensors. In this case, the various sensors receive different continuous-time waveforms due to the different physical channels that separate them from the transmitter. However, after sampling at the symbol rate, the corresponding received discrete sequence can be modeled as the output of a Finite Impulse Response (FIR) Filter.

In the recent years, following the work by Tong, Xu, and Kailath [5] many methods have been proposed in order to equalize such systems, relying on the fact that the received signals have a rank-deficient correlation matrix [1]. Initially, these methods were proposed in block versions. This block formalism does not allow a tracking of the channels (when they are time-varying) and has the drawback that the corresponding arithmetic complexity is required by bursts, a fact which either requires a large hardware or introduces a large delay. Very few methods allow an on-line processing of the data as they come. Furthermore, they usually do not explicitly take into account the effect of noise. Finally, a common feature of these methods is that they rely on *structural* properties of the channel, meaning that they do not use any *a priori* knowledge about the input, which is often available at no cost in a communication situation. For instance, it can be useful to take advantage of the fact that the emitted symbols are taken from a discrete finite alphabet.

The algorithm derived in this paper is an adaptive one, providing at each step an estimate of the impulse responses of the multiple channels (thanks to the combination of two criteria which results in a good tradeoff between residual error and sensitivity to initialization), as well as a Conditional Mean (CM) estimate of the symbols currently stored in the channel memory; The Hidden Markov Model (HMM) is used here in a sub-optimal way so that it does not involve the high computational complexity which is the issue of such an approach. Note that the model formulation is usable at a reasonable cost only by the use of the *a priori* knowledge that the emitted sequence belongs to a finite alphabet. Note that all computations provided in this paper are given in real variables. The extension to complex ones is straightforward.

0-8186-7576-4/96 $5.00

2. Problem formulation

Notations are as follows:
- V^T denotes the transposition of vector V
- $\underline{\theta}^{(n)} = [\theta^{(n)}(1)...\theta^{(n)}(L-1)]^T$ denotes the impulse response of the channel from the transmitter to the n^{th} antenna. (L being the channel memory length of all impulse responses).
- $b^{(n)}(t)$ is the additive noise on the n^{th} antenna. We assume these additive noises to be mutually uncorrelated. N being the number of sensors, B_t, denotes the following vector:

$$\underline{B}_t = [b^{(1)}(t)b^{(2)}(t)...b^{(N)}(t)]^T$$

- The transmitted sequence x is independent and identically distributed (iid), and can take M different values $q_k, k = 1...M$ depending on the modulation. X_t is the vector containing all symbols stored in the channel memory at time t:

$$\underline{X}_t = [x(t)x(t-1)...x(t-L+1)]^T$$

Then, the signal received at time t on the n^{th} antenna is given by:

$$y^{(n)}(t) = \sum_{i=0}^{L-1} \theta^{(n)}(i)x(t-i) + b^{(n)}(t) \qquad (1)$$

Thus, we have the stationary model:

$$Y_t = \underline{\Theta}^T X_t + B_t \qquad (2)$$

with matrix Θ and vector $\underline{Y}_t$ defined as:

$$\Theta = [\theta^{(1)}\theta^{(2)}...\theta^{(N)}] \qquad (3)$$
$$Y_t = [y^{(1)}(t).....y^{(N)}(t)]^T \qquad (4)$$

The model described by (2) defines a Hidden Markov Model in which X_t is the state vector of a Markov process described by the following state equation (T being a shift matrix):

$$X_{t+1} = TX_t + x(t+1) * [10...0]^T$$

This hidden Markov process is only reachable through the observation equation (2) which is identical to that corresponding to a transmission through a single channel. However, the hidden process can be reached through N different observations, which is the explanation for the improved performances of the multichannel algorithm in terms of BER.

3 Conditional Mean (CM) estimate of the emitted sequence

Suppose that current estimates of the channels $\hat{\underline{\theta}}_t^{(n)}$, $n = 1, \cdots N$ and of the state vector $\hat{\underline{X}}_{t-1|t-1}$ are available at time t. The state probabilities corresponding to the Forward recursion [2] (ie $Pr(X_t^T = [q_{i_0}....q_{i_{(L-1)}}]|\hat{\Theta}_t, Y_1, .., Y_t)$) would be very computationally demanding, even for moderate length channels since it requires the calculation of M^L probabilities at each step. We use here the approximation derived in [4] which allows the amount of such calculations to be only $M * N$: in this approach, instead of computing the joint probability of all components of vector X_t, we evaluate the probability of each component separately, conditioned by the current prediction of the other ones. This prediction is obtained by taking advantage of the shift structure of the process X. $\hat{X}_{t|t-1}^{(j)}$ being the prediction of the j^{th} component of vector X_t, knowing the observations up to time $t-1$, we have $\hat{X}_{t|t-1}^{(j)} = \hat{X}_{t-1|t-1}^{(j-1)}$. Let $\alpha_{t|t}^{(j)}(k)$ be the probability that the j^{th} symbol in the channel memory be equal to q_k, knowing the observations up to time t, the current estimate of the channels parameters ($\hat{\Theta}_t$), and the prediction of all the other symbols stored in the channel memory. Then, thanks to the so-called forward recursion, we can write: and:

$$\begin{aligned}\alpha_{t|t}^{(j)}(k) &= Pr((X_t^{(j)} = q_k)|Y_1, .., Y_t, \hat{\Theta}_t, \hat{X}_{t|t-1}^{(l)}, l \neq j) \\ &= \alpha_{t-1|t-1}^{(j-1)}(k)\mathcal{N}(Y_t - \hat{\Theta}_t^T \hat{X}_{t|t-1}(j, q_k)) \qquad (5)\end{aligned}$$

Where $\mathcal{N}(.)$ is the N dimensional Gaussian distribution, and $\hat{X}_{t|t-1}(j, q_k)$ is the vector $\hat{X}_{t|t-1}$ where its j^{th} component is replaced by k^{th} possible choice q_k in the alphabet :

$$\hat{X}_{t|t-1}(k) = [\hat{X}_{t|t-1}^{(1)}...q_k...\hat{X}_{t|t-1}^{(L-1)}]^T$$

, then, the estimate of vector X_t is given by:

$$\hat{X}_{t|t} = [\hat{X}_{t|t}^{(0)}....\hat{X}_{t|t}^{(L-1)}]^T$$

Where $\hat{X}_{t|t}^{(j)}$ is the Conditional Mean Estimate of x_{t-j}:

$$\hat{X}_{t|t}^{(j)} = \sum_{k=1}^{M} \alpha_{t|t}^{(j)}(k)q_k$$

The estimate of the emitted sequence being performed, we now focus on the update of the multi-channel impulse response.

4. Estimation of the channel parameters by a combination of two criteria

The parameters are estimated by minimizing a criterion which is a linear combination of a criterion based on the spatial diversity of the system $C_t(\Theta)$ on one side, and of $L_t(\Theta)$, the expected log likelihood, on another side. Both criteria are evaluated with an exponential forgetting factor λ, as we wish tracking slow variations of the channels:

$$Q_t(\Theta) = \sum_{i=1}^{t} \lambda^{t-i}[(1-\alpha)L_i(\hat{\Theta}_i, \Theta) + \alpha C_i(\Theta)] \qquad (6)$$

4.1. Calculation of the expected log-likelihood

The criterion we deal with in this section is the so-called Kullback-Leibler function of the Expectation-Maximisation (EM) algorithm, defined as the expectation of the logarithm of the likelihood function for the complete data calculated at time t(see [3] for the terminology):

$$L_t(\hat{\Theta}_t, \Theta) = E(log\mathcal{N}(Y_t, X_t; \Theta)|Y_1, .., Y_t; \hat{\Theta}_t)$$

$$\begin{aligned} L_t(\hat{\Theta}_t, \Theta) &= \sum_{l=1}^{M^L} \Gamma_{t|t}(l) log(\mathcal{N}(X_t = \xi_l|Y_1, .., Y_t, \Theta) \\ &= \sum_{l=1}^{M^L} \Gamma_{t|t}(l)|Y_t - \Theta^T \xi_l|^2 \end{aligned} \quad (7)$$

where ξ_l is one of the M^L possible realizations of vector X_t and $\Gamma_t(l)$ is the conditional probability of the state: $\Gamma_t(l) = Pr(X_t = \xi_l|Y_1, .., Y_t, \hat{\Theta}_t)$

Because of the approximation developped in the previous section, consisting in computing conditional probabilities instead of joint probabilities , we have to deal with to the so-called "pseudo-likelihood": basically, we approximate each $\Gamma_t(l)$ which is defined as the joint probability on every component of X_t, by the product of the conditional probabilities of each component, given the prediction of the other ones:

$$P(X_t^T = [q_{i_0}, q_{i_1}, ..., q_{i_{L-1}}]|Y_1, .., Y_t, \hat{\Theta}_t) = \prod_{n=0}^{L-1} \alpha'^{(n)}_{t|t}(i_n)$$

The expansion of this calculation (for high SNR levels) leads to the following expression of the expected *pseudo-likelihood* at time t :

$$L_t(\hat{\Theta}_t, \Theta) = \sum_{l=1}^{N} |y_t^{(l)} - \hat{X}_{t|t}^T \theta^{(l)}|^2 \quad (8)$$

4.2 Criterion based on the spatial diversity of the system

We use the criterion described in [5]. Under the assumption that the impulse responses of the various channels have no common zeros, it can be shown [5] that this criterion has a single solution. This property is useful in our case since the EM algorithm is known to have local minima. Moreover, even if the log likelihood involved in this multi-channel case takes explicitly into account the effects of the noise, it does not take full advantage of the spatial diversity of the system. It is expected that a suitable weighting of both criteria can solve the local minima problem, while maintaining the robustness towards noise close to that of the EM algorithm. Only simulations support this claim at that time. Physically, the criterion relies on the fact that the output of each antenna corresponds to the filtering of the SAME input vector by different filters. Consider 2 antennas among the N ones: If we filter one received signal by the impulse response of the other antenna and *vice versa*, both outputs will be filtered by the same coefficients, hence should be equal (up to the disturbances introduced by the noise). This criterion can be written in many different ways. We have chosen :

$$C_t(\Theta) = \sum_{n \neq m} |Y_t^{(m)} \theta^{(n)} - Y_t^{(n)} \theta^{(m)}|^2 \quad (9)$$

where $Y_t^{(n)} = [y^{(n)}(t)...y^{(n)}(t - L + 1)]$ Note that other criteria share the same property and could be used in conjunction with the approximate log-likelihood.

4.3 Maximization step

The maximization over each channel $\theta^{(l)}$ is performed by computing the partial derivatives of $Q_t(\Theta)$ according to $\theta^{(l)}$. Finally, the channels estimates are obtained recursively by:

$$\begin{aligned} \hat{\theta}_{t+1}^{(n)} &= \hat{\theta}_t^{(n)} + (1 - \alpha) R_t^{-1} (y_t^{(n)} - \hat{\theta}_t^{(n)T} \hat{X}_{t|t}) \hat{X}_{t|t} \\ &\quad + \alpha I_t^{(n)-1} \sum_{m \neq n} (Y_t^{(m)} \theta_t^{(n)} - Y_t^{(n)} \theta_t^{(m)}) Y_t^{(m)T} \end{aligned}$$

R_t and $I_t^{(n)}$ are defined as follows:

$$\begin{aligned} R_t &= \lambda R_{t-1} + \hat{X}_t \hat{X}_t^T \\ I_t^{(n)} &= \lambda I_{t-1}^{(n)} + Y_t^{(n)T} Y_t^{(n)} \end{aligned} \quad (10)$$

Their inverses can be computed recursively using the inverse matrix lemma ([4]).

5. Experimental results

5.1. Adaptive behavior

First consider the adaptive behavior of the algorithm, on a BPSK modulation, with $N = 2$, on non-minimum phase channels, $h^{(1)} = [0.15\,0.9\,0.3]$ and $h^{(2)} = [0.3\,0.3\,0.3]$,

Usefulness of the criterion based on diversity Fig.1 shows the evolution of the taps of the first channel $h^{(1)}$, using a straightforward extension of the EM algorithm used in [4] (no spatial diversity explicitly taken into account). It is seen that the algorithm converges to a local minimum of the likelihood corresponding to a minimum-phase channel : $\hat{h^{(1)}} = [0.9\,0.35\,0]$. Fig2 corresponds to $\alpha = 0.3$, and the algorithm converges to the true parameters.

Usefulness of the MAP estimate The usefulness of $L_t(H)$ in the criterion is easily seen by comparing on Fig3 the MSE on the parameters estimates for $\alpha = 1$ (C_t only) and $\alpha = 0.3$. As the likelihood takes more efficiently the noise effects into account, the parameters produced by its minimization are more accurate than the one obtained when minimizing $C_t(H)$ only. Both simulations were initialized to the same values $h^{(1)} = h^{(2)} = [1\,0\,0]$, and performed with a $10db$ SNR at the output of the channels.

5.2. BER results

Fig4 compares the BER obtained by using two channels ($h^{(1)}$ and $h^{(2)}$ altogether) to that obtained using a single channel ($N = 1$ on either $h^{(1)}$ or $h^{(2)}$). The improvement is significant, while the computational cost involved for the multi-channel case is still linear with the channel memory.

6. Conclusion

This paper proposes a new algorithm, which couples two different and complementary blind equalization methods.The first one based on ML identification of the channels taps, and detection of the symbols thanks to a HMM formulation, brings robustness towards noise. The other one is based on a criterion involving the spatial diversity of the system and tends to constrain the solution to be unique. The proposed algorithm is shown to take advantage of the complementarity of both criteria, especially avoiding the problem of the local minima of the likelihood, while providing accurate results in case of poor SNR. Moreover, the fact that the algorithm is adaptive allows a real time computation without the high computational complexity of the HMM-based classical methods. This has been obtained by the use of a suboptimal HMM formulation which has nevertheless a good efficiency.

References

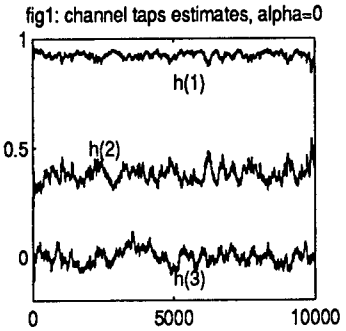
[1] E. Moulines, P. Duhamel, JF. Cardoso, S. Mayrargue *Subspace methods for the blind identification of multichannel FIR filters* IEEE Trans SP, Feb 1995, pp. 516-525

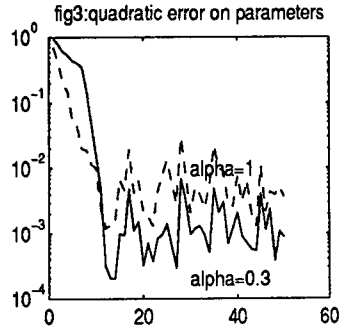
[2] V. Krishnamurthy and J. B. Moore *On-Line Estimation of hidden Markov Model Parameters Based on the Kullback-Leibler Information Measure* IEEE Trans. SP Vol.41 N8 August 1993,pp 1557-1572

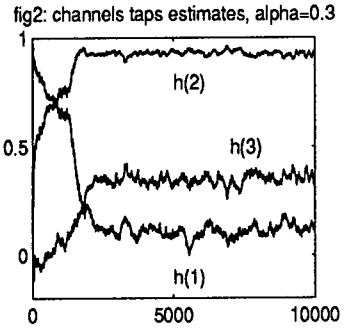
[3] A.P. Dempster, N.M. Laird and D.B Rubin *Maximum Likelihood from incomplete data via the EM algorithm* J.Roy.Soc Vol.6 1977 pp 1-38

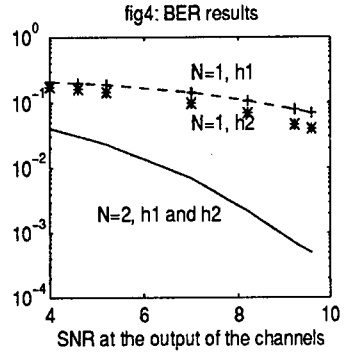
[4] L.B. White, S. Perreau, P. Duhamel *Reduced Computation Blind Equalization for FIR channel input Markov Models* proceedings of ICC, June 1995, Seattle, USA

[5] H. Liu, G. Xu, L. Tong *A deterministic approach to blind identification of multi-channel FIR systems* Proceedings of ICASSP 1994,pp IV.581-584.









BLIND CHANNEL EQUALIZATION VIA MULTIOBJECTIVE OPTIMIZATION

Shiyu Zeng, Hanks H. Zeng and Lang Tong

Dept. of Electrical and Systems Engineering
University of Connecticut
Storrs, U-157, CT 06269-3157

Abstract

We present in this paper an multiple objective optimization approach to fast blind channel equalization. By investigating first the performance (mean-square error) of the standard fractionally spaced CMA equalizer in the presence of noise, we show that CMA local minima exist near the minimum mean-square error (MMSE) equalizers. Consequently, CMA may converge to a local minimum corresponding to a poorly designed MMSE receiver with considerablely large mean-square error. Based on the multiple objective optimization techniques, we propose next a blind channel estimator by exploiting simultaneously the second-order cyclostationary statistics and the constant modulus of QAM-type communication signals. Such a channel estimation-based blind equalization scheme has the advantage of designing FIR minimum mean-square error equalizer with the optimal delay.

1. INTRODUCTION

Blind equalization has the potential to improve the efficiency of communication systems by eliminating training signals. Difficulties of its application in wireless communications, however, are due largely to the characteristics of the propagation media – multipath delays and fast fading. The challenge is achieving blind equalization using only a limited amount of data.

A widely tested algorithm is the constant modulus algorithm (CMA) [5, 10]. In the absence of noise, under the condition of the channel invertibility, the CMA converges globally for symbol-rate IIR equalizers and fractionally spaced FIR equalizers [4, 6]. It is shown in [3] that CMA is less affected by the ill-conditioning of the channel. However, Ding *et. al.* [2] showed that CMA may converge to some local minimum for the symbol-rate FIR equalizer. In the presence of noise, the analysis of convergence of CMA is difficult and little conclusive results are available. Another drawback of CMA is that its convergence rate may not be sufficient for fast fading channels.

Another approach to the blind equalization is based on the blind channel estimation. Some of the recent eigenstructure-based channel estimations (see *e.g.* [7, 8]) require a relatively smaller data size comparing with higher-order statistical methods. However the asymptotic performance of these eigenstructure-based schemes is limited by the condition of the channel [12, 13]. Specifically, the asymptotic normalized mean-square error (ANMSE) is lower bounded by the condition number of the channel matrix. Unfortunately, frequency selective fading channels with long multipath delays often result in ill-conditioned channel matrices.

The key idea of this paper is to combine the approach based on minimizing the constant modulus cost and that based on matching the second-order cyclostationary statistics. The main feature of the proposed approach is the improved convergence property over the standard CMA equalization and the improved robustness for ill-conditioned channels.

2. THE MODEL

Fractionally sampled channel and its equalizer can be represented by the cascade of a single-input multiple-output (SIMO) channel and a multiple-input single-output (MISO) equalizer. The system equations are given by

$$x_k^{(i)} = \sum_{j=0}^{L_h-1} h_j^{(i)} s_{k-j} + w_k^{(i)} \quad , i = 1, \cdots, M \quad (1)$$

$$y_k = \sum_{i=1}^{M} \sum_{j=0}^{L_f-1} f_j^{(i)} x_{k-j}^{(i)} \quad (2)$$

where $h^{(i)}$, $f^{(i)}$ are the ith (sub)channel and its equalizer with length L_h, L_f respectively, s, w, x, y are transmitted symbol, additive noise, received data and equalizer

This work was supported in part by the National Science Foundation under Contract NCR-9321813 and by the Advanced Research Projects Agency monitored by the Federal Bureau of Investigation under Contract No. J-FBI-94-221.

output respectively. In matrix form, we have

$$\mathbf{x}_k = \mathbf{H}\mathbf{s}_k + \mathbf{w}_k \quad (3)$$

$$y_k = \mathbf{f}^H\mathbf{x}_k = \mathbf{q}^H\mathbf{s}_k + \mathbf{f}^H\mathbf{w}_k \quad (4)$$

$$\mathbf{q} \triangleq \mathbf{H}^H\mathbf{f}, \quad (5)$$

where $(\cdot)^H$ denotes Hermitian, $\mathbf{H}$ is channel matrix and $\mathbf{q}$ is the combined channel. We shall make the following assumptions:

A1: The input sequence $\{s_k\}$ is zero-mean and

$$E\{s_k s_l^*\} = \delta(k-l).$$

s_k has the constant modulus property (CM) $|s_k| = 1$.

A2: Noise $w_k^{(i)}$ is zero-mean, white Gaussian with variance σ^2.

3. PROPERTIES OF CMA EQUALIZERS

The analysis in this section is restricted for real parameters and equiprobable binary source. Generalizations of the complex case are readily available. The CMA minimizes,

$$J_c(\mathbf{f}) \triangleq E\{(y_k^2-1)^2\} \quad (6)$$

$$= 3||\mathbf{f}||_{\mathbf{R}}^4 - 2||\mathbf{f}||_{\mathbf{R}}^2 - 2||\mathbf{q}||_4^4 + 1, \quad (7)$$

where $||\mathbf{f}||_{\mathbf{R}}$ is 2-norm defined by $\sqrt{\mathbf{f}^t\mathbf{R}\mathbf{f}}$, $||\mathbf{q}||_4$ is 4-norm defined by $(\sum q_i^4)^{\frac{1}{4}}$, and

$$\mathbf{R} \triangleq E\{\mathbf{x}_k\mathbf{x}_k^t\} = \mathbf{H}\mathbf{H}^t + \sigma^2\mathbf{I}. \quad (8)$$

In [14], it has been shown that CMA equalizers must be in the signal subspace spanned by the columns of $\mathbf{H}$. Therefore the analysis of CMA can be carried out in the combined channel $\mathbf{q}$ defined in (5). The equivalent CMA cost function is then given by

$$J(\mathbf{q}) \triangleq J_c((\mathbf{H}^t)^\dagger\mathbf{q}) = 3||\mathbf{q}||_{\mathbf{\Phi}}^4 - 2||\mathbf{q}||_{\mathbf{\Phi}}^2 - 2||\mathbf{q}||_4^4 + 1, \quad (9)$$

where $\mathbf{\Phi} \triangleq \mathbf{H}^\dagger\mathbf{R}(\mathbf{H}^t)^\dagger$. In the absence of channel noise, it has been shown that the CMA using fractionally-spaced equalizers converges globally [6] to one of zero forcing equalizers, *i.e.*, $\mathbf{q}_c = \mathbf{e}_\nu, \forall 1 \le \nu \le n_q$ where $\mathbf{e}_\nu$ a unit column vector with 1 at the νth entry and zero elsewhere. In the presence of noise, some minima may become local minima. In this section, we study the locations of these CMA equalizers. Specifically, we will study the neighborhoods of MMSE equalizers which minimize

$$J_m^\nu(\mathbf{f}) \triangleq E\{(y - s_{k-\nu+1})^2\}, \quad (10)$$

where ν is the delay of the equalizer. Note that the CMA does not have control of the delay ν due to the nature of the blind equalization.

There are several reasons to choose this type of regions. Since the MMSE equalizer is the optimal linear equalizer, any equalizer which is far away from it has a large MSE. Therefore, if there exist CMA local minima in these regions, one of the minima must be the optimum CMA equalizer which has the minimum MSE. The other reason is the strong relationship between the MMSE equalizer and the CMA equalizer. This can be seen when the noise approaches to zero.

Without loss of generality, let's consider $\mathbf{q}$ in the neighborhood of MMSE equalizer $\mathbf{q}_m$ at delay $\nu = 1$. $\mathbf{q}$ and $\mathbf{\Phi}$ can be partioned into

$$\mathbf{q} \triangleq \theta\begin{pmatrix}1\\ \mathbf{q}_{\mathrm{I}}\end{pmatrix}, \qquad \mathbf{\Phi} \triangleq \begin{pmatrix}a & \mathbf{b}^t\\ \mathbf{b} & \mathbf{C}\end{pmatrix}, \quad (11)$$

where $\mathbf{q}_{\mathrm{I}}$ is the intersymbol interference part, θ represents the signal energy, and $1-\theta$ is the bias between the ZF equalizer and $\mathbf{q}$.

In order to locate the CMA equalizer (the minimum point), we need following definitions, given the MMSE equalizer $\mathbf{q}_m^t = \theta_m[1, \mathbf{q}_{m\mathrm{I}}^t]$,

$$\delta \triangleq ||\mathbf{q}_{\mathrm{I}} - \mathbf{q}_{m\mathrm{I}}||_{\mathbf{C}} \quad (12)$$

$$\theta_r^2 \triangleq \frac{\theta_m}{3 - 2\theta_m^2 - 2\theta_m^2||\mathbf{q}_{m\mathrm{I}}||_4^4} \quad (13)$$

$$c_0 \triangleq \frac{1}{3 - 2\theta_m^2 - 2\theta_m^2||\mathbf{q}_{m\mathrm{I}}||_4^4} \quad (14)$$

$$c_1(\delta) \triangleq -2(\delta^2 + \frac{1}{\theta_m}) \quad (15)$$

$$c_2(\delta) \triangleq 3(\delta^2 + \frac{1}{\theta_m})^2 - 2(1 + (\delta + r_m)^4) \quad (16)$$

$$D(\delta) \triangleq c_1(\delta)^2 - 4c_2(\delta)c_0. \quad (17)$$

The following theorem gives a sufficient condition of the existence local minimum, its location and also gives the size of the region.

Theorem 1 *Under the condition that $J_m^\nu(\mathbf{f}_m) < \frac{1}{3}$, if $D(||\mathbf{q}_{m\mathrm{I}}||_2) < 0$, then there exists a local minimum in*

$$\mathcal{B} \triangleq \{0 \le \delta \le \delta_U^*, \theta_L^* \le \theta \le \theta_U^*\}, \quad (18)$$

where

$$\delta_U^* = \inf_{\delta>0, D(\delta)<0}\{\delta\} \quad (19)$$

$$\theta_L^* = \min_{0\le\delta\le\delta_U^*}\sqrt{\frac{-c_1(\delta) - \sqrt{c_1(\delta)^2 - 4c_2(\delta)c_0}}{2c_2(\delta)}} \quad (20)$$

$$\theta_U^* = \max_{0\le\delta\le\delta_U^*}\sqrt{\frac{-c_1(\delta) + \sqrt{c_1(\delta)^2 - 4c_2(\delta)c_0}}{2c_2(\delta)}} \quad (21)$$

This theorem provides an expression $D(||\mathbf{q}_{m\mathrm{I}}||_2)$ to determine the region of cylinder $\mathcal{B}$ which includes the

CMA equalizer. The procedure only needs the parameters of the MMSE equalizer.

Perhaps the most interesting concern is the MSE of the CMA equalizer. With the result of Theorem 1, we are ready to give the answer.

Theorem 2 *(a) The MSE of the CMA equalizer in $\mathcal{B}$ is bounded by*

$$\underbrace{\frac{(\theta_U-\theta_m)^2}{\theta_m}}_{\Delta\mathcal{E}_L} \le \Delta\mathcal{E} \le \underbrace{\frac{(\theta_L-\theta_m)^2}{\theta_m}+\theta_U^2\delta_U^2}_{\Delta\mathcal{E}_U}, \qquad (22)$$

where $\Delta\mathcal{E}$ is the extra MSE, i.e., $\Delta\mathcal{E} \triangleq J_m^\nu(\mathbf{q}_c) - J_m^\nu(\mathbf{q}_m)$. (b) Let $\Delta\mathcal{E}_U^\nu$ be the upper bound of CMA equalizer associated with delay ν, then the MSE of the optimum CMA equalizer is upper bounded by

$$\mathcal{E}_{c*} = \min_\nu\{\mathcal{E}_m^\nu + \Delta\mathcal{E}_U^\nu\}. \qquad (23)$$

(c) The MSE of CMA is approximated by

$$\widehat{\Delta\mathcal{E}} = \frac{(\theta_r-\theta_m)^2}{\theta_m} = 4\mathcal{E}_m^2 + O(\mathcal{E}_m^2). \qquad (24)$$

The consequence of these theorems is twofolds. (i) The CMA equalizers are very close to the MMSE equalizers; (ii) There may exist a CMA local minimum in the neighbourhood of a MMSE equalizer which has significantly large MSE.

4. THE MULTIPLE OBJECTIVE OPTIMIZATION APPROACH

To avoid the undesirable local minimum of CMA, one can use the channel estimation based equalization approach. Once the channel is estimated, an MMSE equalizer can be constructed by selecting the optimal ν in (10). Furthermore, this approach provides the flexibility to design other types of receivers, such as decision-feed back equalizer, or maximum likelihood sequence estimator.

Considered in this paper are the costs associated with the constant modulus property $J_{CM}(\mathbf{h})$, the second-order statistics $J_{CF}(\mathbf{h})$, and the observed data $J_Q(\mathbf{h})$:

$$J_{CM}(\mathbf{h}) = \sum_k (|y_k|^2-1)^2 \qquad (25)$$

$$J_{CF}(\mathbf{h}) = \sum_{i,j,m} |r_{ij}(m) - \hat{r}_{ij}(m)|^2, \qquad (26)$$

$$J_Q(\mathbf{h}) = \mathbf{h}^H\mathbf{Q}\mathbf{h}. \qquad (27)$$

Note that the optimization of $J_Q(\mathbf{h})$ leads to, among a number of eigenstructure-based algorithms, the least-squares [7] or the subspace channel estimators [8]. Matrix $\mathbf{Q}$ in $J_Q(\mathbf{h})$ can be obtained from the data directly.

Both $J_{CF}(\mathbf{h})$ and $J_Q(\mathbf{h})$ involve the second-order statistics (in different ways) whereas $J_{CM}(\mathbf{h})$ involves the higher-order statistics. We present next the weighting and the constrained approaches, the two frequently used techniques in multiobjective optimization, to the optimization of the above cost functions.

4.1. The CM-CF Algorithm

The CM-CF algorithm is derived from the weighted optimization of the constant modulus cost $J_{CM}(\mathbf{h})$ and the correlation fitting cost $J_{CF}(\mathbf{h})$:

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}\in\mathcal{H}} \underbrace{\alpha J_{CM}(\mathbf{h}) + \beta J_{CF}(\mathbf{h})}_{J(\mathbf{h})}, \qquad (28)$$

where α, β are weights of the two cost functions respectively. $\mathcal{H}$ is the subspace contains the channel vector. In practice, $\mathcal{H}$ may be constructed from the principal component structure of the fading channel [11].

The difficulty of this optimization is that the explicit form of the constant modulus cost $J_{CM}(\mathbf{h})$ as a function of the channel is unknown. Fortunately, from the analysis in Section 3, the constant modulus equalizer can be approximated by the MMSE equalizer which can be obtained once the channel is estimated. A gradient search is used to minimize $J(\mathbf{h})$,

$$\hat{\mathbf{h}}_{n+1} = \hat{\mathbf{h}}_n - \mu\nabla_{\mathbf{h}} J(\hat{\mathbf{h}}_n), \qquad (29)$$

where μ is a step size.

4.2. CMA with Subspace Constraints

In the constrained approach, we consider the following optimization

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} J_{CM}(\mathbf{h}) \ \textit{subject to}\ J_Q(\mathbf{h}) \le \alpha||\mathbf{h}||^2. \qquad (30)$$

When $\mathbf{Q}$ is constructed from the true covariance matrix $\mathbf{R}_{xx}$, the "true" channel is in the null space of $\mathbf{Q}$ and the channel identification becomes one of finding the eigenvector associated the zero eigenvalue. When the estimated covariance matrix is used and the channel is close to be unidentifiable [9], the null space is no longer easy to determine. It is therefore reasonable to extend the subspace to include additional dimensions. Mathematically, we may view this approach as restricting the channel vector in a subspace that the quadratic cost $J_Q(\mathbf{h})$ is constrained by an upper bound. Let $\mathcal{V}$ be the *linear subspace* in which

$$J_Q(\mathbf{h}) \le \alpha||\mathbf{h}||_2^2 \qquad (31)$$

for some pre-specified α. As a suboptimal approach to (30), the channel estimator is then obtained from the following constrained optimization

$$\min_{\mathbf{h}\in\mathcal{H}\bigcap\mathcal{V}} J_{CM}(\mathbf{h}). \qquad (32)$$

The above optimization can then be transformed into an unconstrained optimization. It can be shown that $\mathcal{C} = \mathcal{H} \bigcap \mathcal{V}$ can be obtained from the span of the eigenvectors associated with the smallest several eigenvalues of matrix $\hat{\mathbf{\Phi}} = \mathbf{B}^H \hat{\mathbf{Q}} \mathbf{B}$. A gradient-type optimization is used similar to (29).

5. SIMULATIONS

The class of two-ray multipath fading channels with independently faded components is used in the simulation. The channel impulse response is given by

$$h(t) = \sum_{i=1}^{2} a_i p(t - \tau_i), \tag{33}$$

where $\{a_i\}$ are independent zero-mean complex Gaussian variables; p(t) is the raised-cosine waveform with roll-off factor 0.25 and the length of 6 symbol intervals. Uniformly distributed in [0, 2T] (T is a symbol interval), the delays $\{\tau_i\}$ are statistically independent. The signal is sampled at twice of the symbol rate.

We compared the mean-square error of the equalized channel using (i) the CM-CF approach; (ii) the CMA with Subspace Constraints; (iii) the Least-Squares CMA (LSCMA) [1]; (iv) the MMSE equalizer constructed from the subspace channel estimator. The cumulative percentage of the channel estimates for a fixed MSE is computed and shown in Fig. 1. When compared with the LSCMA algorithm (the dashdot line), the proposed algorithms (the solid and dashed lines) has considerable improvement for the small MSE, such as the MSE less than 0.02, and improvement is reduced as MSE increases.

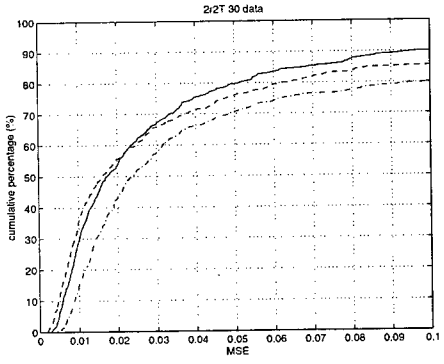


Figure 1: Performance comparison.

REFERENCES

[1] B.G. Agee. "The least-squares CMA: a new technique for rapid correction of constant modulus signals". In *Proc. ICASSP 86 Conf.*, Tokyo, 1986.

[2] Z. Ding, R. A. Kennedy, B. D.O. Anderson, and C.R. Johnson. "Ill-convergence of Godard blind equalizers in data communication systems". *IEEE Trans. Communications*, pages 1313–1327, September 1991.

[3] I. Fijalkow, J.R. Treichler, and C.R.Johnson Jr. "Fractionally Spaced Blind Equalization: Loss of Channel Disparity". In *Proc. IEEE Intl. Conf. Acoust. Speech, Sig. Proc.*, volume 3, pages 1988 – 1991, Detroit, MI, May 1995.

[4] G.J. Foschini. "Equalizing without altering or detecting data". *Bell Syst. Tech. J.*, 64:1885–1911, October 1985.

[5] D.N. Godard. "Self-recovering equalization and carrier tracking in two-dimensional data communication systems". *IEEE Trans. Commun.*, COM-28(11):1867–1875, November 1980.

[6] Y. Li and Z. Ding. "Global convergence of fractionally spaced Godard adaptive equalizers". In *Proc. 28th Asilomar Conference on Signal, Syst., & Comp.*, Asilomar, CA, Mar. 1994.

[7] H. Liu, G. Xu, and L. Tong. "A Deterministic Approach to Blind Identification of Multichannel FIR Systems". In *Proc. 27th Asilomar Conference on Signal, Syst., & Comp.*, Asilomar, CA, Oct. 1993.

[8] E. Moulines, P. Duhamel, J.F. Cardoso, and S. Mayrargue. "Subspace-Methods for the Blind Identification of Multichannel FIR Filters". *IEEE Trans. SP*, SP-43(2):516–525, Feb. 1995.

[9] L. Tong, G. Xu, B. Hassibi, and T. Kailath. "Blind identification and equalization of multipath channels: A frequency domain approach". *IEEE Trans. Information Theory*, 41(1):329–334, January 1995.

[10] J.R. Treichler and B.G. Agee. "A new approach to multipath correction of constant modulus signals". *IEEE Trans. ASSP.*, ASSP-31(2):459–472, April 1983.

[11] H. Zeng and L. Tong. "Blind Channel Estimation Using the Second-Order Statistics: Algorithms". *submitted to IEEE Trans. on Signal Processing.*

[12] H. Zeng and L. Tong. "Blind Channel Estimation Using the Second-Order Statistics: Asymptotic Performance Analysis". *submitted to IEEE Trans. on Signal Processing.*

[13] H. Zeng and L. Tong. "Some New Results on Blind Channel Estimation: Performance and Algorithms". In *Proc. 27th Conf. Information Sciences and Systems*, Baltimore, MD, March 1995.

[14] H. Zeng and L. Tong. "Relationships Between the CMA and Minimum Mean Squares Error Receivers". In *Proc. 28th Conf. Information Sciences and Systems*, Princeton, NJ., March 1996.

On The Inclusion of Channel's Time Dependence in a Hidden Markov Model for Blind Channel Estimation

Carles Antón-Haro, José A. R. Fonollosa and Javier R. Fonollosa.
Dpt. of Signal Theory and Communications. Universitat Politècnica de Catalunya.
c/ Gran Capità s/n. 08034 Barcelona (SPAIN)
Tel: +34-3-4016454, Fax: +34-3-4016447, e-mail: carles@gps.tsc.upc.es

Abstract

In this paper we introduce several modifications to the Baum&Welch (BW) formulas used to reestimate the parameters of a Hidden Markov Model (HMM). The estimated parameter is the channel impulse response (CIR) of a communication system which is known to be time-varying. With these modifications, channel tracking properties of a BW-based algorithm are improved. The resulting algorithm is tested in a specific mobile radio environment (the GSM system), exhibiting good performance at expenses of higher computational complexity.

1. Introduction

It's well known that no high-speed band-limited digital communication can be carried out without the help of an equalizer. Conventional approaches to the adjustment of this equalizer require the transmission of a *training sequence* (i.e. known *a priori* by the receiver and the transmitter), which provides an accurate initial estimate for the equalizer taps; afterwards, slighter adjustments can be made on-line to adapt this first estimate to the, almost always, changing environment. Of course, the transmission of these training sequences, when possible, brings down the capacity of the system. For that reason, there is an increasing interest around blind equalizers [1,2,3] which deal with the problem of the adjustment *without* training sequences (i.e. blindly).

In [3], an Estimation-Modification (EM) Viterbi-based algorithm is proposed to perform jointly a Maximum Likelihood (ML) channel estimation and sequence detection. However, modelling the received signal as a HMM allows us to make use of the complete theory developed for these models. For example, the Baum&Welch (BW) algorithm was proposed in [7] to estimate the parameters of the channel and the characteristics of the modulation. This algorithm is known to lead, at least, to a local maximum of the likelihood function [4], what is not guaranteed by the Viterbi algorithm (VA). In this paper, several modifications to this previously proposed algorithm are introduced to cope with the special features of mobile radio channels.

2. Signal model

As mentioned before, the environment in which the new algorithm is tested is the Paneuropean Mobile Radio System, also known as GSM. In this system, a constant- envelope Gaussian Minimum Shift Keying (GMSK) modulation scheme with equivalent bandwidth (BT) equal to 0.3 is used. The access strategy is TDMA with 8 timeslots per carrier and 156.25 bit-intervals per timeslot in Normal bursts. At the chosen bit rate (270.8 kb/s), multipath propagation leads to deep fades and to uncontrolled Intersymbol Interference (ISI). Besides, and due to the mobile nature of the receiver, Doppler effect is also observed.

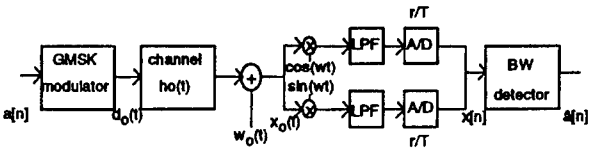


Fig. 1: Transmission subsystem.

Taking into account the above mentioned features, the signal at the input of the BW detector can be modelled as:

$$x[n] = \mathrm{f}(\mathrm{s}[n]) + w[n] \qquad (1)$$

where f(.) is a non-linear function of the present state $\mathbf{s}[n]$, and $\{\mathrm{w}[n]\}$ denotes a sequence of zero-mean Gaussian variables with variance σ^2 (AWGN). If we go on developing an expression for f(.) we get:

$$\mathrm{f}(\mathrm{s}[n]) = \sum_{i=0}^{l_c-1} h[i]\,d[n-i] = \sum_{i=0}^{l_c-1} h[i]\,e^{j\phi[n-i]} \qquad (2)$$

where $\mathbf{h}$ and $\mathbf{d}$ are the baseband equivalences for $\mathbf{h_0}$ and $\mathbf{d_0}$. For a modulation index of 0.5, $\phi[n]$ can be expressed as:

0-8186-7576-4/96 $5.00

$$e_i^2 = \frac{\sum_{n=1}^{L} \gamma_i[n] \left| x[n] - m_i^{(0} - m_i^{(1} \cdot n \right|^2}{\sum_{n=1}^{L} \gamma_i[n]}, \quad 1 \le i \le N \quad (13)$$

Differentiating each component in the sum $\varepsilon^2=\Sigma e_i^2$ with respect to $m_i^{(0}$ and $m_i^{(1}$ we obtain:

$$\nabla e_i^2 = \begin{pmatrix} \frac{\partial e_i^2}{\partial m_i^{(0}} \\ \frac{\partial e_i^2}{\partial m_i^{(1}} \end{pmatrix} = -2 \cdot \begin{pmatrix} \Re\left\{ \sum_{n=1}^{L} \gamma_i[n] \left(x[n] - m_i^{(0} - m_i^{(1} \cdot n \right) \right\} \\ \Re\left\{ \sum_{n=1}^{L} \gamma_i[n] \left(x[n] - m_i^{(0} - m_i^{(1} \cdot n \right) \cdot n \right\} \end{pmatrix} \quad (14)$$

whose Hessian is positive definite unless

$$c1) \qquad \sum_{n=1}^{L} \gamma_i[n] = 0 \quad (15)$$

(i.e. that state was not observed along the timeslot), or

$$c2) \quad \sum_{n=1}^{L} \gamma_i[n] = \gamma_i[n_i] \quad (16)$$

(i.e.that state was observed only once, when $n=n_i$). In those cases, of course, there is no sense in looking for a linear approximation. From equating the gradient to zero and carrying out proper transformations, we find that

$$m_i^{(0} = \frac{A \cdot \left(\sum_{n=1}^{L} \gamma_i[n] x[n] \cdot n \right) - B \cdot \left(\sum_{n=1}^{L} \gamma_i[n] x[n] \right)}{\Delta} \quad (17)$$

$$m_i^{(1} = \frac{C \cdot \left(\sum_{n=1}^{L} \gamma_i[n] x[n] \right) - B \cdot \left(\sum_{n=1}^{L} \gamma_i[n] x[n] \cdot n \right)}{\Delta} \quad (18)$$

$$\hat{\sigma}^2 = \frac{1}{L} \sum_{n=1}^{L} \sum_{i=1}^{N} \gamma_i[n] \left| x[n] - m_i^{(0} - m_i^{(1} \cdot n \right|^2 \quad (19)$$

where

$$A = \sum_{n=1}^{L} \gamma_i[n]; \quad B = \sum_{n=1}^{L} \gamma_i[n] \cdot n; \quad C = \sum_{n=1}^{L} \gamma_i[n] \cdot n^2 \quad (20)$$

$$\Delta = A \bullet C - B^2$$

provide the components of the desired vector and an estimate for the variance of the AWGN.

Finally, special measures should be taken for the cases above mentioned in which $m_i^{(0}$ and $m_i^{(1}$ remain undefined. In the first case, *c1)*, those components of the means vector are not considered in order to obtain **h**. The method adopted is blocking them with a (diagonal) Weighting Matrix to be included in the LS estimate of step 3. The elements of such matrix are a measure of the reliability in the estimation of every component of vector **m**, as a function of the times this state was observed along the sequence. To be precise:

$$\mathbf{W} = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & w_N \end{pmatrix} \qquad w_i = \sum_{n=1}^{L} \gamma_i[n] \quad (21)$$

In the second case, *c2)*, the *static* estimate for **m** replaces the linear approximation. That is:

$$m_i^{(0} = \frac{\sum_{n=1}^{L} \gamma_i[n] x[n]}{\sum_{n=1}^{L} \gamma_i[n]} = x[n_i] \quad (22)$$

$$m_i^{(1} = 0$$

The resulting algorithm will be referred from now on as the Time-Dependent BW (TDBW) algorithm.

5. Simulation results

We tested the performance of the algorithm for the channels described in the ETSI recommendations. The speed for the mobiles in each environment was chosen according to [8]. Among all the cases, the most interesting ones were RA250 and RA100 (Rural Area Environment; speed equal to 100 and 250 km/h), since channel coherence time-intervals are the lowest ones. It should also be remarked that a sampling rate of 2 samples/symbol was considered to compensate for possible timing errors.

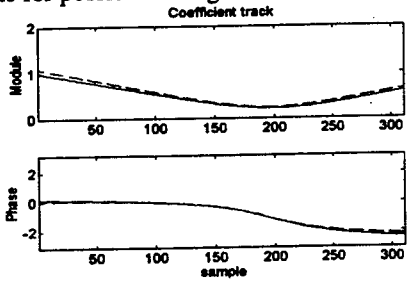


Fig. 2: Tracking for the first tap of a RA250 channel vs. time in amplitude and phase for the proposed algorithm. Dashed lines stand for the true channel; solid lines for the TDBW estimate.

Channel tracking properties of the proposed algorithm are shown in Fig. 2 and Fig. 3. It can be observed that such properties are good as long as the linear approximation for the channel evolution is feasible. Comparing those figures with those obtained with the ABW algorithm (Fig. 4), we conclude that CIR tracking is now much less noisy.

In addition to this, now there is no need for waiting the algorithm to converge within the first samples of each timeslot. Moreover, the TDBW version is far more robust against deep fades which usually make the ABW algorithm to lose tracking. The reason for this robustness is that TDBW is a batch-type algorithm, where every sample in the

timeslot is used to estimate the CIR in every instant (even though in deep local fades), whereas in the ABW version the estimate relies mainly on the previous and, maybe, already-faded samples. Of course, those improvements are conditioned to an approximately linear variation of the channel, what is not required in the ABW algorithm. However, if this requirement was not met, it would always be possible to increase polynomial order to obtain a better approximation for the channel evolution.

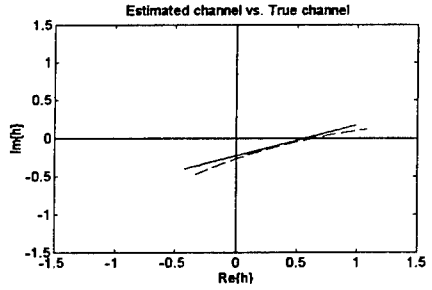


Fig. 3: Tracking for the first tap of a RA250 channel in rectangular coordinates ($Re\{h_1(t)\}$ and $Im\{h_1(t)\}$). Dashed lines stand for the true channel; solid lines for the TDBW estimate.

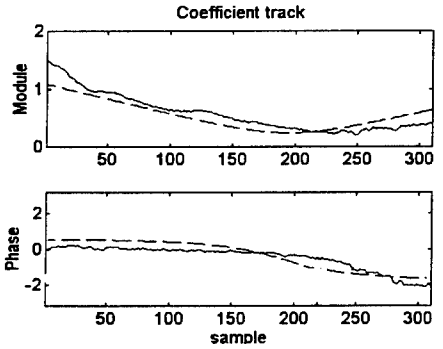


Fig. 4: Tracking for the first tap of a RA250 channel vs. time. Dashed lines stand for the true channel; solid lines for the ABW estimate.

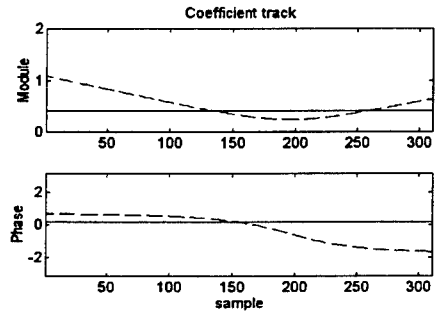


Fig. 5: Tracking for the first tap of a RA250 channel vs. time in amplitude and phase. Dashed lines stand for the true channel; solid lines for the BBW estimate.

The main advantage with respect to the BBW algorithm described in section 3, is the ability to track the evolution of the CIR along the timeslot, instead of approximating each tap by a constant value (Fig. 5). For mobile stations exhibiting rather high speed, it reverts in a lower BER.

On the other hand, the main drawback of the proposed TDBW algorithm is the increase in the computational burden when compared with both the BBW and ABW versions. And, what is more, now the number of parameters to be estimated is double the quantity required before ($\mathbf{m}^{(0}$ and $\mathbf{m}^{(1}$ vs. $\mathbf{m}$), whereas the amount of data available to perform that estimation is just the same (one timeslot). Consequently, for CIRs exhibiting large delay-spreads such as HT (Hilly Terrain environment), the variance increase in the estimation of some components in $\mathbf{m}[n]$ is very severe and the Weighting Matrix cannot prevent the system from unstability. In those cases, the only way to make the algorithm to converge is considering larger timeslots which contain more symbols.

6. Conclusions

A new technique to include the time-varying nature of the parameters of a HMM in the BW reestimation formulas has been presented. The resulting algorithm for blind channel estimation, TDBW, has been compared with those proposed in previous references (BBW and ABW), and its performance qualitatively evaluated in a very concrete environment (the GSM system). The most important drawback of the algorithm is its high computational cost.

Future work is concerned about applying the theory of HMMs and the developed BW-based algorithms to other communication environments such as Underwater Acoustics (UWA), or in other communication problems.

References

[1] J.G Proakis and C.L. Nikias, "Blind Equalization", *Proc. SPIE Adaptive Signal Processing* 1991, Vol. 1565, pp. 76-87, 1991.

[2] John Shynk et al., "A Comparative Performance Study of Several Blind Equalization Algorithms", *Proc. SPIE Adaptive Signal Processing* 1991, Vol. 1565, pp. 102-117, 1991.

[3] M. Ghosh and C.L. Weber, "Maximum Likelihood Blind Equalization", *Proc. SPIE Adaptive Signal Processing* 1991, Vol. 1565, pp. 188-195, 1991.

[4] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proc. IEEE*, Vol. 77, No. 2, pp. 257-286, Feb. 1989.

[5] R. Steele, "Mobile Radio Communications", *Pentech Press Publishers*, London 1992.

[6] T. Aulin and C.W. Sundberg, "Continuous Phase Modulations (Parts I and II)", *IEEE Trans. on Comm.*, Vol COM-29, No. 3, pp. 196-225, March 1981.

[7] José A.R. Fonollosa and J. Vidal, "Application of Hidden Markov Models to Blind Channel Characterization and Data Detection", *Proc. IEEE Int. Conf. Accoust. Speech and Signal Processing*, pp. 185-188, Australia, April 1994.

$$\phi[n] = \pi \sum_{r=-R}^{R} q[r]\, a[n-r] + \theta[n] \qquad (3)$$

where $q[r] \in [0,0.5]$ are the weights corresponding to the (sampled) gaussian-shaping pulse and $\theta[n] \in \{0, \pi/2, \pi, 3\pi/2\}$ accounts for the accumulated phase at instant n [5,6]. Now we conclude that the number of transmitter symbols (bits) involved in a single observation at the receiver is given by:

$$l_t = l_m + l_c - 1 = (2R+1) + l_c - 1 \qquad (4)$$

However, the amount of ISI produced by the GMSK modulator for BT=0.3 can be neglected without significant performance loss. Under this simplifying assumption, which reverts in a lower number of states and a reduced computational complexity, we get that:

$$R = 0 \Rightarrow l_t = l_c \qquad (5)$$

At this point, we can already model each observation in the received sequence, $\mathbf{x}_L = (x[1], x[2], .., x[L])^T$, as a probabilistic function of the present state $\mathbf{s}[n] = (a[n], .., a[n-l_t+1], \theta[n])^T$, obtaining a description of $\mathbf{x}_L$ as a first order HMM with $N = 4 \cdot 2^{l_t}$ states.

3. Overview of the BW algorithm

On the basis of this first order HMM and by means of the BW algorithm, it is possible to obtain a solution to the problem of the identification of the unknown parameters of the model, that is, σ^2 and $\mathbf{h} = (h_0 .. h_{l_c-1})^T$ [7]. To be precise, the parameters really estimated are σ^2 and the means vector $\mathbf{m} = (m_1 .. m_N)^T$ corresponding to the noise-free ISI-corrupted received signal associated with the N states of the system. Assuming a FIR model for the channel, $\mathbf{m}$ is related to $\mathbf{h}$ through the linear constraints:

$$\begin{aligned} \hat{\mathbf{m}} &= \mathbf{D}\hat{\mathbf{h}} \\ \hat{\mathbf{h}} &= \mathbf{D}^{\#}\hat{\mathbf{m}} \end{aligned} \qquad (6)$$

where $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, .., \mathbf{d}_N)^T$ is a $N \times l_c$ full-rank matrix containing in its rows all the l_c-tuples, $\mathbf{d}_i = (d_i^{(1)}, d_i^{(2)}, .., d_i^{(l_c)})^T$, corresponding to the modulator consecutive outputs associated to the N different states of the system. $\mathbf{D}^{\#}$ denotes pseudoinverse. The batch Baum&Welch (BBW) algorithm, thoroughly explained in [7], can be outlined as:

1. Projection of $\mathbf{h}$ on $\mathbf{m}$ by means of the additional linear constraints:
$$\hat{\mathbf{m}} = \mathbf{D}\hat{\mathbf{h}} \qquad (7)$$
2. Reestimation of σ^2 and $\mathbf{m}$ using the BW reestimation formulas.
3. Least Squares (LS) estimation of $\mathbf{h}$ using again linear constraints.
$$\hat{\mathbf{h}} = \mathbf{D}^{\#}\hat{\mathbf{m}} \qquad (8)$$
4. Repeat steps 1..4 until convergence.

The BW reestimation formulas used in step 2, state as follows:

$$\begin{aligned} \hat{m}_i &= \frac{\sum_{n=1}^{L} \gamma_i[n]\, x[n]}{\sum_{n=1}^{L} \gamma_i[n]}, \qquad 1 \le i \le N \\ \hat{\sigma}^2 &= \frac{1}{L} \sum_{n=1}^{L} \sum_{i=1}^{N} \gamma_i[n] \left| x[n] - \hat{m}_i \right|^2 \end{aligned} \qquad (9)$$

where $\gamma_i[n]$ is the probability of being in state i at instant n given the model and the observed sequence. However, these formulas implicitly assume the CIR to be stationary within a timeslot duration, what is not realistic when the timeslot is long enough or the channel varies rapidly. Hence, we will obtain other reestimation formulas to solve this problem.

4. Modified algorithm

Several strategies can be considered to cope with the time-varying nature of the channel: estimating the CIR with a recursive adaptation scheme such as LMS what is referred in [7] as the Adaptive BW (ABW) algorithm; or fragmenting each timeslot in subblocks. As opposed to those methods, we will try to include the time-varying nature of the channel directly in the reestimation formulas.

We can approximate the evolution of every tap in the CIR, h_j, by means of a polynomial in n:

$$\hat{h}_j[n] = h_j^{(0} + h_j^{(1} \cdot n + h_j^{(2} \cdot n^2 + ... \qquad (10)$$

For the channels specified in the ETSI recommendations and assuming the speed of the mobile to be less than 250 km/h, the linear approximation was observed to be good enough. Assembling all the taps in a single column vector:

$$\hat{\mathbf{h}}[n] = \mathbf{h}^{(0} + \mathbf{h}^{(1} \cdot n \qquad (11)$$

and applying the linear transformation above described

$$\hat{\mathbf{m}}[n] = \mathbf{D}\hat{\mathbf{h}}[n] = \mathbf{D}\left(\mathbf{h}^{(0} + \mathbf{h}^{(1} \cdot n\right) = \mathbf{m}^{(0} + \mathbf{m}^{(1} \cdot n \qquad (12)$$

we observe that the evolution for the means is also linear. Vectors $\mathbf{m}^{(0}$ and $\mathbf{m}^{(1}$ will be obtained in order to minimize the MSE given by the following expression:

Robust Blind Joint Data/Channel Estimation Based on Bilinear Optimization

David Gesbert
PAB/RGF/RCN
France Telecom-CNET
92131 Issy-Les-Mlx, France
gesbert@issy.cnet.fr

Pierre Duhamel
Dept. Signal
Telecom Paris
75013 Paris, France
duhamel@sig.enst.fr

Abstract

The problem of identifying/equalizing a digital communication channel based on its temporally or spatially oversampled output has recently gained much attention (single-input/multiple-ouptut - SIMO - deconvolution). In this context, we propose a new joint data/channel estimation method. Our technique relies on the minimization of a bilinear MSE cost function, where the variables to be adjusted are the channel coefficient matrix and a linear equalizer. We show that this a priori *choice of a linear equalization structure allows the derivation of a second-order unimodal criterion, leading to globally convergent identification/equalization schemes. The proposed method is completely blind in that 1) no assumption is required upon the transmitted sequence statistics or alphabet, and 2) it shows some robustness with respect to the channel order estimation problem (thus improving on most previous related works). It also allows the free choice of a delay in the equalizer so that output noise amplification can be optimized.*

1 Introduction

In the context of digital radiocommunications, the signals are transmitted through propagation channels which introduce intersymbol interference (ISI). The channels can be represented as FIR filters which have to be identified and/or equalized for the transmitted symbols to be recovered. Since the pioneering work by Sato [1], all blind equalization techniques (which do not rely on training sequences) have been based on the use of higher-order statistics (HOS) of the received signals, though HOS methods largely suffer from slow and ill convergence problems [2]. Recently, it was shown by Gardner [3] and Tong *et al.* [7] that blind deconvolution based on the sole second-order statistics was a feasible task, provided the observed signals could be seen as the outputs of a SIMO system with sufficient channel disparity (the different channels-polynomials should not have any common zero). In this context, a number of contributions have been made in which the transmitted sequence or the channel coefficients are recovered through subspace decompositions of either the received data matrix (see the so-called deterministic methods [4,5,6]) or the received data correlation matrix (see the stochastic methods [7,8]). Other interesting approaches were also studied in [9,10,11,12]. Here, we introduce a blind and mainly adaptive estimation method in which a multichannel estimate and a linear equalizer are adjusted so as to minimize an observation fitting cost function. The possible local minima of the proposed criterion are investigated and global convergence is established. The presented algorithm shows several attractive features which make it an interesting alternative to most existing methods:

- First, it shows some robustness with respect to the additive noise (though it is optimal only in the noise free case) and to possible errors in the channel order estimation.
- It also allows the use of any reconstruction delays in the equalizer so that noise variance may be optimized at the equalizer's output.

Notations: $\Re$ real part of a complex. $E()$ statistical expectation; $()^*$ complex conjugation; $()^t$ transposition; $()^+$ trans-conjugation; $|\ .\ |$ L_2-norm of a vector or matrix; I identity matrix.

2 Multichannel representation

The SIMO equivalent model of a digital communication system relies on the existence of a number L of different linear time-limited digital filters (channels) h^1,.., h^L, driven by the *same* PAM/QAM sequence s_k, the noisy outputs of which are observed:

$$x_n^i = \sum_{k=0}^{M} s_{n-k} h_k^i + b_n^i \quad \text{for } i = 1..L \qquad (1)$$

s_k and $b_n^i, i = 1..L$, are mutually uncorrelated processes, not necessarily white. We assume (w.l.o.g.) $E\,|s_k|^2 = 1$, $E\,|b_n^i|^2 = \sigma_b^2$. In the context of antenna-array based reception, the channel h^i represents the baud-rate impulse response of the propagation channel linking the transmitter and the i^{th} antenna (spatial

0-8186-7576-4/96 $5.00

diversity). In a mono-antenna scenario, channel diversity can still be obtained by means of temporal oversampling with a factor L (compared to the baud rate) at the antenna output, leading to fractionally-spaced (FS) reception. In the FS context, the channels $h^{(i)}$ correspond to sampled versions (at rate T) of a single propagation channel, at various sampling phases $(i-1)T/L$, $i = 1..L$ (see [7] for more details). Here, M denotes the ISI length. We adopt the following vectorized notations:

$$\begin{aligned} \mathbf{x}_n &= [x_n^1, .., x_n^L]^t, \\ \mathbf{b}_n &= [b_n^1, .., b_n^L]^t, \\ \mathbf{h}_k &= [h_k^1, ..., h_k^L]^t, \\ \mathbf{h} &= [\mathbf{h}_0, .., \mathbf{h}_M] \end{aligned}$$

Then, we have

$$\mathbf{x}_n = \mathbf{h}[s_n, ..., s_{n-M}]^t + \mathbf{b}_n \qquad (2)$$

Consider the space-time samples vectors $X_n = [\mathbf{x}_n^t, \mathbf{x}_{n-1}^t, \cdots, \mathbf{x}_{n-N+1}^t]^t$, $B_n = [\mathbf{b}_n^t, \mathbf{b}_{n-1}^t, \cdots, \mathbf{b}_{n-N+1}^t]^t$ and $S_n = [s_n, \cdots, s_{n-P+1}]^t$, where N is the window size per channel and $P = M+N$ is the number of symbols involved in the expression of vector X_n. The following linear model holds

$$X_n = \mathcal{T}(\mathbf{h})S_n + B_n, \qquad (3)$$

where $\mathcal{T}(\mathbf{h})$ is the so-called $LN \times P$ Sylvester matrix:

$$\mathcal{T}(\mathbf{h}) = \begin{pmatrix} \mathbf{h}(0) & \cdots & \mathbf{h}(M) & \mathbf{0} & \cdots & \mathbf{0} \\ & & & \vdots & & \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{h}(0) & \cdots & \mathbf{h}(M) \end{pmatrix}.$$

To enable blind deconvolution of our SIMO system, we assume throughout the paper that **(H1)** $\mathcal{T}(\mathbf{h})$ *has full column rank* P, with $LN \geq P$. In the following, we concentrate on the joint blind estimation of the pair $(\mathbf{h}, \omega)$ where ω is a $LN \times 1$ linear equalizer satisfying the following condition in the noiseless case: $\omega^t \mathcal{T}(\mathbf{h}) = [0, .., 0, 1, 0, .., 0]$. Note that, although written as a zero-forcing (ZF) equalization problem, the actual (noisy) problem does not fully reduces to some ZF equalizer.

3 The proposed method

Assume in a first step that M is known. Consider an $L \times (M+1)$ matrix $\hat{\mathbf{h}}$ (channels estimate) and an $LN \times 1$ vector $\hat{\omega}$ (equalizer estimate). Our algorithm aims at tuning these channel and equalizer estimates so that the convolution product between each channel estimate and the equalizer output matches the observed signals as sketched in fig.1. Mathematically, this writes:

$$\text{minimize} \quad J(\hat{\mathbf{h}}, \hat{\omega}) = E \mid \mathbf{x}_n - \hat{\mathbf{h}}\bar{X}_n^t\hat{\omega} \mid^2, \qquad (4)$$

where $\bar{X}_n = [X_n, X_{n-1}, .., X_{n-M}]$ is a $LN \times (M+1)$ sample matrix. This observation fitting criterion is *bilinear* in the coefficients of $\theta = (\hat{\mathbf{h}}, \hat{\omega})$. It is similar in spirit to the previously proposed deterministic maximum likelihood (DML) criterion [9]. The DML method, in which the linear equalizer is typically not a free variable (being replaced by a pseudo-inverse of $\mathcal{T}(\mathbf{h})$), is however subject to ill-convergence and is computationally demanding as well. This is not the case here, as will be shown in the following.

3.1 Criterion minima

Assume a noise free ($\sigma_b^2 = 0$) situation. Consider any solution of the form $\theta = (\alpha\mathbf{h}, \omega/\alpha)$, where ω is some ideal zero-delay equalizer and α is an arbitrary complex scalar. Clearly, θ achieves global minimization of our criterion, thus provides a stationary point of $J()$, since $J()$ is positive. Conversely, it may be shown that $J = 0$ leads to channel equalization and identification in the absence of noise:

Lemma 3.1 *Let $\mathbf{z}_n$ be the $L \times 1$ residual error process for some $\theta = (\hat{\mathbf{h}}, \hat{\omega})$, defined as $\mathbf{z}_n = \mathbf{x}_n - \hat{\mathbf{h}}\bar{X}_n^t\hat{\omega}$. Assume $\{s_k\}$ is persistently exciting of order at least $2M$. Suppose $J(\theta) = 0$, i.e. $\mathbf{z}_n = 0$ almost surely. Then $\theta = (\alpha\mathbf{h}, \omega/\alpha)$, where ω is an ideal zero-delay equalizer and α is a complex scalar.*

Proof If $\mathbf{x}_n = \hat{\mathbf{h}}\bar{X}_n^t\hat{\omega}$, we also have from (3)

$$X_n = \mathcal{T}(\mathbf{h})S_n = \mathcal{T}(\hat{\mathbf{h}})[X_n, .., X_{n-P+1}]^t\hat{\omega}$$

under the persistent excitation condition, the subspace spanned by the observed vectors X_n is found to be simultaneously $\mathcal{T}(\mathbf{h})$ and $\mathcal{T}(\hat{\mathbf{h}})$. By theorem 2 in [8], we have $\hat{\mathbf{h}} = \alpha\mathbf{h}$. It follows that $s_n = \alpha X_n^t\hat{\omega}$, showing that $\hat{\omega}/\alpha$ is a zero-delay equalizer.□

3.2 Stability of minima

Here, we check the absence of undesired stable local minima. Let $\theta = (\hat{\mathbf{h}}, \hat{\omega})$ be any stationary point (cancelling the first partial derivatives) of $J()$. The stability of θ is investigated through the criterion second order expansion: Let $\delta\theta = (\delta\hat{\mathbf{h}}, \delta\hat{\omega})$ be a small move around θ. Let $\Delta J = J(\theta + \delta\theta) - J(\theta)$, we find, up to the second order:

$$\Delta J \approx E \mid \mathbf{z}_n - \delta\hat{\mathbf{h}}\bar{X}_n^t\hat{\omega} - \hat{\mathbf{h}}\bar{X}_n^t\delta\hat{\omega} \mid^2 - E \mid \mathbf{z}_n \mid^2,$$

where $\mathbf{z}_n$ is the residual of θ, i.e. $(\mathbf{x}_n - \hat{\mathbf{h}}\bar{X}_n^t\hat{\omega})$. It can be inferred that

Lemma 3.2 *θ is a stable minimum if and only if all components in $\mathbf{z}_n$ are decorrelated from $\bar{X}_n^t$. Due to the particular form of $\mathbf{z}_n$, this implies $\mathbf{z}_n = 0$.*

Due to lack of space, the proof of this lemma will be detailed in a forthcoming paper. Now we have established that a gradient-based algorithm based on the noise free criterion in (4) always converges to the true (channels, equalizer) pair, up to an arbitrary scalar constant.

3.3 Robustness

With respect to noise The presence of additive noise causes bias in the results, however the simulations indicate that acceptable channel and equalizer estimates can be obtained from this algorithm under realistic SNR conditions. Bias removal techniques, based on some norm-constrained minimization, can also be adapted to our problem [14].

With respect to model order In practical situations, the multichannel order (M) is probably not well defined, especially in the case where the channels coefficients taper off at the ends. Then, overestimation of M is likely to occur. We stress the robustness of the proposed method with respect to such errors. The proof goes as follows: Let $\hat{\mathbf{h}} = [\mathbf{h}_0, .., \mathbf{h}_K]$ be the channel candidate, with $K > M$. As in the case of correct order estimation, the minimization of $J(\hat{\mathbf{h}}, \hat{\omega})$ leads to $\mathbf{x}_n = \hat{\mathbf{h}}\bar{X}_n^t\hat{\omega}$, this again gives $span(\mathcal{T}(\mathbf{h})) = span(\mathcal{T}(\hat{\mathbf{h}}))$. As a result the channels in $\hat{\mathbf{h}}$ admit a $K-M$ order common polynomial factor, denoted $q(z)$, and $\mathcal{T}(\hat{\mathbf{h}})$ factors into $\mathcal{T}(\mathbf{h})Q$ where Q is the $P \times (N+K)$ Sylvester matrix associated to $q(z)$ [13]. We have

$$\begin{aligned} \mathcal{T}(\mathbf{h})S_n &= \mathcal{T}(\mathbf{h})Q[S_n, .., S_{n-K-N+1}]^t\mathcal{T}(\mathbf{h})^t\hat{\omega} \\ S_n &= Q[S_n, .., S_{n-K-N+1}]^t\mathcal{T}(\mathbf{h})^t\hat{\omega} \end{aligned}$$

and it is clearly seen that such a condition cannot hold with $q(z) = 1$, unless there exists some recurrence relationship between the successive emitted symbols, a fact which is not compatible with the persistent excitation assumption. As a result, the only possible solution of the above equation is $q(z) = 1$ and $\mathcal{T}(\mathbf{h})^t\hat{\omega} = I$. This result is checked below in the simulations section.

4 Equalization with non-zero delay

The performances of a linear multichannel equalizer generally depend on its delay [15]. Hence, it is useful to control the delay introduced by the equalizer. This is easily obtained by rewriting the proposed criterion as : $J(\hat{\mathbf{h}}, \omega) = E \mid x_{n-d} - \hat{\mathbf{h}}\bar{X}_n^t\omega \mid^2$, where d is a chosen delay parameter. Note however that in case of model order overestimation, the *actual* equalizer's delay may not be determined in advance, since the non zero channel coefficents estimates in $\hat{\mathbf{h}}$ are subject to a possible shift. However this problem should not be very severe in presence of noise.

5 Adaptive algorithm

A possible implementation of the proposed method which allows full adaptivity in the context of time varying statistics/channels is as follows, based on a stochastic gradient approximation. Note that other recursive least-square based approaches can also be used.

$$Y_n = \bar{X}_n^t\omega_n,$$

$$\begin{aligned} \hat{\mathbf{h}}_{n+1} &= \hat{\mathbf{h}}_n - \lambda(\hat{\mathbf{h}}_n Y_n - \mathbf{x}_n)Y_n^+, \\ \hat{\omega}_{n+1} &= \hat{\omega}_n - \lambda\bar{X}_n^*\hat{\mathbf{h}}_n^+(\hat{\mathbf{h}}_n Y_n - \mathbf{x}_n), \end{aligned}$$

where λ is a small stepsize.

6 simulations

We consider the context of $L = 2$ randomly chosen channels of length $M+1 = 5$, given by $h^1 = [-0.089 - 0.489j; -0.340 - 0.016j; 0.022 - 0.069j; -0.192 - 0.031j; 0.464 - 0.613j]$, $h^2 = [0.422 + 0.467j; -0.075 + 0.320j; 0.185 - 0.049j; 0.223 + 0.122j; 0.145 - 0.609j]$, driven by a white QPSK sequence. Output SNR is set to $15dB$ on each of the (normalized) channels. We choose $N = 5$ and consider only the zero delay case ($d = 0$). Fig.2 shows the equalization results in terms of output mean square error between the transmitted and the recovered symbols, using the linear equalizer provided by the algorithm $\hat{\omega}_n$, versus the iteration number n. Both cases of a correct model order estimation ($K = M = 4$) and of a severe overestimation ($K = 8$) are illustrated.

Fig.3 shows the identification results in terms of the distance between $\hat{\mathbf{h}}_n$ and the true channels, up to α, where α is defined in lemma 3.1, defined by $\mid \hat{\mathbf{h}}_n/\alpha - \mathbf{h} \mid^2 / \mid \mathbf{h} \mid^2$. Note that channel identification is well achieved despite the additive noise. This is due to the MSE-like structure of the criterion. Robustness with respect to the model order error is confirmed by both equalization and identification results, though the obtained performances seem to degrade in the case $K = 8$. The rise in the steady-state error is due to adaptation noise, which can be compensated for by decreasing the stepsize. This robustness gives advantage over the methods found in [7,8,4,5].

7 Discussion

We have addressed the problem of blind (adaptive) estimation of both the channel coefficients and the imput in a SIMO context. In the proposed criterion, the channel estimate and a linear equalizer are independant variables to be adjusted so as to match the observed signal in a least mean square sense. The minimization of this criterion asymptotically leads the "true" (channel,equalizer) pair in the absence of noise, up to some scalar constant, and is robust to possible model order errors thanks to the particular structure chosen for the equalizer. A gradient descent implementation was proposed, providing however rather slow convergence. Further work will include the study of implementations of the recursive least-squares type, for example by alternating RLS algorithms on the equalizer and the channels estimates.

References

[1] Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation systems," *IEEE Trans. on Comm.*, vol. 23, pp. 679–682, June 1975.

[2] S. Haykin, Editor, *Blind deconvolution.* Prentice Hall, Englewood Cliffs, NJ, 1994.

[3] W. A. Gardner, "A new method of channel identification," *IEEE Trans. on Comm.*, vol. 39, pp. 813–817, June 1975.

[4] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind identification of multichannel FIR systems," in*Proc. ICASSP* , vol. 4, pp.581–584,1994.

[5] H. Liu and G. Xu, "A deterministic approach to blind symbol estimation," in *IEEE SP Letters*, vol 1, n 12, dec. 1994

[6] A. J. Vand der Veen, S. Talwar, and A. Paulraj, "Blind estimation of multiple digital signals transmitted over FIR channels," *IEEE S.P. Letters*, vol. 2, pp. 99–102, May 1995.

[7] L.Tong, G.Xu, and T.Kailath, "Blind identification and equalization based on second-order statistics: a time-domain approach," *IEEE Tr. on Information Theory*, vol. 40, no. 2 , pp. 340–349, March 1994.

[8] E. Moulines, P. Duhamel, J.F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. SP*, vol. 43, no. 2, pp. 516–526, February 1995.

[9] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filter-banks and multichannel linear prediction," in *Proc. ICASSP*, vol. 4, pp. 585-588, 1994.

[10] K. Abed Meraim, P. Duhamel, D. Gesbert, P. Loubaton, S. Mayrargue, E. Moulines, and D. Slock, "Prediction error methods for time-domain blind identification of multichannel FIR filters," in *Proc. ICASSP*, 1995.

[11] D. Gesbert, P. Duhamel, and S. Mayrargue, "Subspace-based adaptive algorithms for the blind equalization of multichannel FIR filters," in *Proc. EUSIPCO*, 1994.

[12] D. Gesbert and P. Duhamel, and S. Mayrargue, "Blind Least-Squares Approaches for joint data/channel estimation," submitted to the 7^{th} *IEEE Digital Signal Processing Workshop*, 1996.

[13] I. Fijalkow, J. R. Treichler, and C. R. Johnson Jr, "Fractionally spaced blind equalization: loss of channel disparity," in *Proc. ICASSP*, 1995.

[14] D. Gesbert, P. Duhamel, and S. Mayrargue, "A bias removal technique for the prediction-based blind adaptive multichannel deconvolution," in *Proc. 29th Asilomar Conf. on Signals, Systems, and Computers*, 1995.

[15] D. Gesbert, P. Duhamel, and S. Mayrargue, "Blind multichannel adaptive MMSE equalization with controlled delay," in *proc. 8^{th} IEEE Signal Processing Workshop on SSAP*, 1996.

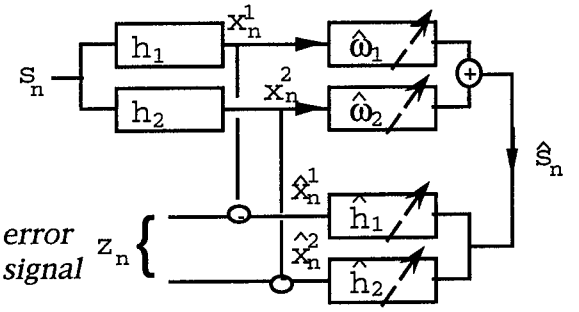


Figure 1: **The joint channels/equalizer estimation setup.**

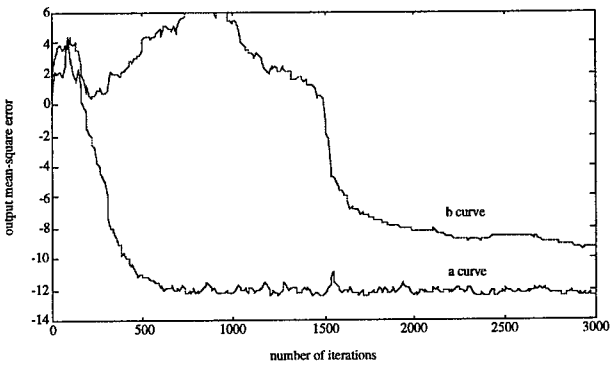


Figure 2: **Output mean square error using the linear equalizer $\hat{\omega}$ by the gradient algorithm in 5. $\lambda = 0.02$. 'a' curve: $K = 4$, 'b' curve: $K = 8$**

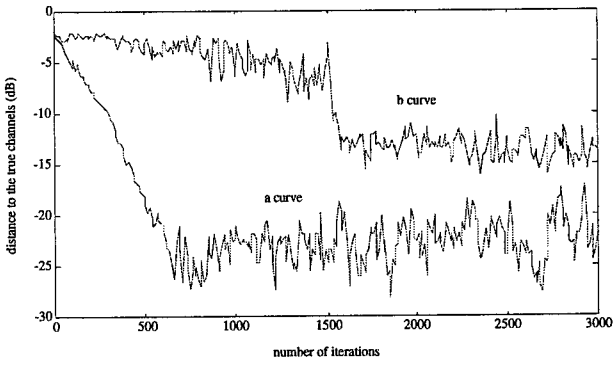


Figure 3: **Identification by the channel estimate $\hat{\mathbf{h}}$ by the gradient algorithm in 5. $\lambda = 0.02$. 'a' curve: K=4, 'b' curve: K=8**

Blind Multichannel Adaptive MMSE Equalization with Controlled Delay

David Gesbert
PAB/RGF/RCN
France Telecom-CNET
92131 Issy-Les-Mlx, France
gesbert@issy.cnet.fr

Pierre Duhamel
Dept. Signal
Telecom Paris
75013 Paris, France
duhamel@sig.enst.fr

Sylvie Mayrargue
PAB/RGF/RCN
France Telecom-CNET
92131 Issy-Les-Mlx, France
mayrargue@issy.cnet.fr

Abstract

This contribution adresses the problem of optimal blind linear symbol recovery, using the channel diversity induced by a sensor array or time oversampling. We present a technique allowing the computation of a minimum mean-square error (MMSE) equalizer, based on the optimization of quadratic second-order functions. The proposed technique improves on existing multichannel equalization methods, in that previous methods generally build on criteria which are optimal in the sole noise free context. Our criterion also allows free choice of the delay for the symbol recovery. As a consequence, MMSE equalization performance can be enhanced through the use of an optimal delay. To this end, a performance analysis is conducted in order to investigate some of the links between the delay and the symbol estimation accuracy.

1 Introduction

Blind multichannel deconvolution exploiting the channel diversity induced by sensor arrays and/or time oversampling has attracted a lot of research efforts in the recent years. Methods can now be found in the literature, based on the minimization of various second-order criteria, which offer promising alternatives to the previously reported higher-order based techniques. These methods are basically multichannel batch deconvolution techniques in which transmitted sequence or channel coefficients are recovered through subspace decompositions of either the received data matrix (see the so-called deterministic methods [3,5,4]) or the received data correlation matrix (see the stochastic methods [1,2]), while other interesting approaches were also studied in [7,8,9,12]. Some methods, performing a channel pre-identification, have to be linked to an extra equalization stage, thus increasing the global cost of the reception scheme. In the (numerous) communications applications where low complexity and/or tracking ability is sought, *direct* online equalization techniques requiring no channel pre-identification are to be favored. Significant improvements can also be gained in this context from channel diversity. These improvements concern two main points : **(i)** convergence reliability (diversity allows the use of second-order unimodal error function for equalization), and **(ii)** equalization accuracy, using a simple linear structure since finite-length zero-forcing equalizers are available in the multichannel context. However, these existing direct equalization methods generally suffer from a lack of robustness in the sense that they build on criteria which are optimal in the noise free context, but not (or even far from) optimal in the practical noisy situations. This typically includes the prediction-based methods [7,8] but also, in a lower extent, the mutually referenced equalizers (MRE) method in [9]. As another lack of optimality, most existing on-line multichannel techniques ([9] being an exception) are unable to exploit the performance gain that stems from the choice of a proper delay for symbol recovery [7,10]. This paper investigates the solutions to these problems. Our contribution is two-fold: **(1)** we present a technique, based on a modification of the criterion initially introduced in [9], allowing the derivation of a blind linear multichannel equalizer, optimal in the MMSE sense. **(2)** By this approach, we show that we may improve on the robustness of the obtained equalizer through a proper tuning of the reconstruction delay : The criterion is optimal in the MMSE sense for a given reconstruction delay, which can be chosen so that the corresponding MMSE is minimal among all possible delays. Finally, a theoretical study is conducted, based on the computation of Cramer-Rao bounds, that provides a guideline for the choice of the optimal delay. **Notations:** $E()$ statistical expectation; $()^*$ complex conjugation; $()^t$ transposition; $()^+$ trans-conjugation; $|\ .\ |$ L_2-norm of a complex scalar, vector or matrix.

2 Problem statement

The multichannel model of a digital communication system relies on the existence of L different channels, modeled by finite degree linear digital filters $h^1,..,h^L$, driven by the *same* PAM/QAM sequence s_k. Their L noisy outputs x_n^i are observed:

$$x_n^i = \sum_{k=0}^{M} s_{n-k} h_k^i + b_n^i \quad \text{for } i = 1..L \qquad (1)$$

We assume throughout the paper that the sequences s_k and $b_n^i, i = 1..L$, are mutually uncorrelated. $b_n^i, i = 1..L$ being *white* processes, $E \mid b_n^i \mid^2 = \sigma_b^2$. Note that the emitted sequence s_k is not required to be white, in contrast with [7,8,10]. In a practical context, the h^i's either represent the baud-rate sampled versions of a single physical channel with various sampling phases (time diversity), or the channel linking the transmitter and the i^{th} sensor on an antenna array (space diversity), see [1,2] for further details. Here, M denotes the ISI length. Consider the $LN \times 1$ space-time signal and noise processes with time window of size N defined by $X_n = [x_n^1, .., x_{n-N+1}^1, ..., x_n^L, .., x_{n-N+1}^L]^t$ and $B_n = [b_n^1, .., b_{n-N+1}^1, ..., b_n^L, .., b_{n-N+1}^L]^t$. Let $P = M + N$ denote the number of symbols involved in the expression of X_n. The following linear model holds

$$X_n = \mathcal{H} S_n + B_n \tag{2}$$

where $S_n = [s_n, ..., s(n-P+1)]^t$ and $\mathcal{H}$ is the so-called $LN \times P$ Sylvester matrix, defined by

$$\mathcal{H} = \begin{pmatrix} h_0^1 & \cdots & h_M^1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & h_0^1 & \cdots & h_M^1 \\ & & \vdots & & & \\ h_0^L & \cdots & h_M^L & 0 & \cdots & 0 \\ \vdots & & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & h_0^L & \cdots & h_M^L \end{pmatrix}$$

The identifiability/equalizability conditions of the single input/multiple output system above are established in [1,6], and can be restated as

- **(H1)** $\mathcal{H}$ has full column rank ($LN \geq P$).

3 Blind MMSE Equalization

In the following, it is shown how noise free adaptive linear equalizers, i.e. vectors satisfying the ZF condition, can first be obtained using the method of mutually referenced filters, and then be exploited to train a blind MMSE equalizer.

3.1 Derivation of adaptive ZF equalizers

It is seen from (2) that a $LN \times 1$ vector (denoted V_j), given by any line of any left-inverse matrix of $\mathcal{H}$ will satisfy the ZF condition up to a delay between 0 and $P-1$, i.e. $V_j^+ \mathcal{H} = [0, .., 0, 1, 0, .., 0]$. The MRE criterion introduced in [9], provides a useful means to compute blindly and adaptively such equalizers, in the noise free case. The main result of [9] can be restated as follows:

Lemma 3.1 *Consider a set of P linear equalizers $V_0, V_1, ..., V_{P-1}$ and adjust them so as to cancel the quadratic function*

$$J_{mre}(V_0, .., V_{P-1}) = \sum_{j=0}^{P-2} E \mid V_j^+ X_n - V_{j+1}^+ X_{n+1} \mid^2$$

under constraint

$$\sum_{j=0}^{P-1} \| V_j \|^2 = 1$$

Then V_j, for $j = 0..P-1$, is a "j-delay" exact ZF equalizer , i.e. $V_j^+ X_n = \alpha s_{n-j}$, where α is an arbitrary constant scalar.

In this result, the information redundancy provided by several independant equalizers associated with different delays is exploited to build a second-order criterion. However, in the noisy situations, the minimizers of the MRE criterion are no longer optimal and the obtained filters $\{V_j\}$ are biased. Fortunately, the the MRE criterion can be modified so that the new performance surface gets noise independant. The proposed bias removal technique, inspired from [11], consists in replacing the former unit-norm constraint by a new quadratic constraint that incorporates the knowledge of the noise covariance matrix structure. The technique is as follows: Let $V = [V_0^+, V1^+, .., V_{P-1}^+]^+$ be the $LNP \times 1$ vector consisting of all the equalizers entries. J_{mre} can be rewritten in a compact matrix form as $J_{mre}(V) = V^+ \mathcal{R} V$ where $\mathcal{R}$ is a sparse $LNP \times LNP$ matrix made from sub-blocks $E(X_{n+1} X_n^+)$ and $E(X_n X_n^+)$. Under the white noise and noise/signal decorrelation assumptions, $\mathcal{R}$ splits into a noise and a signal part as:

$$\mathcal{R} = \mathcal{R}_s + \sigma_b^2 \mathcal{R}_b \tag{3}$$

where $\mathcal{R}_s$ and $\mathcal{R}_b$ are the matrix forms of the criterion J_{mre} in the signal-only case and noise-only case respectively. Closed-form expression of $\mathcal{R}$, $\mathcal{R}_s$, and $\mathcal{R}_b$ are provided in appendix A. Note that $\mathcal{R}_s$ has a non trivial nullspace since J_{mre} can be cancelled in the absence of noise while $\mathcal{R}_b$ has full rank.

Lemma 3.2 *Let $V^{opt} = [V_0^{opt+}, .., V_{P-1}^{opt+}]^+$ be a $LNP \times 1$ vector minimizing $J_{mre}(V) = V^+ \mathcal{R} V$, under constraint $V^+ \mathcal{R}_b V = 1$. Then, for each j, $V_j^{opt+} X_n = \alpha s_{n-j} + V_j^{opt+} B_n$, i.e V_j^{opt} is an unbiased j-delay ZF equalizer.*

Proof It is easily shown using Lagrange multipliers that V^{opt} satisfies:

$$\mathcal{R} V^{opt} - (V^{opt+} \mathcal{R} V^{opt}) \mathcal{R}_b V^{opt} = 0 \tag{4}$$

with $(V^{opt+} \mathcal{R} V^{opt})$ taking the minimal possible value. From (3), the only solutions to this problem are such that $\mathcal{R}_s V^{opt} = 0$ which bring us back to the noise free solutions. Note that this technique do not require the knowledge of σ_b^2. □

3.2 Coupling with a MMSE equalizer

Since the newly constrained mutually referenced equalizers $\{V_j^{opt}\}$ asymptotically provide ISI free symbols estimates, any of these can be used to train a MMSE equalizer as in a fully supervised context: Let the $LN \times 1$ adaptive filter W^{opt} be obtained through the minimization of

$$J_{mse}(W) = E \mid W^+ X_n - V_j^{opt^+} X_{n-d+j} \mid^2 \quad (5)$$

W^{opt} is a d-delay MMSE equalizer, provided the noise contributions in X_n and X_{n-d+j} are decorrelated, so that the correct (unbiased) optimum is attained for W^{opt}. Since the noise is assumed to be temporally and spatially white, this condition is met if $d-j \geq N$. In the following, we provide theoretical hints enlightening the links between the delay and the steady-state performances of W^{opt}. A general guideline is provided for the choice of d.

4 On the Delay for Symbol Recovery

In this section, we investigate analytically the links between the chosen delay and the minimum achievable output MSE, regardless of the equalization criterion. The study is conducted by means of Cramer-Rao bounds. We essentially make the assumption that all transmitted symbols but one, within a temporal window of size N, are known deterministic quantities. Though this assumption may appear unrealistic, it not only greatly simplifies the mathematical developments but also permits the derivation of tractable and interpretable equations. We assume further that the noise is white gaussian. Then, at time (fixed) n, the p.d.f. of a space-time vector X_n writes:

$$\begin{aligned} f(X_n) &= Kexp(-\frac{1}{\sigma_b^2} \mid X_n - \mathcal{H}S_n \mid^2) \\ \frac{\partial log(f(X_n))}{\partial s_{n-d}} &= -\frac{1}{\sigma_b^2}(X_n - \mathcal{H}S_n)^+ \mathbf{h}_{d+1} \\ \frac{\partial}{\partial s_{n-d}^*} \frac{\partial log(f(X_n))}{\partial s_{n-d}} &= \frac{\mid \mathbf{h}_{d+1} \mid^2}{\sigma_b^2} \end{aligned}$$

where $\mathbf{h}_{d+1}$ denotes the $(d+1)^{th}$ column of the channel convolution matrix $\mathcal{H}$. Now we get the classical estimation theory result:

Lemma 4.3 *Let $\mathcal{T}_j(X_n)$ be any unbiased estimator (possibly non-linear) of s_{n-d}, the other symbols being seen as known deterministic variables. Then, we have*

$$E \mid \mathcal{T}_j(X_n) - s_{n-d} \mid^2 \geq CRB(s_{n-d}) = \frac{\sigma_b^2}{\| \mathbf{h}_{d+1} \|^2} \quad (6)$$

Due to the unrealistic assumptions made here, the indicated bounds have very limited practical applicability. For instance the obtained cramèr-Rao bound does not account for the possible lack of channel disparity in $\mathcal{H}$ which may cause some severe degradation in the estimation performance. However, for a relatively "well conditionned" matrix $\mathcal{H}$, the results in (6) provide some insight into the shape of the *distribution* of performances versus the delay d: Due to the Sylvester structure in $\mathcal{H}$, its columns are such that $CRB(s_n) > ... > CRB(s_{n-M}) = ... = CRB(s_{n-N+1}) < ... < CRB(s_{n-P+1})$, as long as $N > M$. This is a channel-independant result giving a simple guideline which consists in favoring the delays close to or, if possible, between $d = M$ and $d = N-1$ symbol durations in order to improve the estimation accuracy. The fact that extreme values for the delay (typically $d = 0$ or $d = P-1$), provide the worst noise enhancement properties was also confirmed in our simulations.

5 Implementation and Simulations

A possible implementation of the blind MMSE equalization technique presented in 3 is based on a straightforward coupled stochastic gradient descent of the criteria J_{mre}, under the bias removal constraint, and J_{mse}: The ZF equalizers estimates V_0,...,V_{P-1} are updated according to the constrained cost J_{mre}, the output of one of them V_j, is selected as a reference signal, the MMSE equalizer estimate W being updated according to the steepest descent of J_{mse}. Due to the lack of space the implementation details are omitted here. A short validation of such a technique is presented in the following communication context: $L = 2$ randomly chosen channels, with degree $M = 4$ (the channel coefficients are those given in [12]). The output SNR is 10dB and the symbols s_n are QPSK-modulated. We take $N = 5$ as the number of snapshots considered altogether. Delay optimization together with the condition $d - j \geq N$ naturally suggests $j = 0$ for ZF equalization delay, and $d = 5$ for the MMSE equalization delay (though, strictly speaking, a 4-delay MMSE equalizer would provide better results).

In fig. 1, we plot a typical learning curve for the output MSE between the transmitted and equalized data, using the estimated 5-delay MMSE equalizer, versus the iteration number. To check the result, the asymptotic MSE achieved by the *true* 5-delay MMSE equalizer is indicated in dashed line. The excess MSE provided by the adaptive equalizer is due to the adaptation noise and can be reduced by decreasing the stepsize in the gradient algorithms .

In fig. 2, we check the relevance of the theoretical results in section 3. The Cramer-Rao bounds provided by expression (6) are plotted for different delays, using '+' symbols. For comparison, the minimum achievable MSE with a linear equalizer is indicated, in 'o'. As expected, the lower bounds provided by the simple performance analysis in 3 are rather optimistic. However, the correspondance indicated between delay and performance is roughly verified, even assuming a sub-optimal linear structure for equalization.

6 Conclusion

In this contribution, it was shown that asymptotically ideal ZF adaptive equalizers could be obtained

through the minimization of the MRE criterion and used to train blindly an adaptive MMSE equalizer. In the presence of additive noise, the MRE criterion needs be modified to allow the derivation of unbiased equalizers. In the proposed algorithm, a non-zero delay can be chosen for the MMSE equalizer. A simplified Cramèr-Rao bounds analysis was used as a means to give practical guidelines for the choice of an optimal delay. The numerical simulations results match rather well with the theoretical derivations.

References

[1] L.Tong, G.Xu, and T.Kailath, "Blind identification and equalization based on second-order statistics: a time-domain approach," *IEEE Tr. on Information Theory*, vol. 40, no. 2 , pp. 340–349, March 1994.

[2] E. Moulines, P. Duhamel, J.F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. SP*, vol. 43, no. 2, pp. 516–526, February 1995.

[3] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind identification of multichannel FIR systems," in*Proc. ICASSP* , vol. 4, pp.581–584,1994.

[4] A. J. Vand der Veen, S. Talwar, and A. Paulraj, "Blind estimation of multiple digital signals transmitted over FIR channels," *IEEE S.P. Letters*, vol. 2, pp. 99–102, 1995.

[5] H. Liu and G. Xu, "A deterministic approach to blind symbol estimation," in *IEEE SP Letters*, vol 1, n 12, dec. 1994

[6] Y. Li and Z. Ding, "Blind channel identification based on second-order cyclostationary statistics," in *Proc. ICASSP*, vol. 4, pp. 81–84, 1993.

[7] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filter-banks and multichannel linear prediction," in *Proc. ICASSP*, vol. 4, pp. 585-588, 1994.

[8] K. Abed Meraim, P. Duhamel, D. Gesbert, P. Loubaton, S. Mayrargue, E. Moulines, and D. Slock, "Prediction error methods for time-domain blind identification of multichannel FIR filters," in *Proc. ICASSP*, 1995.

[9] D. Gesbert, P. Duhamel, and S. Mayrargue, "Subspace-based adaptive algorithms for the blind equalization of multichannel FIR filters," in *Proc. EUSIPCO*, 1994.

[10] G. Giannakis and S. Halford "Blind fractionally-spaced equalization of noisy FIR channels: adaptive and optimal solutions," *Proc. ICASSP*, pp. 1972–1975, 1995.

[11] D. Gesbert, P. Duhamel, and S. Mayrargue, "A bias removal technique for the prediction-based blind adaptive multichannel deconvolution," in *Proc. 29th Asilomar Conf. on Signals, Systems, and Computers*, 1995.

[12] D. Gesbert and P. Duhamel, "Robust blind joint data/channel estimation based on bilinear optimization," in *proc. 8th IEEE Signal Processing Workshop on SSAP*, 1996.

Appendix A

$R_X = E(X_n X_n^+)$, $R_{1X} = E(X_{n+1} X_n^+)$. Define Δ as the $P \times P$ diagonal matrix $diag(1, 2, 2, .., 2, 1)$, and J_P as the $P \times P$ matrix with ones on the diagonal above the main one, and zeroes elsewhere. Then, it is rather straightforward that

$$\mathcal{R} = \Delta \otimes R_X - J_P^+ \otimes R_{1X} - J_P \otimes R_{1X}^+$$

where $\otimes$ is the conventional Kronecker product. We have in a similar way:

$$\mathcal{R}_b = \sigma_b^2(\Delta \otimes I_P - J_P^+ \otimes R_{1\tilde{B}} - J_P \otimes R_{1\tilde{B}}^+)$$

where $\tilde{B}$ refers to the normalized (unit-variance) white noise. I_P is the identity matrix of size P. Let $Y_n = \mathcal{H} S_n$ (noise free observation). The signal-related matrix writes:

$$\mathcal{R}_s = \Delta \otimes R_Y - J_P^+ \otimes R_{1Y} - J_P \otimes R_{1Y}^+$$

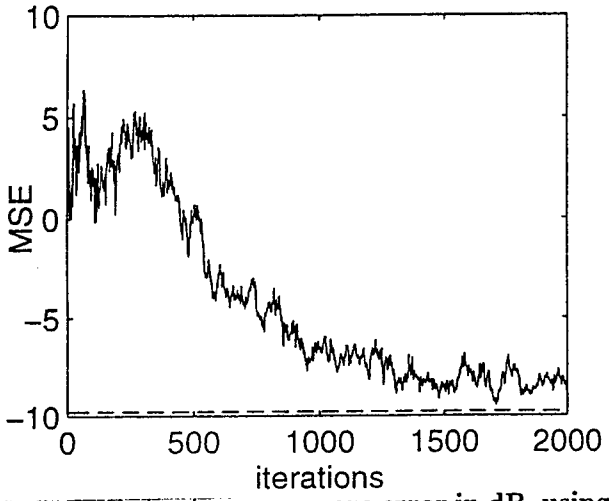


Figure 1: Output mean square error in dB, using the linear MMSE equalizer W by a gradient algorithm versus the iteration number.

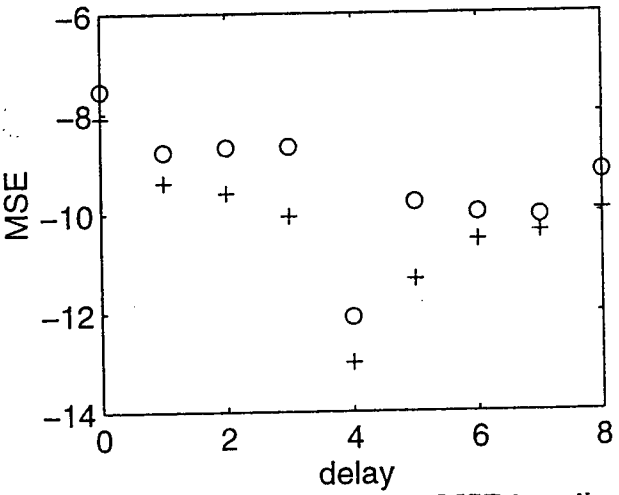


Figure 2: 'o' Minimum achievable MSE by a linear equalizer in dB, versus the delay. '+' Comparison with the derived CR bounds

BLIND IDENTIFICATION OF NONLINEAR CHANNELS EXCITED BY DISCRETE ALPHABET INPUTS

Michail K. Tsatsanis and Hakan A. Cirpan

Electrical Engr. & Computer Sc. Dept.
Stevens Institute of Technology
Castle Point on Hudson
Hoboken, NJ 07030

ABSTRACT

Hidden Markov Models (HMMs) are employed in this paper to describe digital communication channels, and their parameters are estimated in a blind fashion. General nonlinear channels can be accomodated which are not restricted to be of the Volterra type. Contrary to standard HMM parameter estimation techniques, which resort to nonlinear optimization of the likelihood function, the proposed method is based on a graph theoretic approach. We exploit the De Bruijn property of the channel's state transition graph, and develop computationally efficient blind estimation procedures involving shortest path searches. We show identifiability of the associated graph problem and discuss convergence issues. Finally, some illustrative simulations are presented.

1. INTRODUCTION

Most of the existing literature on channel estimation and equalization has focused on linear channels which can be described by an impulse response of finite length. However, in some applications the linearity assumption may not be valid, mainly due to nonlinear amplifiers in the transmitters (or repeaters), as for example in satellite channels [3], [5], [6].

Following the linear channel paradigm, a common method of describing nonlinear channels is by using truncated Volterra models [1]. Although Volterra series provide a general framework for treating nonlinear systems, they may not be perfectly suited for communication channels, as they do not take into account the finite alphabet nature of the input. Moreover, there is no clear indication of what the minimum Volterra order is, that would provide a satisfactory approximation.

In this paper we regard the channel as a general nonlinear mapping with no particular parametrizable form. The instrumental observation however, is that the input (and hence the channel state) can take a finite number of different values. Thus, the channel estimation problem is equivalent to identifying the mapping from each state to the corresponding channel output.

This approach to the channel estimation problem naturally leads to the theory of finite state machines and HMMs. In fact, general maximum likelihood (ML) techniques for blindly estimating the parameters of HMMs are well known [8], and have been applied in the context of communication channels [2], [4], [10]. These approaches however, suffer from increased computational complexity, and convergence problems related to the local minima of the likelihood function [8].

In this paper, we employ graph theoretic techniques in connection with clustering methods to avoid the likelihood maximization procedure. The proposed method is computationally efficient and a unique solution is guaranteed under some identifiability conditions.

2. PROBLEM STATEMENT

Let the received data $y(n)$, $n = 0, \ldots, N-1$ be generated by the communication system shown in Fig. 1, i.e.,

$$y(n) = h[\mathbf{w}(n)] + v(n) \tag{1}$$

where $\mathbf{w}(n) \triangleq [w(n), w(n-1), \ldots, w(n-q)]^T$ and the transmitted sequence $w(n)$ consists of i.i.d., equiprobable numbers taking values from a finite alphabet set $\mathcal{A} = \{a_1, a_2, \ldots a_\alpha\}$ of size α. $h[\cdot]$ is a linear or nonlinear channel of memory order q and $v(n)$ is zero mean, white, additive, Gaussian noise.

The channel $h[\cdot]$ does not have to obey a certain parametric form; however, it is not allowed to map distinct state vectors to identical outputs, as formalized in the following assumption:

(AS1) For every $\mathbf{w}_1 \neq \mathbf{w}_2$, the channel response is $h[\mathbf{w}_1] \neq h[\mathbf{w}_2]$.

Under this assumption, the goal of this paper is to identify the channel mapping $h[\mathbf{w}]$ for every possible state $\mathbf{w} \in \mathcal{A}^{q+1}$. Once the identification step is completed, a ML input estimation procedure can be used (Viterbi algorithm) to recover the input.

The proposed identification procedure consists of two steps: A clustering algorithm is employed first to estimate the α^{q+1} different values the (noiseless) channel output $h[\cdot]$ can take. Then, graph-theoretic techniques are developed to associate each of the cluster centers with the appropriate state vector $\mathbf{w} \in \mathcal{A}^{q+1}$.

It should be pointed out at this point that the channel is uniquely identifiable only up to a permutation of the input alphabet values. For example, in the BPSK case $\mathcal{A} = \{-1, +1\}$ which is indistinguishable from $\mathcal{A} = \{+1, -1\}$ with appropriately permuted response $h[\cdot]$. In linear channels, this inherent ambiguity manifests itself as a scaling ambiguity.

3. CLUSTERING

Clustering techniques have been employed in descriptions of communication channels when training data are available [9]. They are used to provide the channel's (noiseless) response, associated with each HMM state. In the case of negligible additive noise, the clustering step becomes trivial and can be solved by inspection of the received data $y(n)$.

0-8186-7576-4/96 $5.00

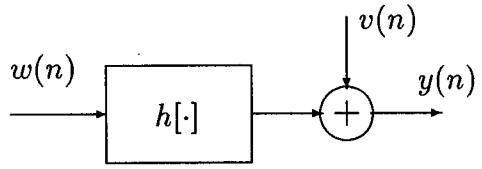


Figure 1. Nonlinear Communication Channel

The most commonly used clustering method is the K-means algorithm (e.g., [11]), which adjusts the cluster centroids $\hat{h}_k$, $k = 1, \ldots, \alpha^{q+1}$ and associates each data point $y(n)$ with a cluster membership set I_k, such that the following minimization problem is solved:

$$\min_{I_k, \hat{h}_k} \sum_{k=1}^{\alpha^{q+1}} \sum_{y(n) \in I_k} |y(n) - \hat{h}_k|^2 \, . \qquad (2)$$

The algorithm proceeds in an alternating fashion, optimizing in turn the centroids $\hat{h}_k$ and the cluster assignments I_k.

4. IDENTIFYING THE MARKOV MODEL

Once the centroids and cluster assignments have been determined by the K-means algorithm, the state transition probabilities need to be estimated to complete the Markov Model description of the channel. For general HMMs, the transition probabilities are typically estimated by maximizing the likelihood of the observed sequence [8]. In the current problem however, we should exploit the a priori information that is available about the channel and avoid the costly maximization step. For every channel state there are only α possible transitions, since the new input $w(n+1)$ can only take α values. Moreover, these transitions are equiprobable.

Let $I(y(n)) \in \{1, \ldots, \alpha^{q+1}\}$ be the cluster membership function obtained from the K-means algorithm (i.e., the nearest neighbor cluster assignment for each data point $y(n)$). We propose to replace the sequence $y(n)$ by $I(y(n))$ and identify the transitions from a state k_0 by recording all the transitions form k_0 in that sequence. In the absence of noise, only the allowable transitions will be present, while in the noisy case, the α most frequently recorded transitions will correspond to the allowable ones. Indeed, if the clusters are separated so that the probability of misclassification in $I(y(n))$ is small (see also (AS1)), then the α most frequently recorded transitions should be distinguishable from the occasional spurious transitions. This simple procedure seems able to complete the graph describing the HMM in a computationally efficient way. However, it cannot provide the association between each state $k \in \{1, \ldots, \alpha^{q+1}\}$ and the corresponding α-ary vector $\mathbf{w}_k \in \mathcal{A}^{q+1}$. In other words, although the noiseless channel outputs h_k have been estimated for every state k, and the transition graph has been completed, the channel mapping $h[\mathbf{w}_k]$ has not been identified yet, because the correspondence $k \leftrightarrow \mathbf{w}_k$ has not been obtained. This association is crucial in using the HMM to decode the input and needs to be recovered.

The problem can be equivalently posed as labeling each state $k \in \{1, \ldots, \alpha^{q+1}\}$, using a $q+1$ length vector $\mathbf{w}_k = [w_{k_1}, w_{k_2}, \ldots, w_{k_{q+1}}]$ of α-ary symbols $w_{k_i} \in \mathcal{A}$, $i = 1, \ldots q+1$, in a way that is consistent with the channel operatrions. The main contribution of our work is in employing graph-theoretic tools and developing an algorithm to solve this association problem.

5. IDENTIFIABILITY

It is clear that for a general finite state model, there is a large number of different labelings possible, and further information is required to complete that task. HMMs describing communication channels however, are of a special form and admit a unique labeling (under certain inherent ambiguities) as shown next. The key observation is that for a communication channel graph, a state transition from $\mathbf{w}_k = [w_{k_1}, w_{k_2}, \ldots, w_{k_{q+1}}]$ to $\mathbf{w}_l$ is valid only if $\mathbf{w}_l = [w_{k_2}, \ldots, w_{k_{q+1}}, \tilde{w}]$, $\tilde{w} \in \mathcal{A}$. In other words, the channel acts as a shift register, at each transition shifting $w_{k_2}, \ldots, w_{k_{q+1}}$ and incorporating the new data point $\tilde{w}$.

Graphs describing such systems are called De Bruijn graphs [7], and have been extensively studied in the area of computer science. They find applications in many diverse problems from coding theory to routing in computer networks. However, to the best of our knowledge, they have not been studied from an identifiability/labeling viewpoint. The main identifiability result developed in this paper is summarized in the following proposition.

Proposition: Every De Bruijn graph admits a unique labeling of its states, within a permutation of the alphabet letters. □

The proof is constructive and actually provides an algorithm to implement the association (labeling) procedure. Before developing the proof, it will be useful to note that every De Bruijn graph has exactly α nodes with self-loops. These correspond to the states $\mathbf{w}_{a_1} = [w_{a_1}, w_{a_1}, \ldots, w_{a_1}]$, $\mathbf{w}_{a_2} = [w_{a_2}, w_{a_2}, \ldots, w_{a_2}]$, ..., $\mathbf{w}_{a_\alpha} = [w_{a_\alpha}, w_{a_\alpha}, \ldots, w_{a_\alpha}]$, and can be identified as the α non-zero entries in the diagonal of the state transition matrix. Since permutations of those α labels simply corresponds to permutations of the input alphabet symbols, we can assume without loss of generality that $\mathbf{w}_{a_1}, \mathbf{w}_{a_2}, \ldots, \mathbf{w}_{a_\alpha}$ are given; we call those nodes the *roots* of the graph. Also, for every node $\mathbf{w}_k$, we call the nodes that are accessible with one transition, the ***children*** of $\mathbf{w}_k$. With this terminology established, the proposition's proof is based on the following lemma.

Lemma: If the following information is given about a De Bruijn graph:
i) the roots' labels,
ii) the label of an arbitrary node,
then the labels of that node's children are unique. □

Proof: Consider an arbitrary node $\mathbf{w}_k = [w_{k_1}, w_{k_2}, \ldots, w_{k_{q+1}}]$, and one of its children $\mathbf{w}_l = [w_{k_2}, \ldots, w_{k_{q+1}}, \tilde{w}]$. Let us assume that there are two different valid labelings for $\mathbf{w}_l$, one corresponding (without loss of generality) to $\tilde{w} = a_1$, and one to $\tilde{w} \neq a_1$. Let us compute the shortest path from $\mathbf{w}_l$ to the roots $\mathbf{w}_{a_1}, \mathbf{w}_{a_2}, \ldots, \mathbf{w}_{a_\alpha}$ (i.e., the minimum number of transitions or shifts required). If $\tilde{w} = a_1$, then clearly at most q shifts are needed to arrive at $\mathbf{w}_{a_1}$, while $q+1$ shifts are needed to arrive at any of the other roots $\mathbf{w}_{a_2}, \ldots, \mathbf{w}_{a_\alpha}$. If however $\tilde{w} \neq a_1$, $q+1$ shifts are needed to arrive at $\mathbf{w}_{a_1}$ which is a contradiction. Hence, both labelings cannot be valid. □

The proof of the lemma indicates a labeling procedure; namely starting by arbitrarily labeling the roots $\mathbf{w}_{a_1}$, $\mathbf{w}_{a_2}, \ldots, \mathbf{w}_{a_\alpha}$, and proceeding by systematically visiting the remaining nodes and labeling them according to the shortest path test. The details of the algorithm are explained

next.

6. PROPOSED ALGORITHM

The proposed algorithm consists of four steps, namely clustering, transition matrix estimation, labeling and decoding:

Step 1: (Clustering) Obtain the α^{q+1} cluster centers $\hat{h}_k$ and cluster membership function $I(y(n))$ from (2), using the K-means algorithm.

Step 2: (Transition matrix) For every state k_0, obtain its children as the α most frequent transitions from k_0 in the sequence $I(y(n))$.

Step 3: (Labeling)
Initialization:
(i) Mark all nodes as `not_visited`
(ii) Find the α non-zero entries in the diagonal of the transition matrix (roots), and label them arbitrarily as $\mathbf{w}_{a_1}, \mathbf{w}_{a_2}, \ldots, \mathbf{w}_{a_\alpha}$.
(iii) Put the roots in the `to_visit` queue.
Recursion:
while `to_visit` queue is non-empty,
(i) Get first node out of the queue
(ii) If `not_visited` then
(iia) Get its α children
(iib) Put them in the `to_visit` queue
(iic) Mark the node as `visited`
(iid) Compute the shortest path between each child and the α roots.
(iie) Label each child using the root corresponding to the smallest shortest path value.

Step 4: Employ the Viterbi algorithm to recover the transmitted symbols.

7. DISCUSSION

Some remarks on the proposed algorithm are now in order:

1. There are many computationally efficient algorithms to compute the shortest path from one node to every other node in the graph (e.g., Dijkstra's algorithm [7]). Their complexity is on the order of $QlogQ$, where $Q = \alpha^{q+1}$. Hence, the complexity of the proposed labeling scheme is $\alpha Q^2 logQ$.

2. The proposed method is suboptimal when compared with the maximum likelihood (ML) solution provided by the Baum-Welch algorithm. However, its computational requirements and convergence properties may be better. In medium to high SNR, clustering methods are known to perform satisfactorily, and may suffer less than ML methods from local minima problems. If further computational power is available, the results from the proposed method can be used as initial conditions for the Baum-Welch algorithm, to improve the estimation accuracy.

3. As a final remark, it should be noted that the ML methods suffer from the same identifiability problem (i.e., ambiguity with respect to permutations of the alphabet symbols), which is inherent to the blind problem at hand.

8. SIMULATIONS

In this section, some simulations results are presented in order to evaluate the performance of the proposed method. In all the simulations, the data were generated by filtering the transmitted sequence (N=800), through a linear channel of order $q = 2$ with $H(z) = 1+(1+0.5i)z^{-1}+(1-0.5i)z^{-2}$, and passing the output through a memoryless nonlinear model of a 'travelling wave tube amplifier' [6] which is employed in satellite communications. The model parameters used were the ones proposed in [6] (see Fig. 5 in [6]).

The transmitted and received BPSK symbols for the nonlinear channel are plotted respectively in Fig. 2a and 2b. Additive Gaussian noise of SNR = 12 dB was added. The output of the clustering algorithm is shown in Fig. 2c, where each data point is assigned to one of eight possible clusters (Original cluster centers, Estimated cluster centers and received data are depicted by o,x,· respectively). Finally, in Fig. 2d the output of the Viterbi algorithm is shown where only one error has occurred. Fig. 3 also depicts the performance of the Viterbi decoder by showing the original and estimated state sequence (Fig. 3a and 3b) as well as the state error sequence (Fig. 3c). Only one excursion from the correct state sequence was observed in this data sequence. Fig. 4 shows the magnitude of the estimation error,

$$\sum_k |h(\mathbf{w}_k) - \hat{h}_k|^2 \ . \qquad (3)$$

experienced by the B-W algorithm at each iteration. In Fig. 4b the results from the proposed method were used as initial conditions, while in Fig. 4a the initial conditions were arbitrarily chosen close to zero according to the suggestions of [10]. Notice the difference in the number of iterations needed for convergence when good initial conditions are provided. Results from 100 Monte Carlo runs (Probability of error versus SNR) of the proposed method (solid line) and the Baum-Welch algorithm (dashed line) with initial conditions from the proposed method are shown in Fig. 5. As expected, the Baum-Welch algorithm is slightly superior to the proposed method when properly initialized.

9. ACKNOWLEDGEMENTS

The authors wish to thank Profs. S. Banerjee, S. Bloom and A. Satyanarana for many valuable comments and fruitful discussions.

REFERENCES

[1] E. Biglieri, S. Barberis and M. Catena, "Analysis and compensation of nonlinearities in digital transmission systems", *IEEE Journal in Selected Areas in Communications*, vol. 6, no. 1, pp. 42-50, January 1988.

[2] J. A. R. Fonollosa and J. Vidal, "Application of hidden markov models to blind channel characterization and data detection", *Intl. Conf. on Acoustics, Speech and Signal Proc. (ICASSP-94)*, vol. IV, pp. 185-188, 1994.

[3] G. F. Herman, "Performance of Maximum-Likelihood Receiver in the Nonlinear Satellite Channel", *IEEE Transactions on Communications*, vol. COM-26, no. 3, pp. 373-378, March 1978.

[4] K. Kaleh and R. Vallet "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels", *IEEE Transactions on Communications*, vol. 42, no. 7, pp. 2406-2413, July 1994.

[5] A. R. Kaye, D. A. George, and M. J. Eric, "Analysis and Compensation of Bandpass Nonlinearities for Communications", *IEEE Transactions on Communications*,pp. 965-972, October 1972.

[6] A. A. M. Saleh, "Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers", *IEEE Transactions on Communications*, vol. COM-29, no. 11, pp. 1715-1720, November 1981.

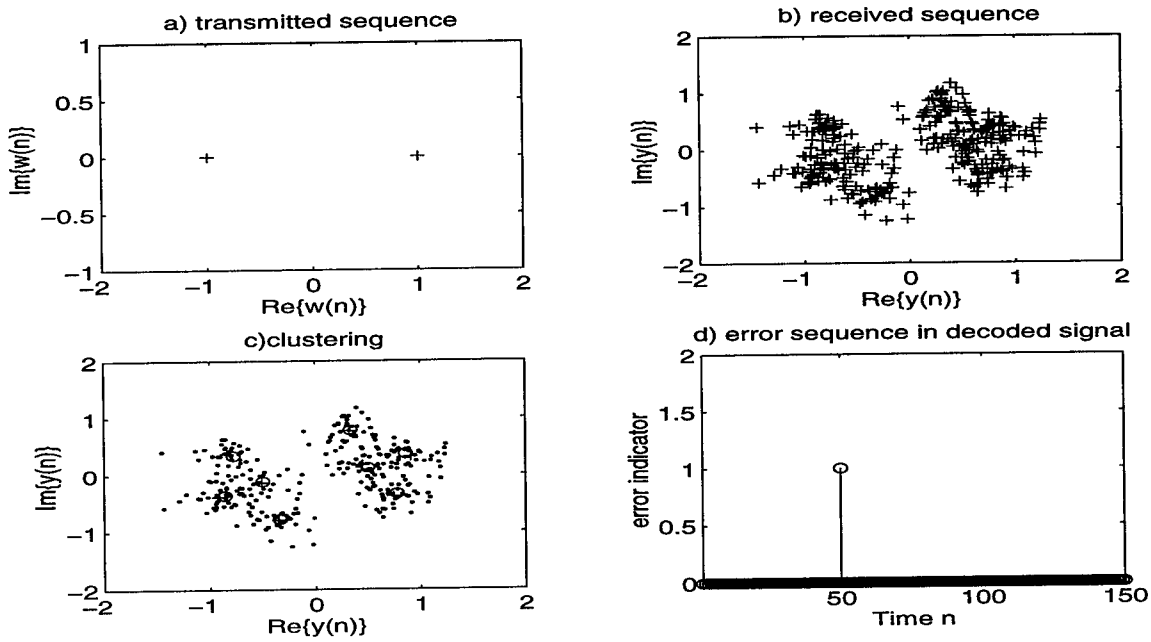


Figure 2. Received and Decoded Data Symbols

[7] M. N. S. Swamy and K. Thulasiraman *Graphs, Networks and Algorithms*, John Wiley & Sons, New York, 1981.

[8] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE*, vol 77, no. 2, pp. 257-274, February 1989.

[9] S. Theodoridis, C. M. S. See and C. F. N. Cowan, "Nonlinear channel equalization using clustering techniques", *Proc. of Intrl. Conf. on Communications (ICC-92)*, pp. 1277-1279, Chicago, 1992.

[10] M. Erkurt, J. G. Proakis, "Joint data detection and channel estimation for rapidly fading channels", *Proc. of IEEE Global Telecommunications Conference (Globecom'92)*, pp. 910-914, 1992.

[11] C. W. Therrien, *Decision, Estimation and Classification*, John Wiley and Sons, New York, 1989.

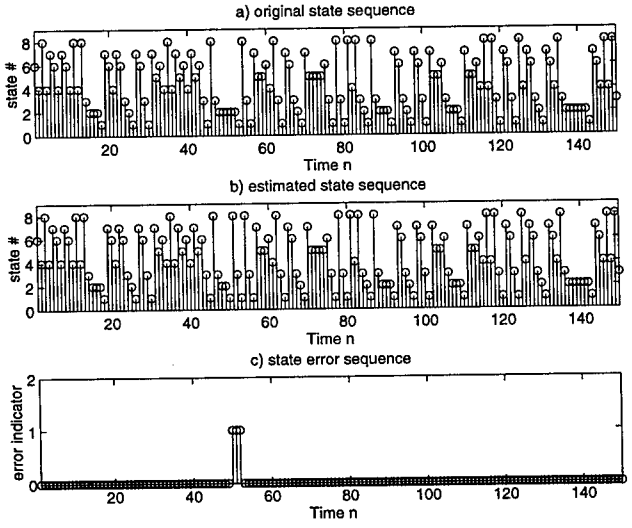


Figure 3. True and Estimated State Sequence

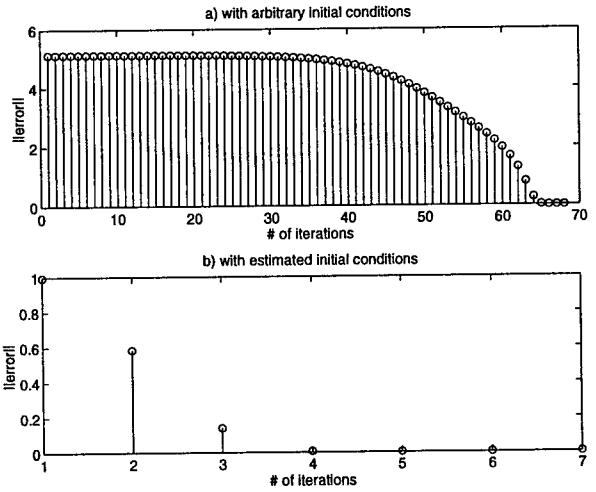


Figure 4. Convergence of Baum-welch Reestimation Algorithm

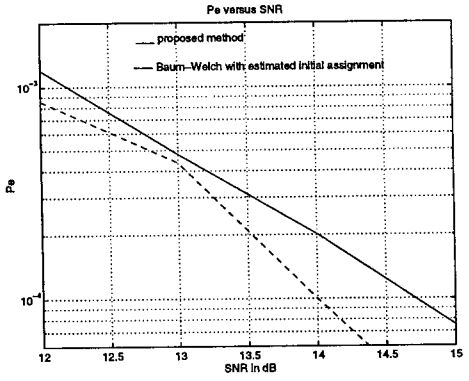


Figure 5. Performance Evaluation of Proposed Method

BLIND EQUALIZATION OF TIME-VARYING CHANNELS: A DETERMINISTIC MULTICHANNEL APPROACH

Georgios B. Giannakis
Dept. of Electrical Engr., Univ. of Virginia, Charlottesville, VA 22903-2442, USA

ABSTRACT

Blind equalization of rapidly changing multipath channels is important for mobile communications, and is addressed here by expanding the time-varying channel over a basis. Time-variation offers diversity and in contrast to time-invariant channels, blind equalization of time-varying channels becomes possible under mild conditions provided that the expansion components can be separated. Multichannel data are needed for the challenging non-separable case where it is shown that unique vector FIR equalizers exist under certain channel co-primeness conditions. Apart from persistence of excitation, no extra restriction is imposed on the input. Order selection methods and blind equalizers are derived directly from the output data. Preliminary simulations are also presented.

1. MOTIVATION AND MODELING

In wireless communications, multipath environments may change with time as the mobile communicators move. If the resulting time-varying (TV) channels exhibit variations which are too rapid for an adaptive algorithm to track, explicit modeling of the variation is well motivated.

Let the discrete-time data $x(n)$, $n = 0, \dots N-1$, be

$$x(n) = \sum_{l=0}^{L} \tilde{h}(n;l)s(n-l) \; + \; v(n)\,, \qquad (1)$$

where the inaccessible input (or source) $s(n)$ is allowed to be deterministic or random (white or colored) and independent of the AWGN $v(n)$. The TV impulse response depends explicitly on time n, and we model it using a basis expansion. Depending on whether the basis modulates the input or the channel, we model $\tilde{h}(n;l)$ respectively as (see also Figs. 1 and 2):

$$\mathbf{M1}: \quad \tilde{h}(n;l) \;=\; \sum_{q=1}^{Q} h_q(l)\, b_q(n-l)\,. \qquad (2)$$

$$\mathbf{M2}: \quad \tilde{h}(n;l) \;=\; \sum_{q=1}^{Q} \bar{h}_q(l)\, \bar{b}_q(n)\,. \qquad (3)$$

Expansion coefficients $h_q(\bar{h}_q)$ are time-invariant (TI), while the *known* basis sequences $b_q(\bar{b}_q)$ capture the time-variation and, depending on the application, are chosen *a priori* as e.g., polynomials, complex exponentials, or, wavelets. To allow for orders that vary with time or lag we define $L = \max_n L_n$ and $Q = \max_l Q_l$. Basis expansions for non-blind TV modeling were reported in [2].

Using single- or multi-channel output data only, we first recast the blind TV equalization as a blind TI source separation problem (Section 2). Structured time variations in (2) and (3) offer what we term *channel diversity* and degrees of freedom which render the blind TV equalization well-posed without use of higher-order statistics [8], or, restrictive assumptions on the input (e.g., whiteness [7]). In Section 3, we consider M-channel data $\mathbf{x}'(n) := [x^{(1)}(n) \dots x^{(M)}(n)]$ with each $x^{(m)}(n)$ obeying (1) and (2) of model **M1**. The resulting vector model is:

$$\mathbf{x}'(n) = \sum_{q=1}^{Q} \left[\sum_{l=0}^{L} \mathbf{h}'_q(l) b_q(n-l) s(n-l) \right] + \mathbf{v}'(n)\,, \qquad (4)$$

where prime stands for transpose, lower (upper) bold is used for vector (matrices), and the $M \times 1$ vectors $\mathbf{h}_q$ and $\mathbf{v}$ are defined similar to $\mathbf{x}$. We establish that $M \times 1$ FIR zero-forcing (or perfect in the absence of noise) equalizers, $\{\mathbf{g}_q^{(d)}(k)\}_{k=0}^{K}$, exist, so that within a delay $d \in [0, L+K]$ (which is non identifiable in the blind case) they yield:

$$\sum_{k=0}^{K} \mathbf{x}'(n-k)\, \mathbf{g}_q^{(d)}(k) \;=\; s_q(n-d)\,, \quad q = 1, \dots, Q\,, \qquad (5)$$

where $s_q(n-d) := b_q(n-d)s(n-d)$ denotes the deconvolved input modulated by the qth basis. Determination of Q and L are addressed, and linear equation based methods to estimate $\mathbf{g}_q^{(d)}(k)$ directly from output data are also developed in Section 3.

Model **M2** was justified based on the mobile kinematics in [8, 7] and can be related to **M1** via: $\bar{b}_q(n) = b_q(n-l)$, $\forall\, l \in [0, L]$. However, as **M2** is more general than **M1**, it requires separate treatment and direct blind equalizers are derived in Section 4 (see also [3] for alternative solutions).

The ideas herein are important generalizations of the TI results in [6, 4, 9, 5, 1] to TV channels. Relative to [7], the present approach allows for deterministic inputs, relaxes identifiability conditions, and achieves the same performance with less data.

2. CHANNEL DIVERSITY

Figs. 1 and 2 illustrate that the TV-SISO models of (2) and (3) can be viewed as TI-MIMO models, if the $\{\tilde{x}_q(n)\}_{q=1}^{Q}$ components can be obtained. For example, if in Fig. 2 the $\tilde{x}_q(n)$ components can be separated in the time-, frequency- or, cyclic-domain, the $x_q(n) = \bar{b}_q^{-1}(n)\tilde{x}_q(n)$ outputs can be recovered by demodulating with the known bases. Blind equalization can then be achieved using existing multichannel approaches (e.g., [4, 9]). Hence, time-variation (not necessarily that arising due to fractional sampling) offers

0-8186-7576-4/96 $5.00

diversity which renders blind equalization of TV channels easier than that of TI channels when separation is achievable. The latter is possible if for example $h_q(n)$ channels are low-pass and $b_q(n)$ are band-pass with center frequencies far apart from each other. Intuitively speaking, a TV channel offers us "different views for each time point n," and hence the term channel diversity.

Subsequently, we focus on the more challenging non-separable case which requires multichannel data $x^{(m)}(n)$, $m \in [1, M]$, $n \in [0, N-1]$. Multichannel data become available either with multiple antennas (see e.g., [4], [3]), or, by oversampling the continuous counterpart of (1).

To illustrate the latter, consider the (baseband) continuous-time data: $x_c(t) = \sum_l s(l) h_c(t; t - lT)$, where T is the symbol period. With oversampling rate M/T, we obtain: $x(n) := x_c(t)|_{t=nT/M} = \sum_l s(l) h(n; n - lM)$, and upon defining the sub-processes $x^{(m)}(n) := x(nM + m - 1)$, we find: $x^{(m)}(n) = \sum_l s(l) h^{(m)}(n; n - l)$ for $m = 1, \ldots, M$. Oversampling creates multiple channels but in contrast to the TI case, $x(n)$ is not necessarily cyclostationary and $x^{(m)}(n)$ is not necessarily stationary. Our results reveal that channel time variation (not necessarily periodic) may be sufficient to deal with blind problems.

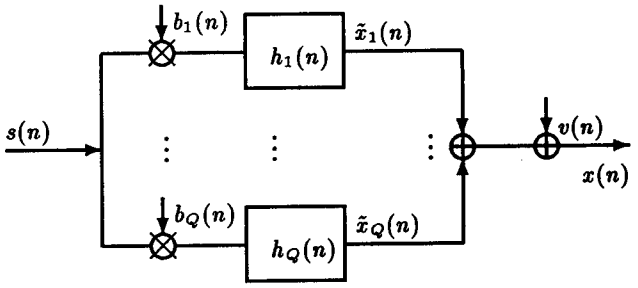


Figure 1. TV - SISO Channel Model M1

3. MODEL M1: BLIND EQUALIZATION

Let $\mathbf{s}'_q(n) := [b_q(n)s(n) \ldots b_q(n-L-K)s(n-L-K)]$ and define for each $q \in [1, Q]$ the $(L+K+1) \times M(K+1)$ block Toeplitz matrix

$$\mathbf{H}_q := \begin{bmatrix} \mathbf{h}'_q(0) & \ldots & \mathbf{0}' \\ \vdots & \ddots & \vdots \\ \mathbf{h}'_q(L) & \ldots & \mathbf{h}'_q(L-K) \\ \vdots & \ddots & \vdots \\ \mathbf{0}' & \ldots & \mathbf{h}'_q(L) \end{bmatrix} .$$

Consider (4) in the noise-free case and form the $(N-K) \times M(K+1)$ block Hankel data matrix

$$\mathbf{X} := \begin{bmatrix} \mathbf{x}'(N-1) & \ldots & \mathbf{x}'(N-1-K) \\ \vdots & \vdots & \vdots \\ \mathbf{x}'(K) & \ldots & \mathbf{x}'(0) \end{bmatrix} = \mathbf{S}_b \mathbf{H} , \quad (6)$$

where the $(N-K) \times Q(L+K+1)$ modulated input matrix $\mathbf{S}_b$ and the $Q(L+K+1) \times M(K+1)$ channel matrix $\mathbf{H}$ are given by

$$\mathbf{S}_b := \begin{bmatrix} \mathbf{s}'_1(N-1) & \ldots & \mathbf{s}'_Q(N-1) \\ \vdots & \vdots & \vdots \\ \mathbf{s}'_1(K) & \ldots & \mathbf{s}'_Q(K) \end{bmatrix} , \quad \mathbf{H} := \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_Q \end{bmatrix} .$$

We assume the following:
(M1.1) $N - K \geq M(K+1)$, which is satisfied by collecting sufficient data;
(M1.2.1)H is at least fat; i.e., (M, L, Q, K) obey

$$M(K+1) \geq Q(L+K+1) . \quad (7)$$

(M1.2.2)H is square; i.e., (7) holds as equality. To satisfy (7), a minimum $M_{min} = Q+1$ channels are required with a minimum equalizer order $K_{min} = QL - 1$ (in the TI case, $M_{min} = 2$ and $K_{min} = L - 1$ [6], [9]).
(M1.3)H is full rank; i.e., $\text{rank}(\mathbf{H}) = Q(L+K+1)$ which implies that transfer functions $\{H_q^{(m)}(z),\ q \in [1, Q],\ m \in [1, M]\}$ are co-prime. Note that $\{H_{q_1}^{(m)}(z)\}_{m=1}^{M}$ may have common zeros provided that for some $q_2 \neq q_1$ those are not also zeros of $H_{q_2}^{(m)}(z)$.
(M1.4)bases $b_q(n)$ are sufficiently varying and $s(n)$ is persistently exciting (p.e.) of sufficient order to assure that $\text{rank}(\mathbf{S}_b) = Q(L+K+1)$. If modulated inputs $\{s_q(n)\}_{q=1}^{Q}$ are linearly independent (sufficiently distinct modes), then $\mathbf{S}_b$ is full rank. Note that relative to the TI case, p.e. conditions on $s(n)$ are relaxed by the modulating bases. We stress that $s(n)$ can be either random or deterministic.

3.1. Order determination

Under **(M1.1)**-**(M1.4)**, matrix $\mathbf{X}$ is full rank $Q(L+K+1)$. With $\tilde{K}_1 > \tilde{K}_2$ denoting *known* upper bounds on K, corresponding matrices $\tilde{\mathbf{X}}_1$, $\tilde{\mathbf{X}}_2$, will have $\text{rank}(\tilde{\mathbf{X}}_i) = Q(L + \tilde{K}_i + 1)$, $i = 1, 2$. It is thus possible to select the orders L and Q using:

$$Q = \frac{\text{rank}(\tilde{\mathbf{X}}_1) - \text{rank}(\tilde{\mathbf{X}}_2)}{\tilde{K}_1 - \tilde{K}_2} , \quad L = \frac{\text{rank}(\tilde{\mathbf{X}}_1)}{Q} - (\tilde{K}_1 + 1) .$$

With Q, L available, K is chosen to satisfy (7) for a given $M \geq Q+1$.

3.2. Existence and uniqueness

Under **(M1.2.1)** and **(M1.3)**, we infer from $\mathbf{X} = \mathbf{S}_b\mathbf{H}$, that a unique linear FIR equalizer exists to yield $\mathbf{GX} = \mathbf{S}_b$. Matrix $\mathbf{G}$ is the pseudo-inverse $\mathbf{H}^\dagger$ which under **(M1.2.2)** becomes $\mathbf{H}^{-1}$. Because (7) is not satisfied with $M = 1$, it follows that blind separation and equalization of TV channels is impossible in the SISO case under **(M1.1)**-**(M1.4)**. The more channels ($M_{min} = Q+1$) required relative to the TI case is the price paid for our ability to equalize (and thus invert) FIR TV channels with linear FIR equalizers.

3.3. Direct blind equalizers

Blind equalizers exist and are unique but in order to find them we first set $n = N-1, \ldots K$ in (5) and collect equations to obtain

$$\mathbf{X}\, \boldsymbol{g}_q^{(d)} = \mathbf{s}_q^{(d)} := \tilde{\mathbf{B}}_q^{(d)}\, \mathbf{s}^{(d)} , \quad (8)$$

where $\boldsymbol{g}_q^{(d)} := [\mathbf{g}_q^{(d)'}(0) \ldots \mathbf{g}_q^{(d)'}(K)]$, $\mathbf{s}_q^{(d)} := [b_q(N-1-d)s(N-1-d) \ldots b_q(K-d)s(K-d)]'$, $\tilde{\mathbf{B}}_q^{(d)} := \text{diag}[b_q(N-1-d) \ldots b_q(K-d)]$, and $\mathbf{s}^{(d)} := [s(N-1-d) \ldots s(K-d)]'$. We use Matlab's notation $\mathbf{X}(i_1 : i_2, :)$ to denote a submatrix of $\mathbf{X}$ formed by the i_1 through i_2 rows and all columns of $\mathbf{X}$. Next we define

$$\mathbf{X}_{0,d} := \mathbf{X}(d+1 : N-K, :) , \quad \mathbf{X}_d := \mathbf{X}(1 : N-K-d, :) , \quad (9)$$

and $\mathbf{B}_q^{(0,d)} := \tilde{\mathbf{B}}_q^{(d)}(d+1 : N-K, d+1 : N-K)$, $\mathbf{B}_q^{(d)} := \tilde{\mathbf{B}}_q^{(d)}(1 : N-K-d, 1 : N-K-d)$.

From (8) and (9) it follows that

$$\begin{aligned} \mathbf{X}_{0,d}\,\boldsymbol{g}_{q1}^{(0)} &= \mathbf{B}_{q1}^{(0,d)}\,\mathrm{s}^{(0)}(d+1:N-K)\,, \\ \mathbf{X}_{d}\,\boldsymbol{g}_{q2}^{(d)} &= \mathbf{B}_{q2}^{(d)}\;\mathrm{s}^{(d)}(1:N-K-d)\,. \end{aligned} \qquad (10)$$

We note that $\mathrm{s}^{(0)}(d+1:N-K) = \mathrm{s}^{(d)}(1:N-K-d)$, which allows us to eliminate the input dependence from the equations in (10) and obtain the cross-relation:

$$\mathbf{B}_{q2}^{(d)}\,\mathbf{X}_{0,d}\,\boldsymbol{g}_{q1}^{(0)} \;=\; \mathbf{B}_{q1}^{(0,d)}\,\mathbf{X}_{d}\,\boldsymbol{g}_{q2}^{(d)}\,. \qquad (11)$$

The pair of equalizers $(\boldsymbol{g}_{q1}^{(0)}, \boldsymbol{g}_{q2}^{(d)})$ will be identifiable (up to a scale) as the eigenvector corresponding to the minimum eigenvalue of $\mathcal{X}_{q1,q2}^{(0,d)}$ in

$$\mathcal{X}_{q1,q2}^{(0,d)}\boldsymbol{g}_{q1,q2}^{(0,d)} \;:=\; \left[\mathbf{B}_{q2}^{(d)}\mathbf{X}_{0,d} \;-\mathbf{B}_{q1}^{(0,d)}\mathbf{X}_{d}\right]\begin{bmatrix} \boldsymbol{g}_{q1}^{(0)} \\ \boldsymbol{g}_{q2}^{(d)} \end{bmatrix} = \mathbf{0}\,, \qquad (12)$$

provided that the nullity $\nu(\mathcal{X}_{q1,q2}^{(0,d)}) = 1$. It turns out that under **(M1.2.2)**, **(M1.3)**, the latter holds for the minimum- and maximum-delay equalizers corresponding to $(0,d) = (0, L+K)$, provided that $\mathrm{rank}(\mathbf{S}_b) = 2Q(L+K+1)$.

Under **(M1.1)**-**(M1.4)**, equalizers corresponding to all possible delays $d \in [0, L+K]$ and $q_1, q_2 \in [1, Q]$ can be found simultaneously; e.g., with $d = L+K$, $q_1 = 1$, and $q_2 = 1 \ldots Q$, we obtain:

$$\begin{bmatrix} \mathbf{B}_1^{(L+K)}\mathbf{X}_{0,L+K} & -\mathbf{B}_1^{(0,L+K)}\mathbf{X}_{L+K} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{B}_Q^{(L+K)}\mathbf{X}_{0,L+K} & \mathbf{0} & \cdots & -\mathbf{B}_1^{(0,L+K)}\mathbf{X}_{L+K} \end{bmatrix} \times \begin{bmatrix} \boldsymbol{g}_1^{(0)} \\ \vdots \\ \boldsymbol{g}_Q^{(L+K)} \end{bmatrix} = \mathbf{0}\,. \qquad (13)$$

To recover $s(n-d)$ from the $\boldsymbol{g}_q^{(d)}$ equalizer's output, we simply demodulate with $b_q^{-1}(n-d)$ to obtain $b_q^{-1}(n-d)s_q(n-d) = s(n-d)$. For each n, it suffices to have $b_{q0}(n) \neq 0$ for at least one $q_0 \in [1, Q]$. If more than one basis are non-zero, we may align and average the corresponding equalizer outputs.

With the input available, one may readily obtain channel estimates (if so desired) by solving (4) using standard regression techniques (see e.g., [2]).

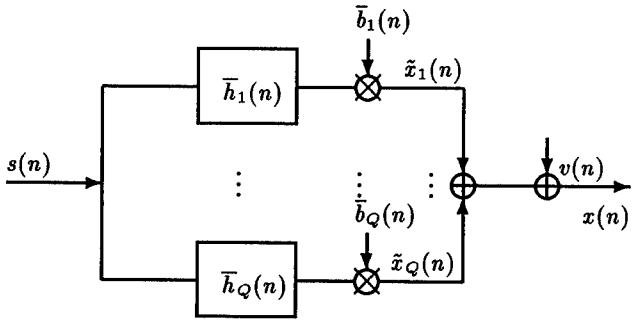


Figure 2. TV - SISO Channel Model M2

4. MODEL M2: BLIND EQUALIZATION

The counterpart of (4) for the model **M2** of (3) is given by (see also Fig. 2)

$$\mathbf{x}'(n) = \sum_{q=1}^{Q}\left[\sum_{l=0}^{L}\bar{\mathbf{h}}_q'(l)\bar{b}_q(n)s(n-l)\right] + \bar{\mathbf{v}}'(n)\,, \qquad (14)$$

but instead of (5), we form the $N \times M$ matrix $\bar{\mathbf{X}} = \bar{\mathbf{S}}_b\bar{\mathbf{H}}$, as follows [3]:

$$\begin{bmatrix} \mathbf{x}'(N-1) \\ \vdots \\ \mathbf{x}'(0) \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{s}}_1'(N-1) & \cdots & \bar{\mathbf{s}}_Q'(N-1) \\ \vdots & \vdots & \vdots \\ \bar{\mathbf{s}}_1'(0) & \cdots & \bar{\mathbf{s}}_Q'(0) \end{bmatrix}\begin{bmatrix} \bar{\mathbf{H}}_1 \\ \vdots \\ \bar{\mathbf{H}}_Q \end{bmatrix}. \qquad (15)$$

The entries in the $N \times Q(L+1)$ input matrix $\bar{\mathbf{S}}_b$ and the $Q(L+1) \times M$ channel matrix $\bar{\mathbf{H}}$ are given by:

$$\bar{\mathbf{s}}_q(n) := \bar{b}_q(n)\begin{bmatrix} s(n) \\ \vdots \\ s(n-L) \end{bmatrix}, \quad \bar{\mathbf{H}}_q := \begin{bmatrix} \bar{\mathbf{h}}_q'(0) \\ \vdots \\ \bar{\mathbf{h}}_q'(L) \end{bmatrix}.$$

Assumptions (M1.1)-(M1.4) are replaced by:
(M2.1) $N \geq M$;
(M2.2) triplet (M, L, Q) obeys:

$$M \;\geq\; Q\,(L+1)\,, \qquad (16)$$

which compared to (7) requires more channels;
(M2.3) $\mathrm{rank}(\bar{\mathbf{H}}) = Q(L+1)$;
(M2.4) $\mathrm{rank}(\bar{\mathbf{S}}_b) = Q(L+1)$.

Under **(M2.1)**-**(M2.4)**, we have $\mathrm{rank}(\bar{\mathbf{X}}) = Q(L+1)$. If $\bar{L}$ is a known upper bound on L, then order L can be obtained as $[\mathrm{rank}(\bar{\mathbf{X}})/Q] - 1$, but Q must be known.

With $\bar{\mathbf{G}} = \bar{\mathbf{H}}^{\dagger}$, **(M2.1)**-**(M2.3)** guarantee that unique linear FIR equalizers exist to recover $\bar{\mathbf{S}}_b$ from (15). If (16) holds as equality, then $\bar{\mathbf{G}} = \bar{\mathbf{H}}^{-1}$.

To derive blind equalizers directly from the data $\bar{\mathbf{x}}(n)$, we follow the notation of Section 3 and start from the zero-forcing condition: $\mathbf{x}'(n)\bar{\mathbf{g}}_q^{(d)} = \bar{s}_q(n) := \bar{b}_q(n)s(n-d)$, which for $n = N-1, \ldots, 0$, leads to the matrix form

$$\bar{\mathbf{X}}\,\bar{\mathbf{g}}_q^{(d)} \;=\; \bar{\mathbf{s}}_q^{(d)} \;:=\; \bar{\mathbf{B}}_q\,\bar{\mathbf{s}}^{(d)}\,, \qquad (17)$$

where $\bar{\mathbf{B}}_q := \mathrm{diag}[b_q(N-1)\ldots b_q(0)]$ and $\bar{\mathbf{s}}^{(d)} := [s(N-1-d)\ldots s(-d)]'$. When compared to (5), the equalizers for model **M2** have order $K = 0$. Defining $\bar{\mathbf{X}}_{0,d}$ and $\bar{\mathbf{X}}_d$ as in (9) with $K = 0$, and $\bar{\mathbf{B}}_q^{(0,d)}$, $\bar{\mathbf{B}}_q^{(d)}$ accordingly, we can eliminate $\bar{\mathbf{s}}^{(d)}$ from (17) and arrive at the cross-relation:

$$\bar{\mathbf{B}}_{q2}^{(d)}\,\bar{\mathbf{X}}_{0,d}\,\bar{\mathbf{g}}_{q1}^{(0)} \;=\; \bar{\mathbf{B}}_{q1}^{(0,d)}\,\bar{\mathbf{X}}_{d}\,\bar{\mathbf{g}}_{q2}^{(d)}\,. \qquad (18)$$

Based on (18) and selecting $M = Q(L+1)$ in (16), the $(0, L)$ pair or all equalizers can be recovered by solving for the minimum eigenvalues of equations similar to (12) or (13); see also [3] for $M > Q(L+1)$ with $q_1 = q_2$. In addition to (16), models **M1** and **M2** have different input matrices ($\mathbf{S}_b$ and $\bar{\mathbf{S}}_b$) with correspondingly different decompositions:

$$\begin{aligned} \mathbf{S}_b \;&:=\; \begin{bmatrix} \mathrm{s}'(N-1)[\mathbf{B}_1(N-1)\ldots\mathbf{B}_Q(N-1)] \\ \vdots \\ \mathrm{s}'(K)[\mathbf{B}_1(K)\ldots\mathbf{B}_Q(K)] \end{bmatrix} \\ \bar{\mathbf{S}}_b \;&:=\; \left[\bar{\mathbf{B}}_1\mathbf{S}\ldots\bar{\mathbf{B}}_Q\mathbf{S}\right]. \end{aligned}$$

Further research is required to characterize and compare p.e. requirements in terms of the spectrum of $s(n)$ and linear independence conditions among the basis sequences. Additional topics include development of adaptive algorithms, model validation studies, and especially for model **M2**, order and basis selection criteria.

Although noise is included in our simulations, the zero-forcing equalizers were derived in the noise-free case. Arguing as in [1], it follows that the minimum norm solution in (12), (13), or, (18), minimizes the noise power at the equalizer output. Future work will include noise explicitly using the linear prediction framework along the lines of [5].

5. SIMULATIONS

We generated $N = 300$ QPSK samples and with $Q = 2$ basis sequences, $b_1(n) = 1$, $b_2(n) = \exp(j2\pi n/50)$, we formed data $x(n)$ according to models **M1** and **M2**. Outputs of complex channels (order $L = 3$) were received by $M = Q + 1 = 3$ antennas for **M1**, and $M = Q(L+1) = 8$ for **M2**. One realization of the eye diagrams before and after equalization are shown in Fig. 3 at SNR=40dB for **M1**. Corresponding diagrams for **M2** at SNR=25dB are depicted in Fig. 4. The equalizer order for **M1** was $K = QL - 1 = 5$, and its coefficients were obtained by solving (12) with $q_1 = 1$, $q_2 = 2$, and $d = L + K = 8$. For **M2**, equalizer coefficients ($K = 0$) were found via (18) with $q_1 = 1$, $q_2 = 2$, and $d = L = 3$. To illustrate the importance of TV modeling, we show in Fig. 5 how the TI equalizers obtained from [9] perform on the data of Figs. 3 and 4. RMS performance of the errors $s(n) - \hat{s}(n)$ vs. SNR is plotted in Fig. 6 for **M1** and **M2** based on $N = 150$ samples and 100 Monte Carlo runs. **M2** was less sensitive to noise than **M1** which also appeared less robust to basis mismatch and p.e. conditions.

REFERENCES

[1] G. B. Giannakis and S. Halford, "Blind fractionally-spaced equalization of noisy FIR channels: adaptive and optimal solutions," *Proc. of Intl. Conf. on ASSP,* vol. 3, pp. 1972-1975, Detroit, MI, May 8-12, 1995.

[2] Y. Grenier, "Time-dependent ARMA modeling of nonstationary signals," *IEEE Trans. ASSP*, pp. 899-911, 1983.

[3] H. Liu, G. B. Giannakis, and M. K. Tsatsanis, "Time-varying system identification: A deterministic blind approach using antenna arrays," *Proc. of 30th Conf. on Info. Scien. & Sys.,* Princeton, NJ, March 1996.

[4] E. Moulines, P. Duhamel, J. Cardoso, S. Mayrargue, "Subspace Methods for the Blind Identification of Multichannel FIR Filters", *IEEE Trans. on SP*, p. 516, 1995.

[5] D. T. M. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," *Proc. of the Intl. Conference on ASSP,* vol. IV, pp. 585-588, Adelaide, Australia 1994.

[6] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE-IT*, p. 340, 1994.

[7] M. K. Tsatsanis and G. B. Giannakis, "Blind identification of time-varying channels using second-order statistics," *Proc. of 29th Asilomar Conf. on Signals, Systems, and Computers,* Pacific Grove, CA, Oct. 29-Nov. 1, 1995.

[8] M. K. Tsatsanis and G. B. Giannakis "Equalization of rapidly fading channels: Self recovering methods," *IEEE Trans. on Communications,* 1996 (to appear).

[9] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. on Signal Proc.*, pp. 2982-2993, 1995.

Acknowledgements: This research was supported by NSF Grant MIP 9424305. The author also wishes to thank Profs. Hui Liu and Michail K. Tsatsanis for discussions on this topic, and Cihan Tepedelenlioglu for his help with the simulations.

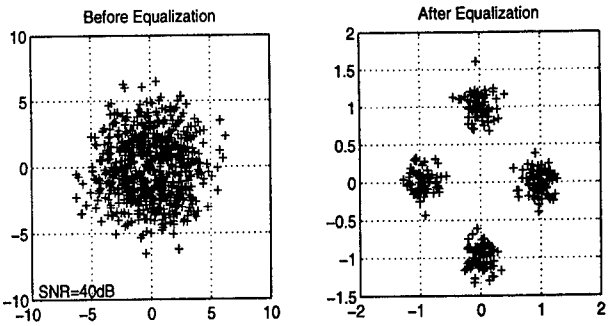


Figure 3. Eye diagrams for model M1

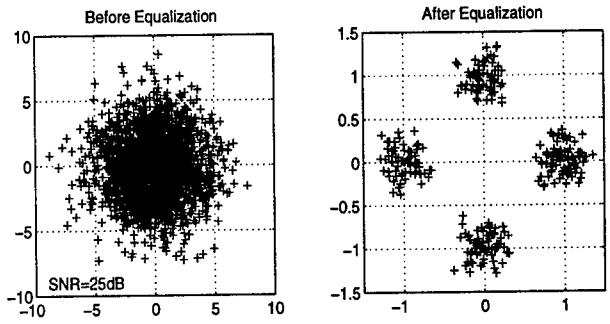


Figure 4. Eye diagrams for model M2

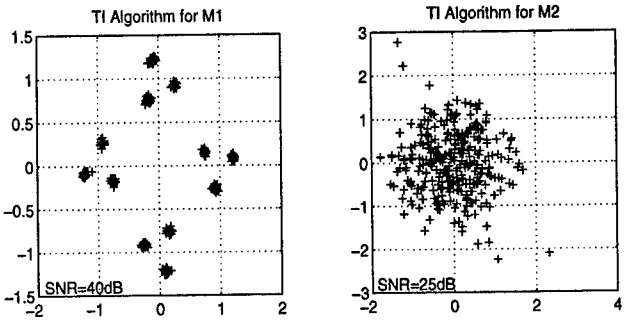


Figure 5. TI equalizers on TV models

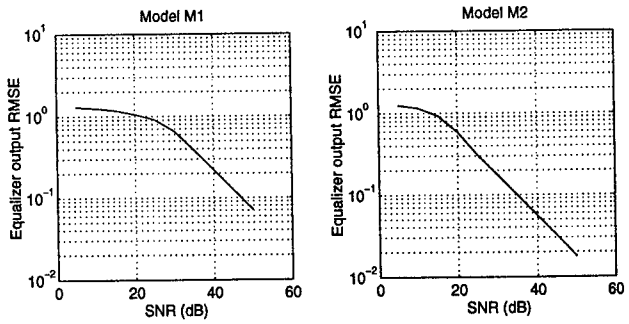


Figure 6. RMSE Performance of M1 and M2

MP-2: NonGaussian Signal Processing

Detection of Human Nerve Signals using Higher-Order Statistics

Barry Upshaw[1],Maria Rangoussi[2], Thomas Sinkjær[3]

[1,3]Center for Sensory-Motor Interaction
Aalborg University,
Aalborg DK-9220, Denmark
E-mail: bu@vision.auc.dk

[2]Dept. of Electrical Engineering
National Technical University of Athens,
Athens GR-15780, Greece
E-mail: maria@softlab.ntua.gr

Abstract

Afferent, whole nerve signals recorded using an implanted nerve-cuff electrode were analyzed using three detectors based on the 1st, 2nd and 3rd order statistical properties of the signals. Results based on standard Rectified, Bin-Integrated (1st order statistical) processing are compared with two algorithms based upon a Singular Value Decomposition (SVD) of the signal's 2nd and 3rd order correlation (cumulant) matrices. Due to the very low signal levels obtainable from nerve-cuff electrodes and the high levels of interference from adjacent muscles, the overall signal-to-noise ratio (SNR) is very poor. In addition, the noise level is non-stationary. The inherent properties of the 3rd order statistics of these signals yield a detector that performs better than the other two.

1. Introduction

It has been known for more than 100 years that animal muscle tissue can be made to contract through application of electrical current. More recently, this has been applied in the development of Functional Electrical Stimulation (FES) systems, with the goal of restoring lost motor function in paralyzed individuals. More than 30 years of FES development have lead to the now generally accepted conclusion that, in order to reduce muscle fatigue and increase reliability, closed-loop systems, in which some sort of "feedback" information is used to control the stimulator's parameters, yield better results than simple open-loop systems. In restoring muscle function via FES, the goal is to emulate, as best possible, the body's lost natural functionality. Given the choice of using artificial sensors (goniometers, strain-gauges, accelerometers, etc.), versus utilizing the subject's still intact sensory system, the latter is likely to provide us with the closest emulation of the body's natural control system. In order for the body's natural sensors to be used effectively, the level of information obtained from them should be comparable to that obtainable from artificial sensors. This requires a reliable, stable, implantable transducer which is able to record the sensory signals (known as "afferent" nerve signals) being passed along the body's nerve fibers, from local touch receptors, to the brain. The only appropriate such device presently suitable for use in humans (where nerve damage must be avoided) is the *nerve-cuff electrode*. Such cuffs are typically constructed from a silicone insulating tube, in which 3 non-insulated rings of stainless-steel or platinum wire act as electrodes. The cuff, which is slit longitudinally, is opened, placed around the nerve, and sutured closed. Lead wires connecting to the ring electrodes are routed to an appropriate exit site and through the skin, where they are attached to an external connector. For our purposes, these electrodes are connected to a special high-gain (110,000x), low-noise amplifier. The resulting amplified nerve signal is commonly referred to as the Electroneurogram (ENG).

We have constructed a prosthetic device utilizing this ENG signal (a "neuralprosthetic") in which a custom designed DSP-based system controls an 8-channel FES stimulator. The entire device is small enough to be easily born by the subject, and uses standard rechargeable batteries. Natural sensory information can be applied to a variety of FES tasks. We have primarily been concerned with two: Hand Grasp Restoration in Tetraplegics, and Hemiplegic Drop-foot Correction. Tetraplegic subjects, who have limited use of their arms, are typically unable to firmly grasp objects. Through stimulation of the muscles in the hand and forearm, simple grasp functions can be restored, using the processed nerve signal as a feedback signal indicating when, due to insufficient stimulation, the grasped object begins to "slip". Subjects suffering from a "drop-foot" are unable to fully activate the muscles which rotate the foot up/down. Thus, because they can not achieve adequate toe clearance, they are unable to walk normally. Stimulation of these muscles can improve such subject's gait, provided it occurs at the correct time in the gait cycle. Timing has, traditionally, been determined via a mechanical switch placed in the subject's shoe, which turns stimulation off upon closure (heel-contact) and on upon opening (heel-lift). We have previously shown that the nerve signal recorded by nerve-cuff electrodes can be used as a sort of "natural" heel-contact switch, [2]. In

0-8186-7576-4/96 $5.00

both applications, *the fundamental problem is the reliable detection of the presence of nerve signal activity in background noise*. Essentially then, the problem reduces to one of pure *endpoint* or transition detection in the drop-foot application.

2. Considerations Specific to this Problem

There are certain aspects of the present problem (in the use of human nerve signals) that complicate detection:

- The *noise* is some non-deterministic combination of tonic nerve firing, electrode thermal noise, and amplifier $1/f$ noise. Although, in the strictest sense, due to the presence of background (tonic) nerve firing, this isn't pure *noise*, in practice, it is dominated by the thermal and $1/f$ components of the electrodes and amplifier. In order to fully activate the paralyzed muscles using FES, it is often necessary to apply stimulation voltage pulses in excess of 140V to the skin's surface. These pulses (typically under 300msec in duration) propagate through the body (acting as a volume conductor) and induce large stimulation *artifact* impulses in the recorded nerve signal. Also, the Electromyographic (EMG) signal from adjacent muscles, either naturally occurring though voluntary activation, or stimulation induced, acts as a high level noise source. In addition, external EMF sources (typically mains power) are often of sufficient intensity to induce large noise potentials. The nerve signal amplitudes typically recorded are in the 1-10 μVolt range for common sensory stimuli. Therefore, the initial SNR of these raw nerve signals is often as low as -60dB! Fortunately, it is known that the majority of nerve signal information is confined to a narrow frequency band, from 1.0 to 3.0kHz. Therefore, an important first step in the detection process is the application of a simple (non-adaptive) bandpass filter. This filter, combined with other processing (windowing, adaptive thresholding, etc.) yields nerve signals with typical SNRs in the range from 0 to +3dB.

- The nerve signals recorded by cuff-electrodes are dominated by the activity from what are termed *fast adapting* sensory receptors. These receptors respond, primarily, to the 1st derivative (i.e. velocity) of applied force. Consequently, during a period of activity, defined by the application of a mechanical stimulus to the skin within the nerve's innervation area, *only the onset and offset of contact initiate detectably increased nerve activity*. Thus, activity occurs in short *bursts* where it is usually not possible to distinguish between force application and force removal. The practical implication of this fact for the use of afferent nerve activity in a drop-foot correction system, is a contact *onset/offset ambiguity* that must be resolved by other means.

- All methods we have tried to-date rely upon a single variable test against a fixed *threshold*. When the value of the processed ENG signal is below the threshold level, the *null* hypothesis H_0 is true, and the present *state* (gait phase) is unchanged. Upon exceeding the threshold, the *alternative hypothesis*, H_1, is indicated, and the present state is toggled (i.e., an *edge* occurred). Of particular significance is the constraint that the number of False Positives (FPs), or erroneous edge detections be, essentially, zero. The consequences of an FP are that the stimulator will be erroneously deactivated while the leg is still in motion, sufficient toe clearance will not be maintained, and the subject may fall. Thus, the detection threshold must be set sufficiently high such that the FP percentage is low. Conversely, if the threshold is too high, resulting in missed detections, the stimulator will not be turned off during the *stance* (standing) phase, the subject's muscles will tire rapidly and, again, the subject may fall. Thus, ideally, the processed ENG signal, as the input to the threshold detector, should have a very high SNR (i.e. the signal amplitude during transitions should be high, while the background level during constant force presence/absence should be close to zero). Given low SNR inputs (+3dB max.), and very non-stationary conditions (variable foot contact pressures, variable gait cycle timing, plus variable muscle and external EMF interference signals), the demands upon the signal processing algorithm for robust ENG processing are, indeed, strict!

- Finally, it is important to note that this is an unconditionally *real-time* processing application. Most ENG processing algorithms have, up until now, primarily been designed to characterize the properties of afferent nerve-cuff recordings *off-line*, and typically used inherently non-real-time methods, such as ensemble averaging, to enhance SNRs. When real-time information is desired, the standard processing method still widely used is to bin-integrate (over the inter-stimulation pulse interval) the rectified, filtered signal. Commonly referred to as the RBI (Rectified, Bin-Integrated) signal, this yields, essentially a standard lp-norm detector (or the *energy* over a window, if the squared signal is integrated), based on the signal's 1st order statistics. Unfortunately, while simple to implement (even with analog circuitry), energy detectors perform poorly on low SNR signals, with non-stationary noise. In order to improve detection reliability, specifically for the drop-foot application, an adaptive noise threshold was incorporated into the standard RBI algorithm, along with a *windowed* detector, [7]. Using these modifications, we obtained an average detection ratio of 85%, with no FPs. Since this was deemed unacceptable, we began investigating more robust detectors, in which a fundamental criterion is the ability to reject non-stationary, wide-band (essentially white) noise.

It has previously been shown, [4], [5], that good detection reliability is achievable using second- and higher-order statistics (HOS) on speech signals with

SNRs in the range mentioned above. This observation has prompted us to investigate the performance of detectors used for speech signals in the present problem. There are many similarities between the problems of detecting speech in noise and nerve-cuff signals in noise, indicating that similar methods may be applicable. However, one fundamental difference between speech and nerve signals is the onset/offset ambiguity issue mentioned above.

2.1 Autocorrelation-based detectors

The first, more advanced, detector investigated is based upon the signal's 2nd-order statistical properties. The method is based on the fact that the autocorrelation matrix **R** of a signal that contains only white noise is *diagonal*, with all diagonal entries equal to the variance of the noise, σ^2. All (say Q) eigenvalues of this matrix are, therefore, equal to σ^2, as well. If an information (non-white) component is also present in the signal, then **R** is no longer diagonal, and consequently its (real, positive) eigenvalues are not all equal. Testing for the presence of activity in the signal thus becomes equivalent to testing for (non)equality of the eigenvalues of **R**, *under the assumption that the additive noise is white*. Given that **R** can be estimated from a record of N samples through the observation matrix **X**, as $\mathbf{R} = \mathbf{X} \cdot \mathbf{X}^T$, the singular values of **X** can be used for the test. These are obtained using a Singular Value Decomposition (SVD). It has been shown that a computationally efficient method of solving the SVD problem, when the data is real-only, is the use of the Jacobi rotation algorithm, [3].

The actual test is performed by comparing the *difference* or the *ratio* of the maximum and the minimum eigenvalues, not to *zero* or *one*, respectively (as would ideally be the case), but to appropriately set *thresholds*. In theory, a significant advantage of this detection method over the RBI (or energy) method is that it is immune to the noise level (variance). This is because the white noise variance acts as a DC offset in the eigenvalue domain, which doesn't affect the eigenvalue difference. In practice, this detector is much more immune to non-stationary noise levels than the RBI detector, and yields better detection SNRs. Yet, since it primarily acts as a *whiteness versus non-whiteness* test, it is sensitive to the *color* of the noise. Note that in our case a significant proportion of the noise is due to the amplifier's colored ($1/f$) noise.

2.2 Cumulant-based detectors

In order to overcome this limitation, detectors based on the higher-order statistics (HOS) of the data were also tested. The 3rd-order statistics of a signal provide a measure of the *skewness* (difference from the Gaussian distribution) in the signal's statistical distribution, whereas the 2nd order statistics (autocorrelation and spectrum) only provide information about the signal's variance. Detectors based on 3rd-order cumulants have been successfully employed for speech signals due to the fact that *quadratic phase coupling*, present in voiced speech due to non-linearities in the vocal tract, [4], [1], can be detected using 3rd-order statistics. Although a precise model for the signals recorded by nerve-cuff electrodes has yet to be developed, it has been shown, [6], that these signals result in the non-linear combination of a series of *action potentials*, themselves modeled by a non-linear combination of sinusoidal functions. Thus it seems reasonable to assume that, in analogy with speech signals, there are significant (i.e. detectable) non-linearities in nerve-cuff electrode signals. In this case, it can be proven that the 3rd order cumulant of such signals cannot be zero for all lags. Thus a detector, using a method similar to that employed in the eigenvalue-spread algorithm, can be designed using only this diagonal vector as follows:

The 3rd order cumulant of a record of data, $x(n)$, is computed as: $c_{3x} = (1/N)\sum_n x(n)x(n+\tau_0)x(n+\tau_1)$ for an appropriate set of *lags* (τ_0, τ_1), lying on the main diagonal ($\tau_0 = \tau_1$) of the 2-D plane. This is, essentially, equivalent to computing the autocorrelation of $x(n)$ and $x^2(n)$. The $Q \times Q$ Toeplitz matrix $\mathbf{C}_3$ is formed from the first Q diagonal lags (where Q is chosen empirically) and its SVD is computed, as in the 2nd order case. In the 3rd order case, however, it is sufficient to simply use the maximum eigenvalue (rather than the difference between maximum and minimum) as the single test parameter. In this case, we are testing the matrix entries against zero as an indication of the presence of skewed components in the data (here, noise is assumed to be colored, but non-skewed). In practice, the maximum eigenvalue is compared against an empirically determined threshold.

The 3rd order method requires slightly more computations than the 2nd order case; yet, it is substantially less sensitive to additive (non-stationary) noise variance than either the RBI or 2nd order methods. This is important in a neuralprosthetic application where noise levels (and signal properties in general) vary not only amongst applications (i.e. the nerve used, its size, the size of the cuff electrode, etc.), but also amongst patients, and even with the time after implantation. Finally, the storage requirements of both the 2nd and 3rd order algorithms are well within the bounds of the *on-chip* memory of most commercial DSPs in contrast to most frequency domain (FFT or wavelet) methods, which generally require the addition of external memory. This is an important consideration for portable (or implantable) systems, where low power consumption is essential.

3. Results, Discussion and Conclusion

Figures 1 and 2 show a comparison of the 3 algorithms described, under non-stationary noise conditions. In Figure 1, linearly increasing white-noise (up to 100% of nominal) was added to a typical afferent nerve-cuff (ENG) signal in the region from 6000 to 10000 samples. The increased amplitude between samples 3000 and 5000 corresponds to increased nerve activity resulting from a single mechanical stimulation of the skin in the innervated area. This is also indicated by the arrow in Figure 2. The ordinate is in Volts. The nerve-cuff output signal was amplified by 220,000, filtered with a 4th order Butterworth bandpass (500Hz-3kHz) filter, and digitized to 12-bits (±5V range) using a sampling frequency of 10,000Hz.

Figure 2 shows detection results when the 3 detectors are applied on the noisy signal in Figure 1. Note that all three detect the true ENG activity (arrow), although the noise baseline, which defines the SNR of the detector (since the data is normalized to the peak value), is highest for the RBI detector and lowest for the cumulant detector. Thus the cumulant detector yields the highest SNR and the RBI detector the lowest, with the eigen-spread detector's SNR falling in between. As is evident in Figure 2, the SNR of the RBI detector decreases markedly with increased noise power. Both the eigen-spread and cumulant detectors continue to function at 100% added noise power.

In order for a natural sensory based device to be accepted in clinical applications, the amount of parameter adjustment required by the user (or physiotherapist) must be minimal. This has proven to be a severe drawback with RBI based detectors. Although we have obtained reasonable success by adding adaptive noise thresholding to the basic algorithm, we have not yet achieved a truly robust RBI implementation that does not require frequent parameter adjustments. Although it cannot be claimed that HOS offer the best solution for all types of signals, our preliminary results show that they hold great promise in the detection of afferent nerve signals in noise. Further improvements are anticipated through the use of (i) automatic thresholding based on a fixed, specified FP ratio, or (ii) a bi-frequency domain bi-coherence magnitude/phase detector, [1]. Further characterization of the statistical properties of nerve-cuff signals will be required to fully optimize future detection algorithms.

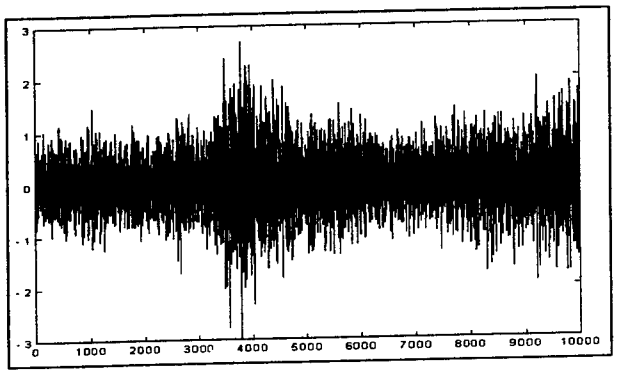


Figure 1. ENG signal plus white noise

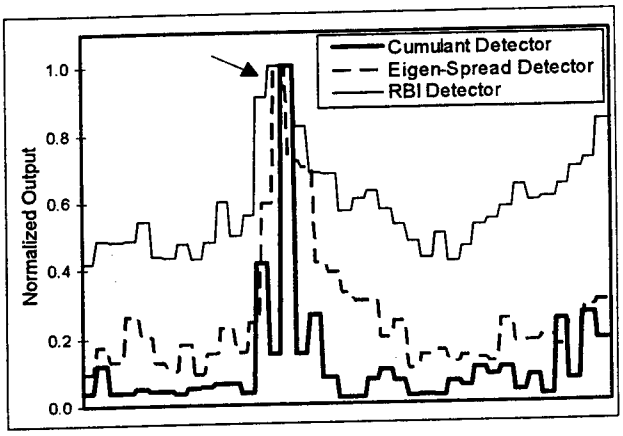


Figure 2. Results from the three detectors

References

[1] Fackrell J., McLaughlin S., "Detecting Phase Coupling in Speech Signals," *IEEE Colloquium Digest on Speech and Image Processing*, pp. 4/1-4/8 (1995).

[2] Haugland M., Hoffer J., Sinkjær T., "Skin Contact Force Information in Sensory Nerve Signals Recorded by Implanted Cuff Electrodes," *IEEE Transactions on Rehab. Engineering*, vol. 2, no. 1, pp. 18-28 (1994).

[3] Haykin S., "Adaptive Filter Theory," 2nd Edition, pp. 418-428, Prentice Hall (1991).

[4] Rangoussi M., Bakamidis S., Carayannis G., "On the use of SVD and HOS for robust endpoint detection of speech," in *Levels in Speech Comm.: relations and interactions*,Ed. C.Sorin, pp 267-279, Elsevier (1995).

[5] Rangoussi M., Carayannis G., "Adaptive Detection of Noisy Speech using Third-Order Statistics," *Intl. J. Adapt. Contr. & Sig. Proc.*, Special Issue on HOS, Wiley, (to appear, Dec. 1995).

[6] Stein R., Oguztöreli M., "The Radial Decline of Nerve Impulses in a Restricted Cylindrical Extracellular Space," *Biol. Cybernetics*, vol. 28, pp 159-165 (1978).

[7] Upshaw B., Sinkjær T., "Natural vs. Artificial Sensors Applied in Peroneal Nerve Stimulation", *Proc. of 5th Vienna International Workshop on Functional Electrostimulation*, pp. 239-242 (1995).

Unsupervised and Non Parametric Bayesian Classifier for HOS Speaker Independent HMM Based Isolated Word Speech Recognition Systems

M. Zribi*, S. Saoudi** and F. Ghorbel*
* Groupe de recherche Images et Formes
Ecole Nouvelle d'Ingénieurs en Communication
Institut National des Télécommunications
Cité Scientifique 59658 Villeneuve d'Ascq, France.
** Département Signal et Communication
Télécom Bretagne, Plouzané 29285 Brest cedex France.

Abstract

Here, we consider a speaker independent Hidden Markov Model (HMM) based isolated word speech recognition system. The most general representation of the probability density function (pdf), in the classical HMM, is a parametric one (i.e, a Gaussian). We intend here to derive an unsupervised, non parametric and multidimensional Bayesian classifier based on the well known orthogonal probability density function (pdf) estimator which does not assume any knowledge of the distribution of the conditional pdf's of each class. Such result becomes possible since this non parametric estimator is suitable and adapted to Expectation Maximization (EM) mixture identification algorithm.

Keywords : Unsupervised non parametric Bayesian classifier, orthogonal probability density function estimate, Expectation Maximization, Cepstrum coefficients, Line Spectrum Pairs, Speech recognition, Hidden Markov Model.

1. Introduction

Let us consider the isolated word speech recognition. For each word of the vocabulary, we want to design a separate M-state HMM. We represent the speech signal of a given word as a time sequence of spectral vectors (i.e, the Cepstrum or the Line Spectrum Pairs (LSP) coefficients). In a recent study [4], we proved that these two different kinds of acoustic analysis set of parameters provide a comparable recognition rate performance. In this paper, we focus our attention on the use of the LSP parameters instead of using the Cepstrum coefficients. The d LSP coefficients are computed with the antisymmetric form of the Split Levinson algorithm. This method is shown to be better, in terms of complexity, than other known algorithms [8], d is chosen to be equal to 10. Thus, for each vocabulary word, we have a training sequence consisting of observations of d-multivariate LSP. The first task is to build individual word models by adjusting the model parameters in order to maximize the likelihood of the observation sequence. The most general representation of the conditional pdf, for which an estimation procedure has been formulated is a Gaussian distribution [1,2,5]. The goal here is to make refinements on the pdf representation so as to improve the capability of modeling the spoken word sequences. When we want to design a speech recognition system, two fundamental procedures are generally required. Firstly, some feature descriptors are extracted from the observed speech signal. Secondly, the signal is labeled using a classification rule in the features space. Different classification algorithms are used for such problem. The statistical approach is recognized as efficient (Hidden Markov model, Bayesian classification rule,..). For automatic speech recognition, the unsupervised classifier is suitable since it is able to be adapted to the speaker. The best statistical classifiers are those based on the Bayesian classification rule since it minimizes the posterior probability of miss-classification which usually needs assumption on the pdf distributions. However, we verify experimentally that the conditional distributions with respect to a given class of the LSP are not close to a parametric one and change considerably according to the speaker, the pronounced word and so on... The common recognition mechanism is based on the HMM and makes use of diagonal Gaussian output distribution for each state. Therefore, it is well easy to show that the usual Gaussian hypothesis is not an efficient approximation. In this paper, we intend to present an efficient unsupervised Bayesian classifier without assumption on the distributions of the LSP coefficients. In this unsupervised context, the Bayesian classification rule which known by its optimality in the mean of the posterior probability of miss-classification criterion usually needs some parametric hypothesis for the conditional probability density function of each classes. Using the orthogonal probability density function estimate [10], the suggested classifier algorithm does not need any

0-8186-7576-4/96 $5.00

assumption on the distribution of the observed data. In this work, the proposed classifier is designed in two steps. The mixture identification is the first step. It consists on the estimation of the mixture parameters : the a priori probability and the conditional probability density functions of each class. It will be done with the Expectation Maximization algorithm (EM) [9]. The second step consists of the application of the Bayesian classification rule. The paper is organized as follows : In section 2, we give some elements of the isolated word Hidden Markov Model. In section 3, we recall the classical EM algorithm. Section 4 is devoted to the presentation of the suggested non parametric classifier.

2. Elements of the isolated word HMM

An isolated word HMM is built up of the following :
i) M, the number of states in the model. We denote S_i, the i^{th} state for i=1,.....,M. For our simulation M is equal to 10.
ii) The d-multivariate pdf $f_i(x)$ for each state S_i ($B=\{f_i(x)\}_{1\le i\le M}$). The observation is a continuous random variable, i.e, the Line Spectrum Pairs (LSP) coefficients.
iii) The state transition probability distribution $A=[a_{ij}]$ with $a_{ij}=Pro[\ q_{t+1}=S_j / \ q_t=S_i\ ]$ for $1 \le i,j \le M$, where q_t denote the state at time t. For our case, i.e, left-right model (see Figure 1), we have $a_{ij}=0$ for $j<i$ or $j>i+2$.
iv) The initial state distribution $\Pi=\{p_i\}$ for $1 \le i \le M$ and $p_i=Pro[\ q_1=S_i]$. For our case, i.e, left-right model, we choose $p_1=1$ and $p_i=0$ for i>1.
The complete specification of an HMM requires then specification of M (the number of states), the M-continuous d-multivariate pdf $B=\{f_i(x)\}_{1\le i\le M}$, the matrix transition $A=[a_{ij}]$, $1 \le i,j \le M$, and the initial state distribution Π. For convenience, we use $\lambda^{(i)}=(A, B, \Pi)$ to denote the HMM model for the i^{th} word. A block diagram of an isolated word HMM recognizer is given in Figure 2.

3. The classical EM algorithm

The classical EM assumes that the observed data is a realization of a mixture of parametric distributions, so that its pdf can be written as :

$$f(x)=\sum_{k=1}^{K}\pi_k f(x/\theta_k),\ with\ 0 \le \pi_k \le 1\ and\ \sum_{k=1}^{K}\pi_k = 1$$

where $f(x/\theta_k)$ is the conditional pdf of class k and π_k is the probability a priori of each class of a LSP vector. This algorithm is iterative and has three main steps. We propose to describe it here in the case of the Gaussian hypothesis (i.e : $\theta_k=(\mu_k, \Gamma_k)$ where μ_k is the mean vector and a Γ_k is the covariance matrix of the class k).

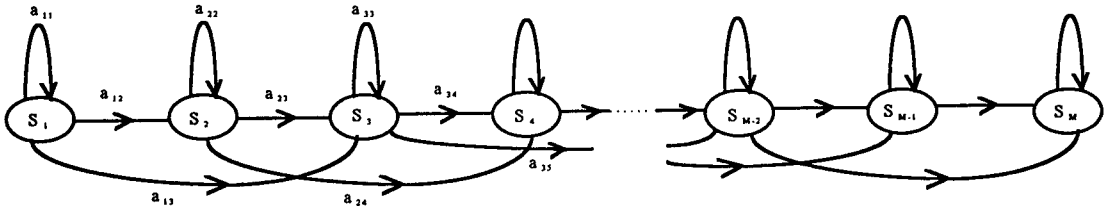


Figure 1. A M state left-right HMM model

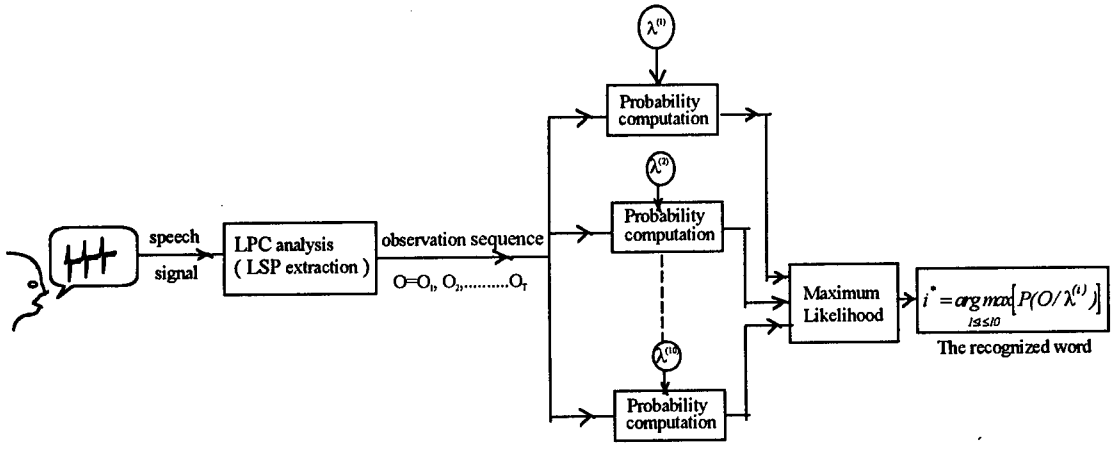


Figure 2. block diagram of an isolated word HMM recognizer ($\lambda^{(i)}$: the HMM model for the i^{th} word)

- *Initialization step* : We suppose the number of classes K is known and then an initial solution of the parameters of the mixture are extracted from the histogram.
- *Expectation step* : It consists on the estimation of the a posterior probability $\hat{\pi}_k^n(x_i)$ for the realization x_i belonging to the class k at the n[th] iteration :

$$\hat{\pi}_k^n(x_i) = \frac{\hat{\pi}_k^n f(x_i / \hat{\theta}_k^n)}{\sum_{j=1}^{K} \hat{\pi}_j^n f(x_i / \hat{\theta}_j^n)}$$

- *Maximization step* : We build here the parameters needed for the next step, in the follow way :

$$\hat{\pi}_k^{n+1} = \frac{1}{N}\sum_{i=1}^{N} \hat{\pi}_k^n(x_i),\ \hat{\mu}_k^{n+1} = \frac{\sum_{i=1}^{N} x_i \hat{\pi}_k^n(x_i)}{\sum_{i=1}^{N} \hat{\pi}_k^n(x_i)},$$

$$[\Gamma_k]^{n+1} = \frac{\sum_{i=1}^{N} (x_i - \hat{\mu}_k^{n+1})(x_i - \hat{\mu}_k^{n+1})^T \hat{\pi}_k^n(x_i)}{\sum_{i=1}^{N} \hat{\pi}_k^n(x_i)}$$

for $k=1,\ldots,K$.

4. The proposed non parametric EM

4.1. Estimation based on orthogonal expansions

The estimation of the pdf based on methods of Fourier analysis is suitable for this situation. Let X be a random vector taking values in the d-dimensional Euclidean space $\mathbb{R}^d$ and suppose that the distribution of X is described by a probability density function f. Given a sample $X_1,\ldots,X_N$ of N independent observations of X, the $\hat{f}_{K_N}$ estimator of f is the probability density function :

$$\hat{f}_{K_N}(x) = \sum_{j=1}^{d} \sum_{m_j=0}^{K_N^j} \hat{a}_{(m_1,\ldots,m_d)} e_{(m_1,\ldots,m_d)}(x) \text{ where}$$

$$\hat{a}_{(m_1,\ldots,m_d)} = \frac{1}{N}\sum_{i=1}^{N} e_{(m_1,\ldots,m_d)}(X_i) \text{ and}$$

$\left\{e_{(m_1,\ldots,m_d)}(x)\right\}_{(m_1,\ldots,m_d)\in \mathbb{N}^d}$ is a normal complete basis of $L_2(]a,b[^d)$.]a,b[is an interval of the real line. For simplicity we consider the same $K_N^j = K_N$ for all j=1,....,d. This assumption does not induce a bad behavior on the estimation of parameters since we use as orthogonal basis functions in the multivariate case the product one of the univariate basis.

4.2. Description of the non parametric EM

This approach do not assume a knowledge on the kind of the conditional pdf of LSP parameters, so that :

$$\theta_j = (a_{(0,\ldots,0),j},\ldots,a_{(K_{N_j},\ldots,K_{N_j}),j}) \text{ and } f(x/\hat{\theta}_j) = \hat{f}_{K_{N_j}}(x).$$

a. Initialization step : We suppose the number of classes K is known and then an initial solution of the parameters of the mixture are extracted from the histogram.
b. Expectation step : In this step, we estimate the a posterior probability $\hat{\pi}_k^n(x_i)$ for the LSP vector x_i belonging to the class k at the n[th] iteration by :

$$\hat{\pi}_k^n(x_i) = \frac{\hat{\pi}_k^n f(x_i / \hat{\theta}_k^n)}{\sum_{j=1}^{K} \hat{\pi}_j^n f(x_i / \hat{\theta}_j^n)}$$

c. Maximization step : The a posterior probability $\hat{\pi}_k^n(x_i)$ of each x_i is computed. So that, at (n+1)[th] iteration, we have :

$$\hat{\pi}_i^{n+1} = \frac{1}{N}\sum_{j=1}^{N} \hat{\pi}_i^n(x_j),$$

$$K_{\hat{N}_i^{n+1}} = int\left[(\hat{N}_i^{n+1})^{1/s}\right] \text{ where } \hat{N}_i^{n+1} = N\hat{\pi}_i^{n+1}$$

$$\hat{a}_{(m_1,\ldots,m_d),k}^{n+1} = \frac{\sum_{j=1}^{N} e_{(m_1,\ldots,m_d)}(x_j)\hat{\pi}_k^n(x_j)}{\sum_{j=1}^{N} \hat{\pi}_k^n(x_j)} \text{ for } m_j = 0,\ldots,K_{\hat{N}_k^{n+1}}$$

The Bayesian rule : After the mixture identification, the Bayesian rule are applied in order to classify the speech signal according to their LSP vector x_i :

$$k(x_i) = \text{Arg}\left[\max_{1\le k\le K}\left\{\pi_k f(x_i/\theta_k)\right\}\right]$$

where $k(x_i)$ represents the label of the class of the vector x_i.

For the database, the set of speech sequences is separated into two parts : one for training, the other for testing. The database contains 10 digits (0 to 9) pronounced by 25 speakers with 150 utterances in the training set and 100 in the test. The analog voice signal is digitized at 8 khz. The signal is multiplied by a 32 ms Hamming window. The LSP coefficients are computed every 16 ms. Finally, once the set of Word HMMs has been designed and optimized, recognition of an unknown word is performed by computing the probability of the observation sequence for each word model and select the word whose model score

is highest (i.e., the highest likelihood). As we have seen here, the non parametric aspect comes from the use of the orthogonal density estimates in the mixture identification step which is reduced to the estimation of the first Fourier coefficients of these densities.

5. Conclusions

We have considered a speaker independent HMM based isolated word speech recognition system. The speech signal has been represented as a time sequence of LSP coefficients. We have shown that the conditional distributions, with respect to a given state, are note close to a parametric one (i.e, a Gaussian). We have suggested an unsupervised and non parametric estimator based on orthogonal expansions to improve the pdf representation.

References

[1] S. Furui, " Digital speech processing synthesis and recognition", Marcel Dekker, NY, 1989.

[2] D. Jouvet, "Application des modèles de Markov à la reconnaissance de la parole", CNET Lannion France, 1995.

[3] E. Keller, " Fundamentals of speech synthesis and speech recognition", J. Wiley, NY, 1991.

[4] E. Martinez and S. Saoudi, " La reconnaissance de la parole HMM-LSP", Internal report, Oct 94, ENST-Br, France.

[5] Qiang Huo and Chrkin Chan,Contextual Vector Quantization for Speech Recognition with Discrete Hidden Markov Model, Pattern recognition, Vol. 28, n°4, pp 513-517 (1995) Pergamon.

[6] L.R. Rabiner, "A tutorial on Hidden Markov models and selected applications in speech recognition", Proc. IEEE, Vol. 77 N°2, Fev. 89, pp 257-286.

[7] S. Saoudi, F. Ghorbel and A. Hillion, Non parametric probability density function estimation on a bounded support; application to shape classification and speech coding, Applied Stochastic models and data analysis, Vol. 10 215-231 (1994) John Wiley & Sons, Ltd.

[8] S. Saoudi, J. M. Boucher and A. Le Guyader, A new efficient algorithm to compute the LSP parameters for speech coding, Signal processing journal, Vol. 28, N 2, August 92, pp 201-212.

[9] C. Soubiran and al, "analyse de mélange gaussiens pour de petits échantillons: application à la cinématique stellaire," Revue de statistique appliquée, Vol 39 n°3, pp 17-35, 1991.

[10] M. Zribi and F. Ghorbel, An unsupervised and Non parametric Bayesian image segmentation, 8th International Conference On Image Analysis and Processing, IAPR, Sanremo, Italy, September 1995.

Fast HOS Based Simultaneous Voiced/Unvoiced Detection and Pitch Estimation Using 3-Level Binary Speech Signals

Ali Alkulaibi, J.J.Soraghan, T.S.Durrani
Signal Processing Division
Dept. of Electronic and
Electrical Eng., University of Strathclyde
Glasgow G1 1XW, Scotland,UK.
ali@spd.eee.strath.ac.uk

Abstract

A new fast and robust HOS based algorithm for simultaneous voiced/unvoiced detection and pitch estimation using 3-level binary speech signals is presented. An accurate and reliable voiced/vnvoiced detection of a speech signal and associated pitch period estimation from the voiced part is made in coloured noise environments with low SNR. The use of the 3-level binary speech signals dramatically reduces the computational effort required in evaluating the third order cumulant. The superior performance of the new algorithm to the conventional autocorrelation method using real speech signals in low SNR environments is demonstrated.

1. Introduction

Accurate and reliable voiced/unvoiced detection of a speech signal and associated pitch period estimation for the voiced part are crucial preprocessing steps in many speech processing applications and are essential in most analysis and synthesis (vocoder) systems. These include automatic detection of the beginning and ending of an utterance in a long recording, speech segmentation and automatic isolated word recognition (AIWR) [1, 4, 7]. Many algorithms have been reported in the literature for solving the detection and estimation problem using second order statistics such as autocorrelation, cepstrum and average magnitude difference function (AMDF) [1, 4, 5]. A common problem with these second order statistics algorithms is that they are sensitive to various noises. Third order statistics have been shown to be particularly insensitive to various noises such as Gaussian and coloured, sinusoidal and car noise [6, 9]. HOS have been applied in [6] to speech signals for pitch determination using autocorrelation of the third order cumulants. In [7] HOS have been used for end point detection of a speech signal by using the maximum singular value of appropriate cumulant matrix. A voiced/unvoiced decision in the frequency domain using HOS has been reported in [8] that uses the bispectrum properties which approximate to zero for the fricative phonemes and a complex structure for the voiced phonemes. A main concern in using HOS in practice is the excessive computation involved in its estimation. In this paper we proposed a new fast and robust 3-level binary HOS based algorithm for simultaneous voiced/unvoiced detection and pitch estimation of speech signals that can work satisfactory in low SNR environments. In section 2 the new algorithm is described. In section 3 the simulation results using real speech signals are presented and its performance is compared to the conventional second order methods.

2. The Algorithm

The block diagram of the 3-level binary HOS based detection and estimation system is shown in Fig.1. The speech signal is segmented into overlapping 30 ms frames. The system uses center clipping and infinite peak clipping as a non linear spectrum flattening on the speech signal [2, 3]. For each frame a clipping threshold is computed as follows:

$$c_l = K min[c_{l_1}, c_{l_2}] \tag{1}$$

where from computer simulations an appropriate value for K is found to be .2, c_{l_1} and c_{l_2} are the maximum amplitude in the first and last third of the frame respectively. Thus a 3-level binary speech signal is produced by center clipping and infinite peak clipping the speech signal with values of $-1, 0, +1$ depending on the relation of the original speech sample to the clipping thresholds as follows:

$$x(n) = \begin{cases} 1 & if \quad s(n) \geq c_l \\ -1 & if \quad s(n) \leq -c_l \\ 0 & otherwise \end{cases} \tag{2}$$

0-8186-7576-4/96 $5.00

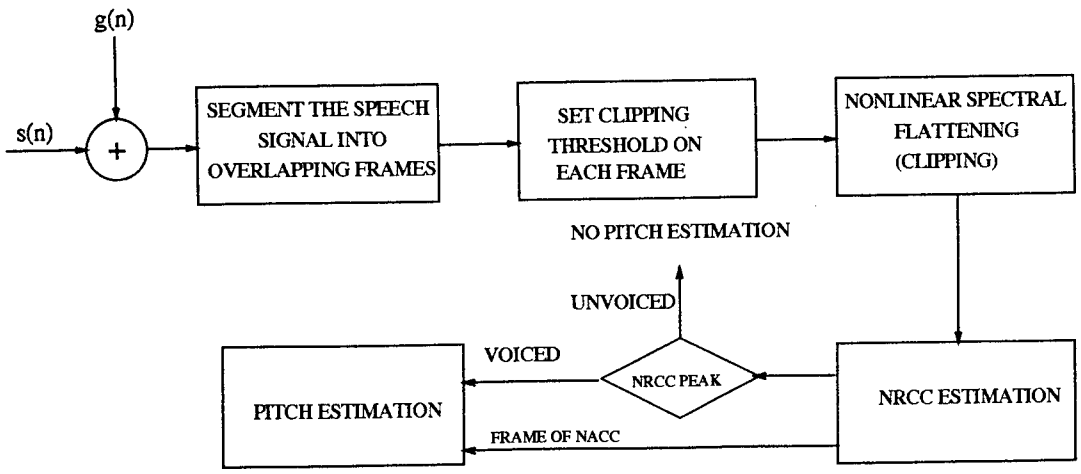


Figure 1. Block Diagram of the $NACC$ System

where $s(n)$ is the speech sample. Since the 1-d slice of the third order cumulants is defined as:

$$c_3(\tau) = E[x(n)x(n)x(n+\tau)] \tag{3}$$

each combination in Eq.(3) can assume the following 3-level binary values as:

$$x(n)x(n)x(n+\tau) = \begin{cases} 0 & if\{\ x(n)=0, \\ & or\ \ if\ \ x(n+\tau)=0\} \\ 1 & if\ \ x(n+\tau)=1 \\ -1 & if\ \ x(n+\tau)=-1 \end{cases} \tag{4}$$

Thus, a simple combinatorial logic circuit is only required in computing each term in the third order cumulant and an up-down counter to accumulate the actual third order cumulant value of Eq.(3). The 3-level binary HOS based detection and estimation system uses a normalized autocorrelation function of the 1-d slice of the third order cumulants $NACC$ defined as:

$$NACC(\tau) = \frac{\left[\sum_{n=0}^{N-1} c_3(n)c_3(n+\tau)\right]^2}{\sum_{n=0}^{N-1} c_3^2(n+\tau)} \tag{5}$$

The numerator and the denominator of Eq.(5) involves simple logical operation. To simultaneously detect the voiced/unvoiced region and the associated pitch period estimation for each frame the peak value of the $NACC$ is compared to a threshold as shown in Fig.1. If it is a voiced frame the pitch and its period are estimated directly from the positions where the $NACC$ has its maximum peaks.

3. Experimental Results

To demonstrate the performance of the 3-level binary HOS $NACC$ system for simultaneous voiced/unvoiced detection and pitch estimation, the utterances of 'six' is used where the utterance has three unvoiced/voiced regions. Additive coloured Gaussian noise of 5dB and 0dB SNR are used for the simulations as shown in Fig.2. For voiced/unvoiced detection the maximum peak of the $NACC$ in Eq.(5) is recorded for each frame as shown in (a) and (c) of Fig.3 respectively. From the figures it clear that a level close to zero signifies an unvoiced region while a significant value signifies a voiced region. From the voiced region the pitch is simultaneously estimated from the periodicity of the $NACC$ in Eq.(5) where for a voiced frame, the complete $NACC(\tau)$ from that frame is plotted for the utterance as shown in Fig.3(b). Clearly the pitch period and location can be simultaneously estimated from the index where the $NACC(\tau)$ takes its maximum value.

To assess the performance of the 3-level binary HOS $NACC$ system for low SNR such as 5dB and 0dB with the conventional second order statistics (autocorrelation $auto$) method [2, 5], the voiced/unvoiced regions for the utterances of 'six' is plotted in Fig.3(a), (c) respectively and for the conventional $auto$ method in Fig.3(d) and (e) respectively. Comparing these figures to (a) (c) of Fig.3 we can see that the conventional $auto$ method has failed to identify the voiced/unvoiced part while the new 3-level binary HOS $NACC$ system maintains its good performance in the presence of a high level noise. The use of the normalized autocorrelation in the new system works better than the direct autocorrelation since it accounts for the non-stationarity in the speech signal [5]. This will reduce the possibility of pitch doubling or tripling encountered in autocorrelation based algorithms due to more similarities in these lags than that of the pitch period.

4. Conclusions

Fast and robust 3-level binary HOS $NACC$ system of a speech signals has been described for accurate and reliable voiced/Unvoiced detection and simultaneous pitch period estimation for the voiced part. The algorithm can easily be implemented in digital hardware using simple combinatorial logic, i.e., an up-down counter can be used to compute each cumulant point. The performance of the new algorithm has been assessed using real speech signal in the presence of low SNR. The robustness of the $NACC$ algorithm has been demonstrated and compared to conventional second order algorithm for high level coloured Gaussian noise.

References

[1] A R. Rabiner and M.Cheng, "A comparative performance study of several pitch detection algorithms," IEEE, Trans. Acoust., Speech, Signal Processing, ASSP-24, No.5, pp. 399-418, Oct 1976.

[2] J J. Dubnowski, R W. Schafer and A R. Rabiner, "Real Digital Hardware Pitch Detector," IEEE. Trans., Acoust., Speech, Signal Processing, ASSP-24, No.1, pp. 2-8, Feb 1976.

[3] M M. Sondhi, "New Methods of Pitch Extraction," IEEE. Trans. Acoust., Speech, Signal Processing, ASSP-16, No.2, pp. 262-266, June 1968.

[4] J.Deller ,J G.Proakis and J H. Hansen, "Discrete time processing of speech signal," Macmillan Publishing Company, 1993.

[6] A.Moreno and J A.Fonollosa,"Pitch Detection of noisy speech using higher order statistics," Proc., IEEE Int., Conf., Acoust., Speech, Signal Processing, pp. 133-136, March 1992.

[7] M.Rangoussi,A.Delopoulos and M.Tsatsanis, "On the use of higher-order statistics for robust endpoint detection of speech," IEEE Signal Processing Workshop on HOS, pp. 56-60, 1993.

[8] B.Wells,"Voice/unvoiced detection based on th bispectrum," Proc., IEEE Int., Conf., Acoust., Speech, Signal Processing, pp 41.12.1-41.12.4, May 1985.

[9] C.Nikias and M.Raghuvers, "Bispectrum Estimation A digital Signal Procession Framework," Proc., IEEE, v.75, pp. 869-891, 1987.

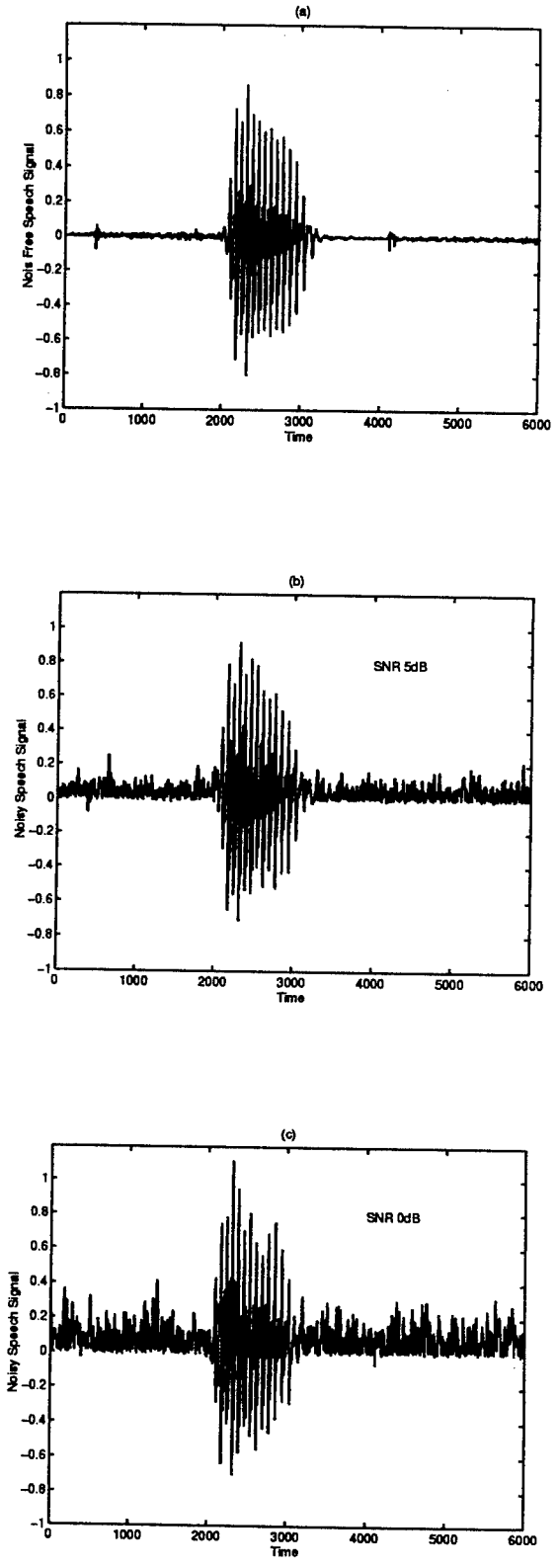
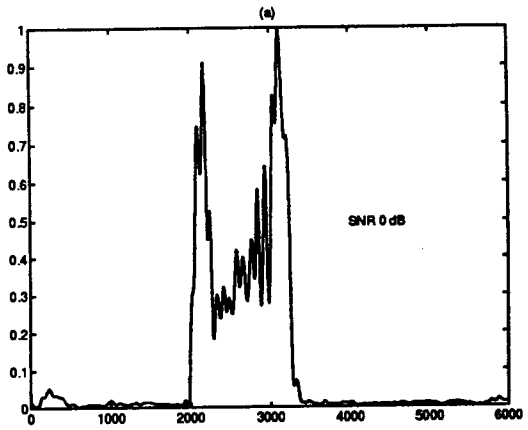
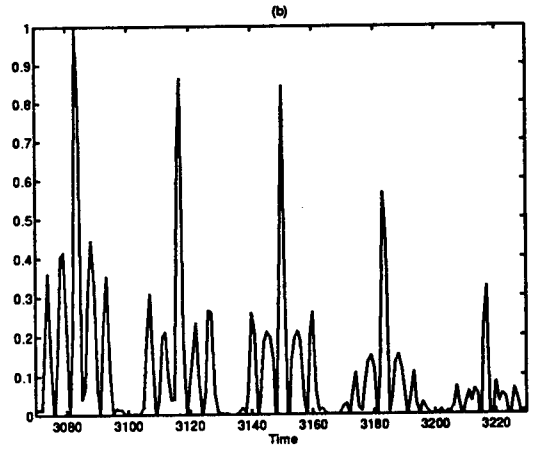
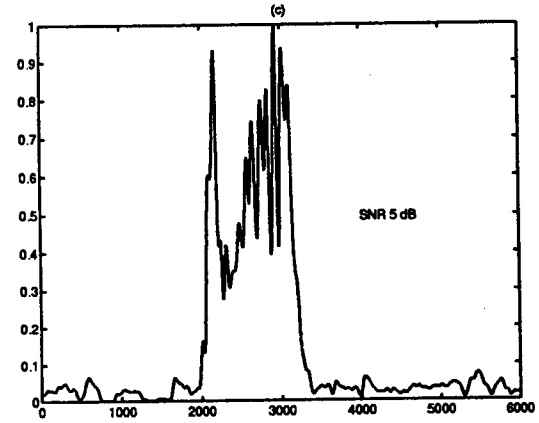
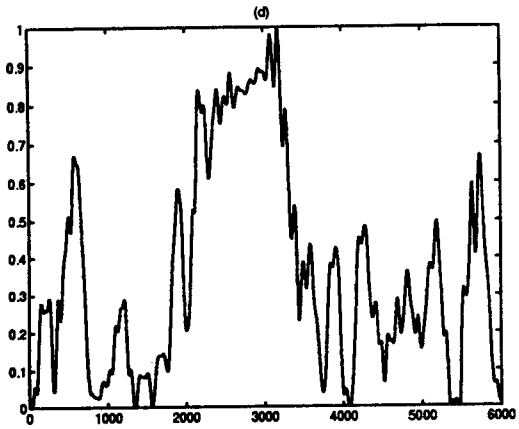


Figure 2. (a) Clean Real Speech Signal For the Word six. (b) With 5dB Coloured Noise. (c) With 0dB Coloured Noise.









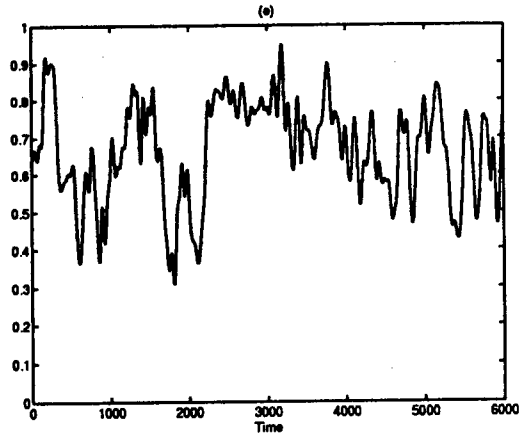


Figure 3. Simulation Results for the word 'six' Using $NACC$ With 0dB Coloured Noise.
(a)$NACC$ V/U Detection.
(b) Pitch Estimation.
(c)$NACC$ V/U Detection With 5dB.
(d) $Auto$V/U Detection With 5dB.
(e)$Auto$V/U Detection With 0dB.

ON OPTIMAL SOURCE SEPARATION BASED ON SECOND AND FOURTH ORDER CUMULANTS

J.-F. Cardoso
ENST Dept SIG/ CNRS
46 rue Barrault. 75634 /Paris. France.
cardoso@sig.enst.fr

S. Bose, B. Friedlander
Dept. ECE, University of California
Davis, CA 95616-5294, USA
sbose,friedlan@ece.ucdavis.edu

ABSTRACT

This paper[1] addresses performance issues in the source separation problem. By drawing on the theory of optimal statistic matching, we derive new contrast functions which are optimal among those involving a given set of cumulants. In low noise, the optimal combination of a particular set of cumulants are shown to be parameter independent and can be pre-computed. We give specific exemples in close form for several choices of 2nd and 4th order cumulants. The resulting performance is investigated as a function of the SNR and of the non gaussianity of the source signals and further compared to suboptimal approaches.

1. INTRODUCTION

Source separation algorithms assume a linear model for a vector $x(t)$ of observations:

$$x(t) = \underline{A}s(t) + n(t) \tag{1}$$

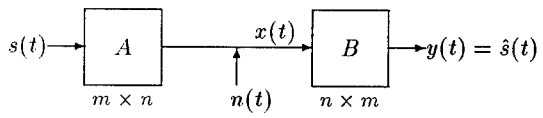
where matrix $\underline{A}$ is $m \times n$ with full column rank, $n(t)$ is additive noise and $s(t)$ is a vector of $n \times 1$ independent components, $s_1(t), \ldots, s_n(t)$: the so-called 'source signals'.

Source separation consists of recovering the source signals and/or estimating the 'mixing matrix' $\underline{A}$ *without using a priori information* about the latter. In this paper, we focus on approaches based on cumulant matching and on contrast functions. These two approaches are briefly reviewed below

Contrast functions have been introduced for source separation by Comon in [1]. The solution to source separation is defined by the separating matrix B such that its output $y = Bx$ shows the largest possible 'contrast'. For instance, Comon in his ICA approach [1] suggests to maximize

$$c(B) = \sum_i |\mathrm{Cum}(y_i, y_i^*, y_i, y_i^*)|^2 \tag{2}$$

subject to $\mathrm{E}yy^H = I_n$ (The constraint must be modified to take noise into account). A similar contrast is optimized by the joint diagonalization algorithm described in [2] as JADE.



[1] The work of S. Bose and B. Friedlander was supported by the National Science Foundation under grant NSF MIP-90-17221, and in part by the university of California MICRO program and Applied Signal Technology, Inc., and the Office of Naval Research under contract N00014-91-J-1602.

Cumulant matching approaches [3, 4] to source separation are a specific case of statistic matching. Denote $\hat{T}$ a vector of statistic such that $E_\theta \hat{T} = T(\theta)$ where θ is an unknown vector parameterizing the distribution of $\hat{T}$. An estimate $\hat{\theta}$ of θ may be obtained as $\hat{\theta} = \arg\min_\theta c(\theta)$ where $c(\theta)$ is some measure of discrepancy between $\hat{T}$ and $T(\theta)$, like

$$c(\theta) = (\hat{T} - T(\theta))^H W (\hat{T} - T(\theta)) \tag{3}$$

with W a positive matrix.

Contrasts and matching. This paper draws on the links between these two approaches. If $\hat{T}$ is a vector of sample cumulants of x_t and the unknown parameter θ can be identified with matrix A, then an objective like (3) may be turned into a contrast function. The main benefit of this perspective is that, given a particular set of cumulants, the theory of optimal statistic matching indicates the optimal weighting to apply to these cumulants.

This paper extends the work presented in [5] by considering optimal blind estimation of the mixing matrix from both 2nd and 4th order cumulants (ref. [5] considered only 4th order cumulants). Including 2nd order statistics is important in the case where little information is available in higher-order statistics. It is also more robust to the effects of noise.

2. CUMULANT MATCHING AND CONTRASTS

2.1. Assumptions.

To keep the exposition as simple as possible, we assume that the noise covariance matrix and the cumulants of orders 4, 6 and 8 of the sources are known. They are denoted

$$k_p = \mathrm{Cum}(s_p, s_p^*, s_p, s_p^*), \tag{4}$$

$$h_p = \mathrm{Cum}(s_p, s_p^*, s_p, s_p^*, s_p, s_p^*), \tag{5}$$

$$o_p = \mathrm{Cum}(s_p, s_p^*, s_p, s_p^*, s_p, s_p^*, s_p, s_p^*), \tag{6}$$

for $p = 1, \ldots, n$. We also assume that $n = m$ (the case of $m > n$ can be handled by first estimating the signal subspace, which has little effect in the source separation problem). We note that the source signals may be assumed to have unit variance:

$$E|s_p(t)|^2 = 1 \quad p = 1, \ldots, n \tag{7}$$

because the amplitude of each independent component can be integrated in the corresponding column of $\underline{A}$. The following moments will appear in the sequel

$$\alpha_p = o_p + 8h_p + 17k_p^2 + 20k_p + 4, \tag{8}$$

0-8186-7576-4/96 $5.00 © 1996 IEEE

$$\gamma_p = h_p + 5k_p + 2, \quad (9)$$
$$\lambda_p = h_p + 4k_p. \quad (10)$$

For further reference, we mention that $\alpha_p = 0$ for constant modulus sources (*i.e.* $|s_p|^2 = 1$ a.s.).

2.2. Optimal matching and contrast functions.

It is well known (see for instance [3]) that, under the appropriate assumptions, the optimal matrix W for weighting in (3) is the inverse of the (asymptotic) covariance of the vector of statistics $\hat{T}$.

$$W_{\text{opt}} = \text{Cov}_\theta^{-1}\{\hat{T}\}. \quad (11)$$

Let then $\hat{T}_x$ denote a vector of sample cumulants of x and consider how optimal cumulant matching instantiates in the source separation case where, with the above assumptions, the unknown parameter is the mixing matrix, *i.e.* $\theta = A$.

The optimal way of matching estimated cumulants $\hat{T}_x$ to their theoretical values $T_x(A)$ is to minimize

$$c(A) = \left(\hat{T}_x - T_x(A)\right)^H \text{Cov}_A^{-1}(\hat{T}_x)\left(\hat{T}_x - T_x(A)\right). \quad (12)$$

We now make a key step: thanks to the multi-linearity of the cumulants, matrix A factors out to some extent in (13). As a matter of fact, setting $B = A^{-1}$ and $y = Bx$, criterion (13) may be rewritten [6] as

$$c(B) = \left(\hat{T}_y - T_s\right)^H \text{Cov}^{-1}(\hat{T}_z)\left(\hat{T}_y - T_s\right). \quad (13)$$

which depends on B via the random vector $y = Bx$ and via the random vector

$$z \stackrel{\text{def}}{=} s + Bn. \quad (14)$$

The net result is that an optimal criterion measuring cumulant mismatch at the array output (*i.e.* for the r.v. x) has been turned into a contrast function measuring the mismatch between the 'true' cumulants T_s of the sources and the sample cumulants $\hat{T}_y$ estimated at the output $y = Bx$ of the separator.

The beauty of this manoeuver is that for high enough SNR, we have $z \approx s$, so that the criterion (13) is approximately equal to:

$$c(B) = \left(\hat{T}_y - T_s\right)^H \text{Cov}^{-1}(\hat{T}_s)\left(\hat{T}_y - T_s\right) \quad (15)$$

with the key feature that the optimal weighting matrix $\text{Cov}^{-1}(\hat{T}_s)$ does not depend on A: it is, as a matter of fact, a *constant* matrix which can be evaluated once for all for a given distribution of the sources.

Further analysis is possible because, thanks to the assumption of independent sources, matrix $\text{Cov}(\hat{T}_s)$ has a nearly diagonal structure when $\hat{T}_s$ is a vector of sample cumulants [7]. It follows that it can be 'manually' inverted. This leads, once a specific set of cumulants T has been chosen, to simple contrast functions in which cumulant mismatch is weighted on a statistically sound basis.

3. OPTIMAL CONTRAST FUNCTIONS.

Some examples are investigated in the next section where we consider a cumulant statistic $\hat{T}$ containing both 2nd and 4th order cumulants (extending the analysis of an earlier paper [5] where $\hat{T}$ could include only 4th order cumulants.) The empirical cumulants of vector y are denoted

$$\hat{r}_{ij} = \widehat{\text{Cum}}(y_i, y_j^*) \quad (16)$$
$$\hat{q}_{ik}^{jl} = \widehat{\text{Cum}}(y_i, y_j^*, y_k, y_l^*) \quad (17)$$

where $\widehat{\text{Cum}}$ is a standard cumulant estimator.

In order to carry out a detailed investigation, we make the following simplifying assumptions. All the processes are assumed to be i.i.d. and circularly distributed ; sources have non-zero kurtosis: $k_p \neq 0$ for $p = 1, \ldots, n$; the noise is normally distributed, independent of the signals with covraince matrix $\sigma^2 I$.

On this basis, we consider various sets of cumulants. We do not present general analytical results when T is the whole set of 2nd and 4th order cumulants (except in sec. 3.1) because we prefer to focus on more specific cases which can be detailed and because room is lacking for an exhaustive report. For the same reason, we leave out the hard core computations, namely explicit inversion of $\text{Cov}(\hat{T})$. More details will be found in [6].

3.1. The normal limit

When the sources are close to being normally distributed, our analysis leads to a strikingly simple conclusion because the limit form of $\text{Cov}(\hat{T})$ is itself very simple. The optimal criterion involving all 2nd and 4th order cumulants is

$$c(B) = \sum_{pq} 4|\hat{r}_p^q - \delta_p^q|^2 + \sum_{pqrs} |\hat{q}_{pr}^{qs} - k_p \delta_{pr}^{qs}|^2 \quad (18)$$

i.e. the mismatch of 2nd order cumulants receives a 4 times heavier penalty than the mismatch of 4th order cumulants. We have used the δ symbol which evaluates to 1 when all its indices are equal and to 0 otherwise.

Another limit case, which is in some sense complementary to the normal limit, is when the sources have a maximally low kurtosis. This is obtained when the sources have a constant modulus. In no noise, there is infinite weight on the auto-cumulant terms as well as on those containing cross-cumulants of the form q_{ij}^{ij}, q_{ik}^{jk} and on certain linear combinations of the 2nd and 4th order cumulants r_{ij}, q_{ij}^{ii} and q_{ij}^{jj}. It would be interesting to determine if the CMA criterion which involves only 2nd and 4th order moments and is super-efficient in the constant modulus case, could be obtained as the limit of an optimally weighted criterion involving a specific subset of 2nd and 4th order cumulants.

3.2. Autocumulants

Matching only *auto*-cumulants is to take

$$\hat{T}_y = [\hat{r}_1^1, \hat{r}_2^2, \ldots, \hat{r}_n^n, \hat{q}_{11}^{11}, \hat{q}_{22}^{22}, \ldots, \hat{q}_{nn}^{nn}],$$
$$T_s = [1, 1, \ldots, 1, k_1, k_2, \ldots, k_n].$$

The best criteria based on these cumulants turns out (maybe not surprisingly) to be a sum of criteria, each term being concerned with a particular output, *i.e.* $c_{\text{auto}}(B) = \sum_{p=1}^n c_p(B)$ with

$$c_p(B) = (k_p + 1)\,|\hat{q}_{pp}^{pp} - k_p|^2 + \alpha_p\,|\hat{r}_{pp} - 1|^2 - 2\lambda_p\,(\hat{r}_{pp} - 1)(\hat{q}_{pp}^{pp} - k_p) \quad (19)$$

which can also be written as sums of squares:

$$c_p(B) = |\xi_p(\hat{q}_{pp}^{pp} - k_p) - \zeta_p(\hat{r}_{pp} - 1)|^2 + \rho_p|\hat{r}_{pp} - 1|^2 \quad (20)$$

where $\xi_p^2 = k_p + 1$, $\zeta_p = \lambda_p/\xi_p$ and $\rho_p = \alpha_p - \zeta_p^2$. Again, if the source distributions are close to normal, then all cumulants are close to 0 and, according to (8), $\alpha \approx 4$. Thus, we have in the normal limit

$$c_{\text{auto}}(B) = \sum_{p=1}^{n} 4|\hat{r}_{pp} - 1|^2 + |\hat{q}_{pp}^{pp} - k_p|^2 \qquad (21)$$

where the coupling between 2nd and 4th order cumulant estimates has disappeared.

3.3. All 2nd and auto-4th order cumulants

This criterion is interesting because it relates to Comon criteria in that the same set of cumulants are used, except that these are combined in an optimal way. The performance relationship between the two is illustrated in the following section. The criteria itself, optimally involving the whole 2nd order information and only the auto 4th order cumulants is

$$c_{2+\text{auto4}}(B) = c_{\text{auto}}(B) + \sum_{p\neq q}^{n} |r_{pq}|^2. \qquad (22)$$

It is seen that the cross-correlation terms add very simply to the $c_{\text{auto}}(B)$ criterion.

3.4. 4th order cumulants only

The case where the whole 4th order cumulant set

$$\hat{T}_y = \{\hat{q}_{ij}^{kl} \mid 1 \leq i,j,k,l \leq n\} \qquad (23)$$

is involved in the estimation was investigated in [5]. For two identically distributed sources, we obtained

$$c_4(B) = \frac{|\hat{q}_{11}^{11} - k|^2 + |\hat{q}_{22}^{22} - k|^2}{\alpha} + \frac{|\hat{q}_{12}^{12}|^2}{(k+1)^2} + \frac{2\,|\hat{q}_{11}^{22}|^2}{(k+2)^2} + \frac{2(|\hat{q}_{12}^{11}|^2 + |\hat{q}_{22}^{21}|^2)}{\gamma + k^2} + \frac{2k^2\,|\hat{q}_{12}^{11} - \hat{q}_{22}^{21}|^2}{\gamma^2 - k^4}. \qquad (24)$$

For instance, in the case of two QAM16 sources, one has $k = -0.68$, $h = 2.08$, $o = -13.5184$ (we assume that the phase of these constellations is randomized and we recall that, by convention, the sources have unit variance). The optimal criterion based on 4th order cumulants for the source separation of two QAM16 signals is then approximately, at low noise:

$$c_4(B) = 0.72(|\hat{q}_{11}^{11} + 0.68|^2 + |\hat{q}_{22}^{22} + 0.68|^2) + 9.77|\hat{q}_{12}^{12}|^2 + 1.15|\hat{q}_{11}^{22}|^2 + 1.75(|\hat{q}_{12}^{11}|^2 + |\hat{q}_{21}^{22}|^2) + 3.72|\hat{q}_{12}^{11} - \hat{q}_{21}^{22}|^2.$$

This shows that in this case the cross-cumulant $\hat{q}_{12}^{12}$ is a more reliable measure of independence than, say, $\hat{q}_{11}^{22}$.

3.5. Link to suboptimal criteria

Now as pointed out before, a number of algorithms in the literature (ICA, JADE) can be interpreted in terms of statistic matching. However these do not use the optimum weighting. Rather they are based on a hard prewhitening in the sense that the (empirical) covariance matrix of the signals at the output of the separating matrix is constrained to be *exactly* the identity matrix, leaving no room for an 'approximate decorrelation'.

This can be interpreted as the weighted statistic matching in which virtually infinite weights are put on the second order statistic terms and flat weights on the 4th order terms. In practice the limit as one increases the weight on the second order terms can be taken. Results in the next section demonstrate the equivalence of this weighted statistic with the ICA/JADE contrasts.

Use of this suboptimal weighting results in a performance loss which we illustrate in the next section. In addition we will consider the flat weighting which is simply statistic matching with equal weights applied to all the statistics.

4. ASYMPTOTIC PERFORMANCE

We shall be using the interference rejection at the output of the separating matrix as the measure of performance in our analysis. The relevant figure of merit is the ISI (inter symbol interference) which is defined pairwise between two sources p and q as the ratio of the power of source j to that of source p at the channel output corresponding to q; since this is proportional to $1/N$ where N is the number of samples, we shall really be considering the ISI rate, given by $N \times$ ISI in the subsequent expressions and plots.

Indeed the performance can be expressed in terms of the perturbation of the global system from the identity, $\hat{B}A = I + \mathcal{E}$. Then $\text{ISI}_{pq} = \text{E}|\mathcal{E}_{pq}|^2$ and this may be computed from the covariance of $\mathcal{E}$ which itself is given by $(W^{1/2}D)^{\#}(W^{1/2}\text{Cov}(\hat{T})W^{1/2})(W^{1/2}D)^{\#H}$ where D is the derivative $\partial T/\partial \mathcal{E}$.

Two performance bounds.

Denote ρ_{pq} the (p,q) entry of matrix $(A^HA)^{-1}$ and let σ be the noise power. Any source separation using hard-whitening has a pair-wise lower bounded rejection rates [8]. We call this the 'pre-whitening bound'. For $n = m$, it is

$$\text{ISI}_{pq} + \text{ISI}_{qp} \geq \frac{1}{2}(1 + \sigma\rho_{pp})(1 + \sigma\rho_{qq}) \qquad (25)$$

Another bound is provided by computing what would be the ISI if matrix A was identified knowing the source signals. This is the so-called I/O (input/output) bound. It is:

$$\text{ISI}_{pq} \geq \sigma\rho_{pp} \qquad (26)$$

Some numerical evaluaions are given below for illustration.. They are computed for a $2times2$ matrix $A = [a_1, a_2]$ such that $|a_1| = |a_2|$ and with the values of ρ_{ij} indicated on the plot. We use QAM4 and QAM16 distributins with a randomized phase (to ensure circularity).

Optimal contrasts. Figure 1 for the case of two identical sources shows the effect of strictly increasing information as we go from the optimal criterion using the 4th order cumulants to that involving entire second and 4th order cumulants to the input-output bound. The optimal criterion involving all the second order and the 4th order auto cumulants is also shown. We note that for QAM4 at good SNR, optimal matching of 2nd and 4th order cumulants is close to I/O performance (there is a ratio of 2 in terms of ISI); in this situation includind 2nd order information seems crucial for good performance (see how the curve for 4th order levels off at increasing SNRs). These conclusions do not apply to the QAM16 case (there is clearly a 'constant modulus effect' here).

Suboptimal contrasts The next figure 2 indicates the performance of actual JADE/ICA contrast based algorithms compared to the suboptimum criterion employing flat weighting. The bound for pre-whitening is included for

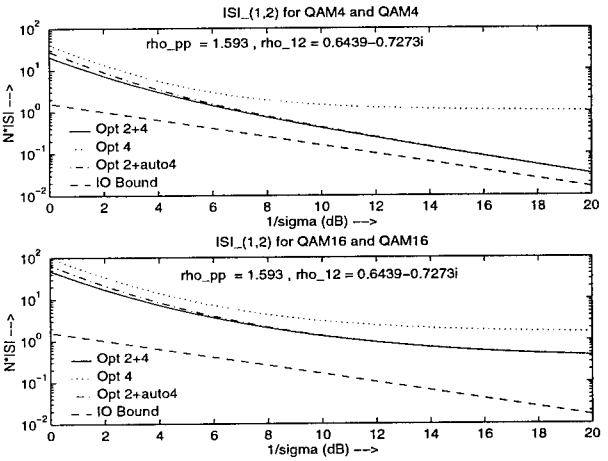


Figure 1. ISI:qam:4+4:16+16:

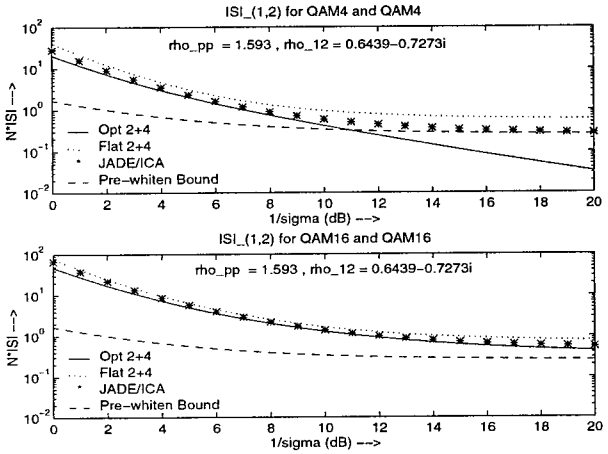


Figure 2. ISI:4+4:16+16

reference. We note that the use of the optimum weights instead of the hard weights for pre-whitening overcomes that bound. Moreover the actual performance of the above algorithms can be dominated by this effect in high SNR, a point which is more evident in the upper panel for two QAM4 sources.

Effect of source distributions We round off this discussion with an illustration of the effect on performance as the source distribution varies from the constant modulus limit to the Gaussian limit. Figure 3 illustrates the variation of ISI rate as we start with two QAM4 sources and make them progressively more Gaussian by adding a Gaussian component to the source with a relative amplitude of t, which can then be treated as a "Gaussianity parameter". Note how the pre-whitening loss constrains the performance of the JADE/ICA algorithms at small values of t while the optimum 4th order criterion does uniformly worse. Further note the flattening of the optimum 2+4 curve near the CM limit and the transition near $t = 1$ corresponding to the Gaussian component effectively smearing out the discrete nature of the distribution. On the other hand, we see that ICA/JADE contrasts do as well as optimal matching of 2nd and 4th order cumulants as the distribution of the sources gets close to normality.

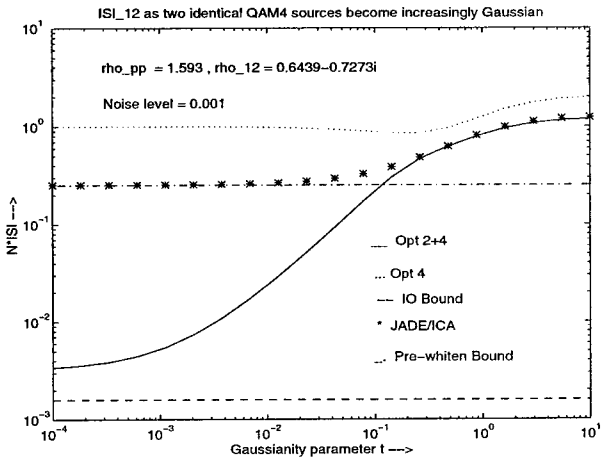


Figure 3. ISI:ToGauss.eps

CONCLUSION

This paper develops the link between contrasts for source separation and criteria based on optimal statistic matching. A number of optimal and sub-optimal criteria are proposed, studied and compared. The ICA contrasts of Comon and of Cardoso are investigated within the same framework; it is seen from the examples that the primary cause of performance loss of those algorithms relative to the optimum criteria at high SNR is caused by the hard pre-whitening. The effect of source distributions on performance is also illustrated; we find in particular that JADE/ICA constrasts are very suboptimal for constant modulus sources but tend to be optimal as the source distributions are pulled from this limit case. More illustrative examples will be presented at the workshop.

REFERENCES

[1] P. Comon, "Independent component analysis, a new concept ?," *Signal Processing*, vol. 36, pp. 287–314, Apr. 1994.

[2] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, pp. 362–370, Dec. 1993.

[3] B. Friedlander and B. Porat, "Asymptotically optimal estimation of MA and ARMA parameters of non Gaussian processes from high-order moments," *IEEE Tr. on AC*, vol. 35, no. 1, pp. 27–37, 1990.

[4] G. Giannakis and S. Shamsunder, "Modelling of non Gaussian array data using cumulants: DOA estimation of more sources with less sensors," *Signal Processing*, vol. 30, pp. 279–297, July 1993.

[5] J.-F. Cardoso, S. Bose, and B. Friedlander, "Output cumulant matching for source separation," in *Proc. IEEE SP Workshop on H.O.S., Aiguablava, Spain*, pp. 44–48, 1995.

[6] S. Bose and J.-F. Cardoso, "Optimal and sub-optimal contrast functions for source separation." In preparation, 1995.

[7] P. McCullagh, *Tensor Methods in Statistics*. Monographs on Statistics and Applied Probability, Chapman and Hall, 1987.

[8] J.-F. Cardoso, "On the performance of source separation algorithms," in *Proc. EUSIPCO*, pp. 776–779, Sept. 1994.

Peak Filter and Notch Filter for Harmonic Retrieval Using Higher-Order Statistics

Chii-Horng Chen and Chong-Yung Chi

Department of Electrical Engineering
National Tsing Hua University, Hsinchu, Taiwan, Republic of China
cychi@ee.nthu.edu.tw

Abstract

In this paper, an algorithm using the well-known notch filter and an algorithm using a peak filter are proposed to estimate the frequencies of sinusoidal signals with a given set of Gaussian noise corrupted measurements $y(n)$ provided that the number of sinusoids is known in advance. The former processes $y(n)$ such that a single fourth-order cumulant of the notch filter output is minimum in absolute value, while the latter processes $y(n)$ such that the same fourth-order cumulant of the peak filter output is maximum in absolute value. Then the unknown frequencies are obtained from the optimum notch filter and the optimum peak filter, respectively. A performance analysis of the proposed two algorithms is then presented followed by some simulation results for a performance comparison of the proposed algorithms and Swami and Mendel's SVD low-rank approximation method.

1. Introduction

Estimation of parameters of sinusoidal signals is a problem to estimate frequencies $0 < \omega_i < \pi$ and amplitudes $A_i > 0$ with a given set of noisy measurements modeled as follows:

$$y(n) = \sum_{i=1}^{p} A_i cos(\omega_i n + \phi_i) + w(n) \qquad (1)$$

where p is the total number of sinusoids, ϕ_i's are random phases and $w(n)$ is additive noise. This is a well defined problem in some statistical signal processing areas such as noise and interference cancellation and estimation of direction of arrival (DOA) of narrowband source signals in sonar and radar arrays. Usually, frequency estimation is followed by amplitude estimation because the former often resorts to a nonlinear search procedure while the latter can be solved from a set of linear equations once ω_i's are estimated. There have been a number of correlation (second-order statistics) based algorithms reported for the estimation of ω_i's such as Pisarenko's harmonic decomposition procedure [1], Tufts and Kumaresan's method [2], overdetermined Yule-Walker method [3] and maximum-likelihood method [4]. Chicharo and Ng [5] proposed an adaptive notch filtering approach for the enhancement and tracking of sinusoids in additive noise. The transfer function of notch filters (IIR filters) of order equal to $2p$ is given by

$$H_p(z) = \frac{\prod_{i=1}^{p}(1 + \beta a_i z^{-1} + \beta^2 z^{-2})}{\prod_{i=1}^{p}(1 + \alpha a_i z^{-1} + \alpha^2 z^{-2})} \qquad (2)$$

where $0 \leq \beta \leq 1$ and $0 \leq \alpha < \beta$. The ω_i's are obtained by solving roots of the numerator polynomial of the adaptive notch filter.

Higher-order (≥ 3) statistics, known as cumulants, have been used for frequency estimation of sinusoidal signals when measurement noise is Gaussian because all higher-order cumulants of Gaussian noise are equal to zero. Thus cumulant based frequency estimation algorithms [6-8] are insensitive to additive Gaussian noise. In this paper, the notch filter and a peak filter, using a single fourth-order cumulant are proposed for frequency estimation of sinusoidal signals. A performance analysis of the proposed frequency estimation algorithms (one using the notch filter and the other using a peak filter) is presented followed by some simulation results.

2. Cumulant based harmonic retrieval using notch filters and peak filters

Assume that we are given a set of noisy measurements $y(n), n = 0, 1, \cdots, N-1$ modeled by (1) under the following assumptions:

(A1) The number p of sinusoids is known *a priori*; amplitudes $A_i > 0$ and frequencies $0 < \omega_i < \pi, i = 1, \cdots, p$ are unknown.

(A2) Measurement noise $w(n)$ is Gaussian with unknown statistics.

This work is supported by the National Science Council under Grant NSC 85-2213-E-007-012.

0-8186-7576-4/96 $5.00 © 1996 IEEE

(A3) Phase ϕ_i's are i.i.d. random variables with a uniform probability density function over $[-\pi, \pi)$ and they are statistically independent of $w(n)$.

Let $C_{M,e}(k_1, \cdots, k_{M-1})$ denote the Mth-order cumulant function of a non-Gaussian signal $e(n)$. We need the following proposition on which the two frequency estimation algorithms to be presented are based.

Proposition 1. Let $e(n)$ be the output of a linear time-invariant system $H(z)$ with input $y(n)$ given by (1) under the assumptions **(A1)** through **(A3)**, i.e.,

$$e(n) = y(n) * h(n) = \sum_{k=-\infty}^{\infty} h(k)y(n-k) \qquad (3)$$

where $h(n)$ is the impluse response of the system. Then

$$C_{4,e}(0,0,0) = -\frac{3}{8}\sum_{i=1}^{p} A_i^4 \cdot |H(e^{j\omega_i})|^4 \qquad (4)$$

A. Notch filter based algorithm:

By **Proposition 1**, one can infer the following fact:

(F1) Let $e(n)$ be the output signal given by (3) of the notch filter $H_p(z)$ with $\beta = 1$ given by (2). Then $|C_{4,e}(0,0,0)| = min\{|C_{4,e}(0,0,0)|\} = 0$ occurs only when $|H_p(e^{j\omega_i})| = 0$ for all i, i.e.,

$$a_i = -2 \cdot cos(\omega_i) \qquad (5)$$

Let $\widehat{C}_{4,e}(0,0,0)$ denote the fourth-order sample cumulant associated with $C_{4,e}(0,0,0)$. By **(F1)**, we propose the following frequency estimation algorithm:

Algorithm 1:

(S1) Let $e(n)$ be the output signal given by (3) of the notch filter $H_p(z)$ ($\beta = 1$) given by (2). Find the optimum parameters $\widehat{a}_i, i = 1, \cdots, p$ of $H_p(z)$ such that $|\widehat{C}_{4,e}(0,0,0)|$ is minimum.

(S2) Obtain $\widehat{\omega}_i$ by (5), i.e.,

$$\widehat{\omega}_i = arccos(-\widehat{a}_i/2) \qquad (6)$$

B. Peak filter based algorithm:

The peak filter used for frequency estimation is an IIR filter with transfer function

$$V_p(z) = \frac{\prod_{i=1}^{p}(1 + \rho\alpha a_i z^{-1} + \rho^2\alpha^2 z^{-2})}{\prod_{i=1}^{p}(1 + \alpha a_i z^{-1} + \alpha^2 z^{-2})} \qquad (7)$$

where $0 < \alpha < 1$ and $0 \leq \rho < 1$. The peak filter differs from the notch filter in that each pair of complex conjugate poles (with magnitude α) are closer to the unit circle than the associated pair of complex conjugate zeros (with magnitude $\alpha\rho < \alpha$).

Again, by **Proposition 1**, one can infer the following fact:

(F2) Let $e(n)$ be the output signal given by (3) of the peak filter $V_p(z)$ given by (7). Then $|C_{4,e}(0,0,0)| = max\{|C_{4,e}(0,0,0)|\}$ occurs when $a_i, i = 1, \cdots, p$ of $V_p(z)$ are given by (5).

The following frequency estimation algorithm is due to **(F2)**:

Algorithm 2:

(S1) Let $e(n)$ be the output signal given by (3) of the peak filter $V_p(z)$ given by (7). Find the optimum parameters $\widehat{a}_i, i = 1, \cdots, p$ of $V_p(z)$ such that $|\widehat{C}_{4,e}(0,0,0)|$ is maximum.

(S2) Obtain $\widehat{\omega}_i$ using (6).

To find the optimum $\widehat{a}_i$ required in (S1) of the proposed two algorithms, we have to resort to iterative optimization algorithms because

$$\widehat{C}_{4,e}(0,0,0) = \frac{1}{N}\sum_{n=0}^{N-1} e^4(n) - 3\left(\frac{1}{N}\sum_{n=0}^{N-1} e^2(n)\right)^2 \qquad (8)$$

is a highly nonlinear function of a_i. A gradient type iterative algorithm is used to search for the optimum $\mathbf{a} = (a_1, \cdots, a_p)^T$. At the nth iteration, $\widehat{\mathbf{a}}$ is updated by

$$\widehat{\mathbf{a}}(n) = \widehat{\mathbf{a}}(n-1) \pm \eta \frac{\partial|\widehat{C}_{4,e}(0,0,0)|}{\partial \mathbf{a}}\Big|_{\mathbf{a}=\widehat{\mathbf{a}}(n-1)} \qquad (9)$$

where η is a small positive constant and "$-$" is for **Algorithm 1** and "$+$" is for **Algorithm 2**, respectively. An initial condition for $\widehat{\mathbf{a}}(0)$ is needed to initialize the iterative algorithm given by (9). Swami and Mendel's method [6] can be used to obtain an estimate for each ω_i and the associated a_i computed by (5) can be used for $\widehat{\mathbf{a}}(0)$.

3. Performance analysis

To illustrate the performance of the proposed two frequency estimation algorithms, let us assume that $p = 1$, $A_1 = 1$, $\omega_1 = 0.5\pi$ and $w(n)$ is white with variance σ_w^2. Then

$$|C_{4,e}(0,0,0)| = \begin{cases} (3/8)|H_1(e^{j\omega_1})|^4 & \text{for } \textbf{Algorithm 1} \\ (3/8)|V_1(e^{j\omega_1})|^4 & \text{for } \textbf{Algorithm 2} \end{cases}$$

with the same optimum solution $a_1 = a = 0$ by (5). Figure 1 (a) shows $log_{10}|C_{4,e}(0,0,0)|$ associated with the peak filter used by **Algorithm 2** for $\rho = 0.9$ and $\alpha = 0.9$ (dashed line), 0.95 (dotted line) and 0.99 (solid line), respectively, and Figure 1 (b) shows $|C_{4,e}(0,0,0)|$ instead of $log_{10}|C_{4,e}(0,0,0)|$ associated with the notch filter used by **Algorithm 1** for $\beta = 1$ and $\alpha = 0.9$ (dashed line), 0.95 (dotted line) and 0.99 (solid line), respectively. One can see, from these two figures, that a single peak (whose magnitude is larger for larger α) in Figure 1 (a) and a single notch ($|C_{4,e}(0,0,0)| = 0$) in Figure 1 (b) located at $a = 0$ are associated with

each curve, and that the larger α, the narrower is the peak for the former and the notch for the latter.

It can be shown that

$$\widehat{C}_{4,e}(0,0,0) \approx C_{4,e}(0,0,0) + \widehat{C}_{4,w'}(0,0,0) \qquad (10)$$

where $\widehat{C}_{4,w'}(0,0,0)$ is the fourth-order sample cumulant of the Gaussian noise $w'(n)$ in the filter output $e(n)$ due to the presence of $w(n)$. Note that $\widehat{C}_{4,w'}(0,0,0)$ itself is a random variable. For the notch filter, it can be shown that for $a = 0$

$$E[\widehat{C}^2_{4,w'}(0,0,0)] \geq \sigma_1^2 = \left[\frac{6}{N}\sum_{l=-N}^{N}(1-\frac{|l|}{N})r^2_{w'}(l)\right]^2$$

where $r_{w'}(l)$ (autocorrelation function of $w'(n)$) is given by

$$r_{w'}(l) = \begin{cases} \sigma_w^2 \cdot \frac{2}{1+\alpha^2}, & l = 0 \\ \sigma_w^2 \cdot \frac{(\alpha^2-1)\alpha^{|l|}cos(l\pi/2)}{(1+\alpha^2)\alpha^2}, & l \neq 0 \end{cases}$$

Therefore, $min\{|C_{4,e}(0,0,0)|\} = 0$ is easily smeared by $\widehat{C}_{4,w'}(0,0,0)$ if $\sigma_1 \gg 0$ (low SNR). On the other hand, for the peak filter, it can be shown that for $a = 0$

$$E[\widehat{C}^2_{4,w'}(0,0,0)] \leq \sigma_2^2 = 1050 \cdot \left(1 + \frac{(\rho^2-1)^2\alpha^4}{1-\alpha^4}\right)^4 \cdot \sigma_w^8$$

One can easily infer that if $max\{|C_{4,e}(0,0,0)|\}/\sigma_2 = (3/8)|V_1(e^{j0.5\pi})|^4/\sigma_2 \gg 1$, the optimum $a = 0$ can be accurately estimated even if SNR is low. For instance, $max\{|C_{4,e}(0,0,0)|\} = 4316 \gg \sigma_2 = 28.6$ for $SNR = 0\ dB$, $\rho = 0.9$ and $\alpha = 0.99$. Therefore, the previous performance analysis leads to following fact:

(F3) **Algorithm 2** outperforms **Algorithm 1** for finite data, because the former is more robust to additive noise than the latter.

4. Simulation results

As mentioned in Section 2, SM method [6] was used to provide an initial condition for the proposed two frequency estimation algorithms. In the simulation, thirty independent runs were performed to compute the mean square error (MSE) defined as

$$MSE = \frac{1}{30}\sum_{j=1}^{30}\{\sum_{i=1}^{p}(\widehat{f}_{ij} - f_i)^2\} \qquad (11)$$

where $f_i = \omega_i/2\pi$ and $\widehat{f}_{ij}$ is the obtained estimate for f_i at the jth run. Two sets of simulation results ($p = 1$ and $p = 2$, $A_1 = A_2$) for measurement noise $w(n)$ assumed to be white Gaussian were obtained using **Algorithm 1** with $\beta = 1$ and $\alpha = 0.99$ and **Algorithm 2** with $\rho = 0.9$ and $\alpha = 0.99$, respectively.

Let $SNR = A_i^2/(2\sigma_w^2)$ where σ_w^2 is the variance of $w(n)$. Table 1 shows the simulation results for $p = 1$, $A_1 = 1$, $f_1 = 0.2$, $N = 1024$, 2048, 4096 and $SNR = 0$, 5, 10, 15, 20 dB. From this table, one can see that **Algorithm 2** performs best, SM method performs second and **Algorithm 1** performs worst. On the other hand, Table 2 shows the corresponding results for $p = 2$, $A_1 = A_2 = 1$, $f_1 = 0.1$ and $f_2 = 0.2$. From Table 2, one can see that **Algorithm 2** performs best except for the case that $SNR = 0\ dB$ when $N = 1024$ and 2048 while SM method performs best for this case. These simulation results indicate that the latter may perform better than the former for small N and low SNR. However, **Algorithm 1** always performs worst as predicted by **(F3)**, and its performance for low SNR may not improve even when N is increased (see the results for $N = 2048$ and 4096 when $SNR = 0\ dB$, 5 dB and 10 dB in Table 2). The reason for this is that although N was doubled, the notch of $min\{|C_{4,e}(0,0,0)|\} = 0$ in some realizations was severely smeared by $\widehat{C}_{4,w'}(0,0,0) \approx \widehat{C}_{4,e}(0,0,0)$ at the vicinity of $(a_1,a_2)^T = (-2cos(0.2\pi),\ -2cos(0.4\pi))^T$ where $w'(n)$ was the Gaussian noise in the notch filter output due to measurement noise $w(n)$.

5. Conclusions

We have presented two frequency estimation algorithms with a given set of noisy sinusoidal signals under the three assumptions **(A1)** through **(A3)**. **Algorithm 1** uses the notch filter and **Algorithm 2** uses the peak filter, while the former tries to minimize but the latter tries to maximize the same single absolute fourth-order cumulant. A performance analysis for the proposed two algorithms was also presented. Then some simulation results obtained by the proposed two algorithms and Swami and Mendel's method were presented for a performance comparison. The presented simulation results support that **Algorithm 2** performs best for the case of $p = 1$, but for the case of $p = 2$ it performs best except that when N is small and SNR is low, Swami and Mendel's method performs best.

6. References

[1] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Roy. Astron. Soc.*, vol. 33, pp. 347–366, 1973.

[2] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood," *Proc. IEEE,* vol. 70, no. 9, pp. 975–989, Sept. 1982.

[3] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE,* vol. 70, no. 9, pp. 907–939, Sept. 1982.

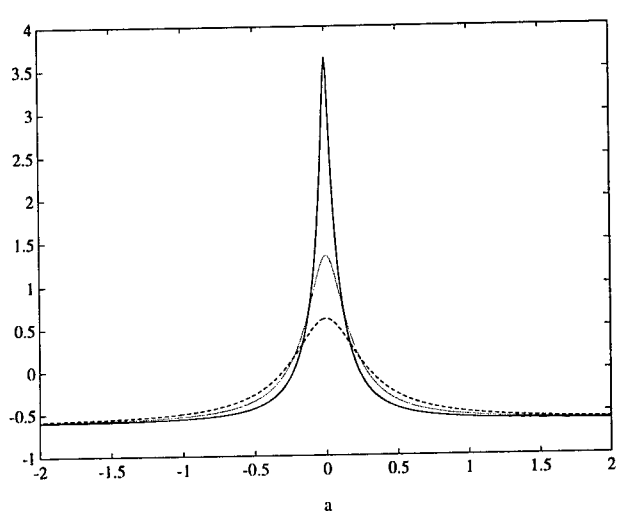
[4] P. Stoica, R. L. Moses, B. Friedlander and T. Söderström, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. ASSP,* vol. 37, no. 3, pp. 378–392, March 1989.

[5] J. F. Chicharo and T. S. Ng, "Gradient-based adaptive IIR notch filtering for frequency estimation," *IEEE Trans. ASSP*, vol. 38, no. 5, pp. 769–777, May 1990.

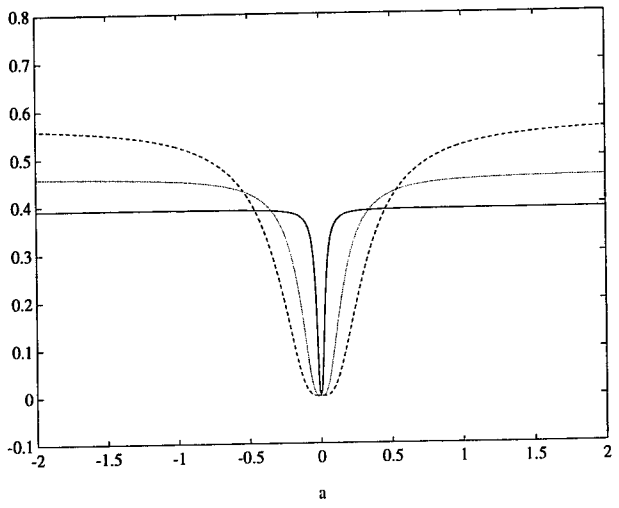
[6] A. Swami and J. M. Mendel, "Cumulant-based approach to the harmonic retrieval and related problems," *IEEE Trans. ASSP*, vol. 39, no. 5, pp. 1099–1109, May 1991.

[7] Z. Shi and F. W. Fairman, "Harmonic retrieval via state space and fourth-order cumulants," *IEEE Trans. Signal Processing*, vol. 42, no. 5, pp. 1109–1119, May 1994.

[8] J. M. M. Anderson, G. B. Giannakis and A. Swami, "Harmonic retrieval using higher order statistics: A deterministic formulation," *IEEE Trans. Signal Processing*, vol. 43, no. 8, pp. 1880–1889, Aug. 1995.



(a)



(b)

Figure 1. (a) $log_{10}|C_{4,e}(0,0,0)|$ associated with the peak filter for $\rho = 0.9$ and $\alpha = 0.9$ (dashed line), 0.95 (dotted line) and 0.99 (solid line), respectively; (b) $|C_{4,e}(0,0,0)|$ associated with the notch filter for $\beta = 1$ and $\alpha = 0.9$ (dashed line), 0.95 (dotted line) and 0.99 (solid line), respectively.

N	SNR	$MSE(\times 10^{-7})$		
		SM Method	**Algori -thm 2**	**Algori -thm 1**
1024	20 dB	0.0171	0.0026	0.0471
	15 dB	0.0266	0.0122	0.1401
	10 dB	0.0634	0.0310	0.3955
	5 dB	0.3141	0.1838	1.6149
	0 dB	4.4869	2.3268	8.3789
2048	20 dB	0.0026	0.0003	0.0478
	15 dB	0.0059	0.0026	0.1439
	10 dB	0.0214	0.0112	0.5197
	5 dB	0.1717	0.0856	1.7674
	0 dB	3.7787	1.6415	8.2338
4096	20 dB	0.0010	0.0003	0.0199
	15 dB	0.0019	0.0004	0.0617
	10 dB	0.0060	0.0015	0.1888
	5 dB	0.0480	0.0152	0.5698
	0 dB	0.9703	0.3434	1.7159

Table 1. MSE's associated with the SM method, **Algorithm 1** (using the notch filter) and **Algorithm 2** (using the peak filter) for $p = 1$ and $f_1 = 0.2$.

N	SNR	$MSE(\times 10^{-7})$		
		SM Method	**Algori -thm 2**	**Algori -thm 1**
1024	20 dB	1.4042	0.0035	0.5792
	15 dB	1.5070	0.0041	2.0163
	10 dB	1.8507	0.1120	100.93
	5 dB	4.8170	0.0266	1394.9
	0 dB	86.835	148.69	3034.0
2048	20 dB	0.2858	0.0010	0.5308
	15 dB	0.3233	0.0011	1.7446
	10 dB	0.5162	0.0023	42.250
	5 dB	2.3045	0.0066	269.14
	0 dB	39.381	41.777	1605.0
4096	20 dB	0.0974	0.0002	0.5556
	15 dB	0.1133	0.0005	1.8608
	10 dB	0.2016	0.0006	239.27
	5 dB	0.9268	0.0013	518.37
	0 dB	10.896	0.0035	10596

Table 2. MSE's associated with the SM method, **Algorithm 1** (using the notch filter) and **Algorithm 2** (using the peak filter) for $p = 2$, $f_1 = 0.1$ and $f_2 = 0.2$.

Robust Frequency-Domain Bicoherence Estimation

Justin W A FACKRELL*, Achilleas G STOGIOGLOU† and Steve McLAUGHLIN‡
Department of Electrical Engineering, University of Edinburgh, King's Buildings, Edinburgh EH9 3JL, UK.
Tel/Fax: +44 [131] 650 5655 / 650 6554. Email: jwaf@ee.ed.ac.uk

Abstract

Estimates of Higher Order Statistical quantities (such as the bicoherence) have higher variances than their second-order counterparts. Reliable estimates can be obtained by using longer data records, but in practice this is often not possible. In direct-method bicoherence estimation, estimates from shorter records can be highly dependent on measurement errors and background noise. To try to get around these problems, a new bicoherence measure based on the $\alpha-$trimmed mean bispectrum is described. Simulations indicate how well this new measure performs compared to the standard bicoherence measure.

1 Introduction

The discrete bispectrum of a discrete, stationary, stochastic process $x(n)$ can be estimated using a segment-averaging approach [4]; the signal $x(n)(n = 1, .., N)$ is divided into K non-overlapping segments $(m = 1, .., K)$, each of length N_{DFT} $(N = N_{DFT}K)$. The $N_{DFT}-$point DFT $X_m(k)$ is computed in each segment m, and the bispectrum is estimated using

$$\hat{B}(k,l) = \frac{1}{K}\sum B_m(k,l) = \frac{1}{K}\sum X_m(k)X_m(l)X_m^*(k+l), \quad (1)$$

in which $\sum \equiv \sum_{m=1}^{K}$. As it stands the variance of this estimate is different in each bifrequency bin (k,l) [3]. The variance can be (approximately) flattened by *normalising* the bispectrum to form the *squared bicoherence* $b^2(k,l)$ [4]

$$\hat{b}^2(k,l) = \frac{\mid \hat{B}(k,l) \mid^2}{\hat{S}(k,l)\hat{P}(k+l)}, \quad (2)$$

in which

$$\begin{aligned}\hat{S}(k,l) &= \frac{1}{K}\sum S_m(k,l) \\ &= \frac{1}{K}\sum \mid X_m(k)X_m(l) \mid^2,\end{aligned}$$

$$\begin{aligned}\hat{P}(k+l) &= \frac{1}{K}\sum P_m(k+l) \\ &= \frac{1}{K}\sum X_m(k+l)X_m^*(k+l), \quad (3)\end{aligned}$$

are the denominator components and again $\sum \equiv \sum_{m=1}^{K}$. The methods used in this paper are equally applicable to other bispectrum normalisations, and similar results (not shown) have been achieved for the skewness function.

Previous applications of the bicoherence for Quadratic Phase Coupling (QPC) detection [4, 1] have considered coupled sinusoids in white Gaussian background noise, but there has been no investigation into how well the bicoherence detects QPC if the background noise includes disturbances such as transients. Furthermore, previous analyses have typically used long data records such that $K \approx N_{DFT}$, but in practical applications the data length N (and thus K) may be limited.

Under these more demanding conditions, new problems can arise. Although the bispectral estimate (Eqn. 1) is asymptotically complex normal [3], if K is small the distributions of $\Re[\hat{B}(k,l)]$ and $\Im[\hat{B}(k,l)]$ may be non-normal. Furthermore, bispectral estimates from short records are small-sample estimates of large-variance quantities, and so occasional large values (possibly due to estimation errors, or to external transients) can exert a strong influence over the bispectral estimate. In other words, the distributions of bispectral estimates based on small, noisy samples may have long tails, and so bispectral averages formed using the mean estimator (as in Eqn. 1) may be susceptible to outliers.

The new method developed in this paper is based on forming a bispectral estimate without using the values in the tails of the distribution. Obviously this will *reduce* the variance of the estimate, and in the case where the sources of error described above are small the new estimate will be worse than the raw estimate. However, in cases where the sources of error are significantly large (and this can often be gleaned from inspection of the time series and power spectrum) the new method can result in improved estimates.

The key assumption in this new method is that the sources of error described above influence the bispectral estimate *in a small number of segments only*. i.e. that extreme bispec-

*supported by EPSRC and BT Laboratories.
†supported by NERC.
‡supported by The Royal Society.

tral values (due to either measurement errors or transients) occur only in a small number of segments. By excluding these it is hoped that the resulting bispectral estimate, and subsequent bicoherence estimates will be more robust. Previous applications of robust techniques to HOS have been limited to time domain parameter estimation problems [5].

2 An α-trimmed mean estimator for the bispectrum

The key steps in the computation of the $\alpha-$trimmed-mean bispectrum estimate will now be described. These are based on the $\alpha-$trimmed mean algorithm described in [6]. The $\alpha-$trimming is applied to the real and imaginary parts of the bispectral estimate separately, because $\alpha-$trimming is only appropriate on signals which are symmetrically distributed [6] (For this reason it cannot be applied directly to the bicoherence estimate). The algorithm is described below.

1. Divide the time series $x(n)(n = 1, ...N)$ into K segments (for clarity it is assumed that there is no overlapping of the frames).

2. Compute the raw bispectral estimates $B_m(k,l)(m = 1, .., K)$ (see Eqn. 1).

3. For each (k,l) form two vectors $\mathbf{r} = [r_1, .., r_K]^T$ and $\mathbf{i} = [i_1, .., i_K]^T$, each containing K integers. Each integer in $\mathbf{r}$ (or $\mathbf{i}$) identifies a segment m, and hence a value of $\Re[B_m(k,l)]$ (or $\Im[B_m(k,l)]$). The integers in $\mathbf{r}$ and $\mathbf{i}$ are arranged so that

$$\begin{aligned} \Re[B_{r_1}(k,l)] &= \min_m \Re[B_m(k,l)], \\ \Re[B_{r_K}(k,l)] &= \max_m \Re[B_m(k,l)], \\ \Im[B_{i_1}(k,l)] &= \min_m \Im[B_m(k,l)], \\ \Im[B_{i_K}(k,l)] &= \max_m \Im[B_m(k,l)]. \end{aligned} \quad (4)$$

   Note that the ordering is done separately for real and imaginary parts. $\mathbf{r}$ and $\mathbf{i}$ thus determine the *order statistics* [6] of the real and imaginary parts of the raw segmental bispectrum estimates.

4. The α-trimmed mean estimate at a particular bifrequency (k,l) is then evaluated as the sum of the $\alpha-$trimmed real and imaginary parts.

$$\begin{aligned} \hat{B}_\alpha = \frac{1}{K(1-2\alpha)} \Big\{ & \\ & (1-r)\big[\Re[B_{r_{g+1}} + B_{r_{K-g}}] + \\ & j\Im[B_{i_{g+1}} + B_{i_{K-g}}]\big] \\ + & \sum_{m=g+2}^{K-g-1} \Re[B_{r_m}] + j\Im[B_{i_m}] \Big\}, \end{aligned} \quad (5)$$

   where g is the largest integer less than or equal to αK, $r = \alpha K - g$ and the (k,l) has been dropped for clarity. Eqn. 5 is a summation over the segments identified by the *middle* $K - 2g$ values of $\mathbf{r}$ and $\mathbf{i}$ (i.e. $[r_{g+1}, .., r_{K-g}]$ and $[i_{g+1}, .., i_{K-g}]$).

5. This estimate is thus formed for all bifrequencies (k,l) of interest.

This estimate is based on the *absolute* values of the real and imaginary parts of the bispectrum estimates. It discards the contributions to $\hat{B}$ of a fraction of segments. If α is increased then the contributions from more segments will be discarded.

It is important to stress that the list of segments for which bispectral values are discarded can be different at different bifrequencies. This is intended to accommodate interference such as bandlimited transients, which will affect the raw bispectral estimates at some bifrequencies only.

Furthermore, the choice to apply the α-trimming algorithm *separately* to the real and imaginary parts of the bispectral estimate means that at any given bifrequency, the segments from which the real part of the bispectral estimate is discarded may be different from the segments from which the imaginary part is discarded. The tacit assumption here is that the real and imaginary parts of the segmental bispectral estimates $B_m(k,l)$ are independent of each other. Although this is asymptotically true (because the estimator is asymptotically complex normal[3]), it is not at this time clear how valid this assumption is in practical situations.

2.1 Normalisation

Since the contributions to the bispectrum estimate from the tails of the sampling distribution are excluded by the $\alpha-$trimming technique described above, the denominator of the normalisation in Eqn. 2 also needs to be changed. In order to try to preserve the magnitude of the bicoherence, the following, slightly ad-hoc approach is taken at each bifrequency. Since both real and imaginary parts are treated in the same way, only the real part will be considered here.

- Form the vector $\mathbf{r}$, which lists the segment numbers associated with $\Re[B_m]$, ordered according to the size of $\Re[B_m]$, as described above.
- The segments listed at the top and bottom of $\mathbf{r}$ (i.e. $r_1, ..r_g, r_{K-g+1}, ..r_K$) are *excluded* from the $\alpha-$trimmed estimate B_α (Eqn. 5).
- *Halve* the contributions of these outlying segments $(r_1, ..r_g, r_{K-g+1}, ..r_K)$ to the estimates on the denominator of Eqn. 3.

In this last step, the reason for halving, rather than excluding altogether, contributions to denominator estimates is explained as follows. Consider one segment m in which

the raw estimates of the numerator and denominator of the bicoherence are $B_m(k,l)$, $S_m(k,l)$ and $P_m(k+l)$. Now the contribution of segment m to the final bispectral estimate will be zero *only if* both $\Re[B_m(k,l)]$ and $\Im[B_m(k,l)]$ are extreme values (i.e. so both the real and the imaginary parts are trimmed). In such circumstances it is desirable that the contributions to $\hat{S}(k,l)$ and $\hat{P}(k+l)$ are also zero. If, on the other hand, $\Re[B_m(k,l)]$ is trimmed but $\Im[B_m(k,l)]$ is not, then the segment m does contribute something to the numerator estimate $\hat{B}_\alpha$, and so it should also contribute something to the denominator estimates $\hat{P}_\alpha$ and $\hat{S}_\alpha$.

Although this method seems to work well, it is not satisfactory from a mathematical perspective, and finding a better way of doing this is a topic of current research.

2.2 An efficient implementation

The algorithm presented above can be very memory consuming, since the raw bispectral estimate for every segment has to be stored before the order statistics can be computed. Since the trimming factor α is typically small ($0 \leq \alpha \leq 0.2$) a more efficient algorithm can be constructed by rewriting the estimation equation; instead of a summation over the segments which are *not* in the distribution tails (as in Eqn. 5), rewrite this as a summation over *all* segments followed by a subtraction of the tail values. Using this implementation, only the $2g+2$ (real and imaginary) tail values need to be stored, resulting in a large saving in memory needs. However, the new implementation requires a local sort[1] on each segment in turn.

The form for this algorithm becomes evident by writing

$$\sum_{m=1}^{K} X_m = \sum_{m=1}^{g+1} X_m + \sum_{m=g+2}^{K-g-1} X_m + \sum_{m=K-g}^{K} X_m, \quad (6)$$

and subsituting for the summation $\sum_{m=g+2}^{K-g-1}$ in Eqn. 5.

The estimate is then written as

$$\begin{aligned} \hat{B}_\alpha \;=\; & \frac{1}{K(1-2\alpha)} \Bigg\{ \sum_{m=1}^{K} B_m \\ & - \; r\left[\Re[B_{r_{g+1}} + B_{r_{K-g}}] + \right. \\ & \left. j\Im[B_{i_{g+1}} + B_{i_{K-g}}]\right] \\ & - \; \sum_{m=1}^{g} \Re[B_{r_m}] + j\Im[B_{i_m}] \\ & - \; \sum_{m=K-g+1}^{K} \Re[B_{r_m}] + j\Im[B_{i_m}] \Bigg\}. \end{aligned} \quad (7)$$

Four vectors $\mathbf{R}^{(LH)}$, $\mathbf{R}^{(RH)}$, $\mathbf{I}^{(LH)}$ and $\mathbf{I}^{(RH)}$ each store $g+1$ extreme values of the real and imaginary parts of the raw estimates as $m = 1,..K$. For example, $\mathbf{R}^{(LH)} = [R_1^{(LH)}, .., R_{g+1}^{(LH)}]^T$, stores the left-hand extreme values (i.e. values in the left hand tail of the distribution of $\Re[B_m(k,l)]$). When $B_m(k,l)$ is calculated for a new segment m, $\Re[B_m(k,l)]$ is placed as the $(g+2)$th element of $\mathbf{R}^{(LH)}$. A sort is carried out, so that $R_1^{(LH)} = \min_{\tilde{m}=1,..,g+2} R_{\tilde{m}}^{(LH)}$ and $R_{g+2}^{(LH)} = \max_{\tilde{m}=1,..,g+2} R_{\tilde{m}}^{(LH)}$. The value in $R_{g+2}^{(LH)}$ is discarded. When all K segments have been processed in this way, these four vectors $\mathbf{R}^{(LH)}$, $\mathbf{R}^{(RH)}$, $\mathbf{I}^{(LH)}$ and $\mathbf{I}^{(RH)}$ will contain the quantities needed to calulated the trimmed mean estimate from Eqn. 7.

Since α is typically about 0.05 this represents a storage saving of roughly 80%[2] over the standard method of computing the trimmed mean.

3 Results

We propose that the modified bicoherence, described above, be used as a detector of Quadratic Phase Coupling (QPC) in signal processing environments influenced by background noise and transients.

In order to see how well this measure works in practice, several simulation signals have been analysed. In common with other simulations used to measure the performance of the bicoherence as a QPC detector, the signal $[m(n) = x(n)+v(n)]$ is modelled as the summation of an underlying sinusoidal component $[x(n)]$ and an additive disturbance $[v(n)]$. $x(n)$ exhibits QPC, but this may be difficult to detect with the ordinary bicoherence because of extraneous noise.

The signal of length N is generated segment by segment as follows;

$$x(n) = \sum_{j=1}^{3} \cos(2\pi f_j n + \phi_j), \quad (8)$$

with $f_3 = f_1 + f_2$ and $\phi_3 = \phi_1 + \phi_2$. The phases ϕ_1, ϕ_2 are re-randomised $U[0, 2\pi)$ in each frame (this satisfies the Phase Randomisation Assumption which renders the bicoherence *magnitude* suitable for QPC detection[2]).

In previous applications of the bicoherence [4, 1] $v(n)$ was white Gaussian noise. In this paper, $v(n) = t(n)+g(n)$ is a summation of randomly occurring short-lived transients (modelled by damped sinusoids) and white Gaussian noise;

$$v(n) = \sum_{j=1}^{N_t} \cos(2\pi h_j n + \theta_j) e^{-\eta_j n} + g(n), \quad (9)$$

in which $h_r = U[0, 0.5)$, $\theta_j = U[0, 2\pi]$,$\eta = U[0, 1)$ and $g(n)$ is white Gaussian noise. The transients are triggered randomly, with the probability of a transient beginning at any one time sample is controlled by a paramter γ.

[1] A simple bubble sort was used in the current work.

[2] based on an assumption that $K = 64, \alpha = 0.05, g = 3$, the saving is $1 - (2g+2)/K = 87\%$.

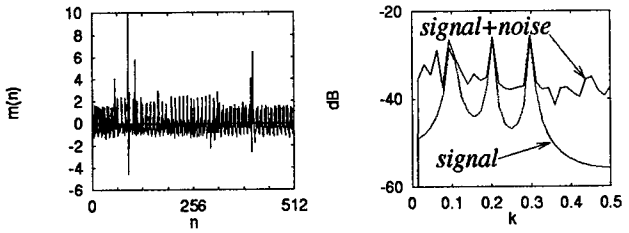


Figure 1. *Left: Time series of $m(n)$. Right: Power spectra of $m(n)$ (signal plus noise) and $x(n)$ (clean signal no noise).*

α	$b^2_\alpha(0.2, 0.1)$	$\beta(\%)$	notes
0	0.82	1.7	standard method
0.05	0.82	3.2	$\approx$ 2 segments discarded
0.20	0.84	7.3	$\approx$ 6 segments discarded

Table 1. *Performance measure of new technique.*

Fig. 1 shows the time series of one example of the noisy signal $m(n)$ ($N = 1024$, $\gamma = 0.02$, SNR= $10\log_{10}\sigma_x^2/\sigma_v^2 = 0$dB), with $f_1 = 0.1, f_2 = 0.2$, together with the Power Spectrum ($N_{DFT} = 64, K = 16$) for both the signal with no noise [$x(n)$], and the signal with the transient and steady state noise added [$m(n) = x(n) + v(n)$]. Clearly the noise has a very detrimental effect on the power spectrum, almost obscuring the spectral peaks. Fig. 2 shows the squared bicoherence of $m(n)$ with different levels of $\alpha-$trimming. The bicoherence should peak at $(0.2, 0.1)$ (which is equivalent to $(0.1, 0.2)$ because of symmetry). The top plot $\alpha = 0$ corresponds to the ordinary squared bicoherence estimate - the peak at $(0.2, 0.1)$ is barely visible above the noise floor. However, the $\alpha-$trimmed estimates show much lower noise floors.

The improvement in performance can be measured by $\beta = b^2(0.2, 0.1)/\sum_{IT} b^2(k, l) \times 100\%$ - the percentage of total bicoherence "energy" which occurs in the correct bin[3]. Better QPC detectors will have higher values of β. Table 1 shows how this varies for a typical example of this simulation. It is clear that the $\alpha-$trimmed estimates perform better as QPC detectors than the ordinary squared bicoherence. Further simulation results will be shown at the conference.

4 Discussion and Conclusions

The proposed QPC detector based on an $\alpha-$trimmed bispectral estimate appears to give reduced noise floors in the simulations investigated so far, and peaks due to QPC are easier to pick out using this detector than using the standard bicoherence. In particular the new detector is robust to interference from additive transients. The normalisation scheme used in this paper appears to work successfully,

[3] $\sum_{IT}$ denotes a summation over the Inner Triangle[3].

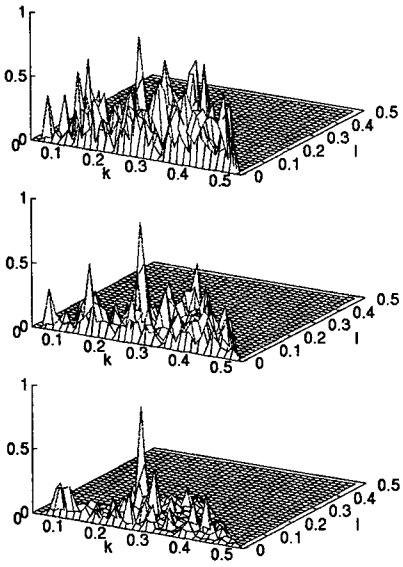


Figure 2. *Squared bicoherences of $m(n)$. Top: Ordinary (standard approach). Middle: $\alpha = 0.05$. Bottom: $\alpha = 0.20$.*

although it does not have a rigorous mathematical basis. The performance of the new estimator as a QPC detector for other types of interference (such as Amplitude Modulation) is a topic of current work, which will also be described at the conference.

References

[1] S. Elgar and G. Sebert. Statistics of bicoherence and biphase. *Journal of Geophysical Research*, C94:10993–10998, 1989.

[2] J. W. A. Fackrell, S. McLaughlin, and P. R. White. Bicoherence estimation using the direct method : Part 1 - theoretical considerations. accepted for publication in Applied Signal Processing, 1996.

[3] M. J. Hinich. Testing for Gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis*, 3(3):169–176, 1982.

[4] Y. C. Kim and E. J. Powers. Digital bispectral analysis and its applications to nonlinear wave interactions. *IEEE Transactions on Plasma Science*, PS-7(2):120–131, 1979.

[5] A. K. Nandi. Robust estimation of third-order cumulants in applications of higher-order statistics. *IEE PROCEEDINGS F. Radar and Signal Processing*, 6:380–389, December 1993.

[6] J. L. Rosenberger and M. Gasko. Comparing location estimators : Trimmed means, medians and trimean. In D. C. Hoaglin, F. Mosteller, and J. W. Tukey, editors, *Understanding Robust and Exploratory data analysis*, pages 297–338. John Wiley, New York, 1983.

The Asymmetric Generalized Gaussian Function: a New HOS-Based Model for Generic Noise Pdfs

A. Tesei, and C.S. Regazzoni
DIBE - University of Genoa, Italy - e-mail: tale@dibe.unige.it

Abstract

The work is addressed to provide realistic modelling of generic noise probability density functions (pdfs), in order to optimize signal detection in non-Gaussian environments. The target is to obtain a model depending on few parameters (quick and easy to estimate), and so general to be able to describe many kinds of noise (e.g., symmetric or asymmetric, with variable sharpness). To this end, a new HOS-based model is introduced, which derives from the generalized Gaussian function and depends on three parameters: kurtosis (fourth order), for representing variable sharpness, and left and right variances (whose combination provides the same information of skewness - third order) for describing deviation from symmetry. The model is applied in the design of a LOD test for detecting signals corrupted by real underwater acoustic noise in a low-frequency range.

1. Introduction

Realistic and simple statistical modelling of generic background noise is addressed in order to optimize *signal detection* in *non-Gaussian environments*. Detection purpose is to decide between the two hypotheses of the presence (H_1) or the absence (H_0) of a transmitted deterministic signal $\{s_k, k=1, .., K\}$ (the approach can be extended to the stochastic case), on the basis of acquired observations $\{y_k, k=1, .., K\}$ (application of *binary hypothesis testing*) [1]; the noise, $\{n_k, k=1, .., K\}$ corrupting the signal during the propagation is assumed additive, independent and identically distributed, stationary, and generally non-Gaussian and unimodal.

The work main target is to design a detector characterized by: *(a)* high performances in the case of weak signals; *(b)* easy applicability to real cases (in particular, easy and realistic estimation of needed parameters, realistic noise modelling, and robustness to variable boundary conditions); *(c)* algorithmical simplicity.

Detection optimization in the case of low/middle values of the Signal-to-Noise Ratio (SNR) (in the range [-30,0] dB) (property *(a)*), is reached by selecting the *Locally Optimum Detector* (LOD) [1] as statistical inference approach.

For satisfying conditions *(b)* and *(c)*, the investigation is addressed to express generalized noise pdf models, usually depending on parameters difficult to be estimated from real data samples, in terms of Higher-Order-Statistics (HOS) parameters [2], which are very *easy and quick* to be extracted from data and are particularly suitable for quantifying deviation from Gaussianity in terms of asymmetry (with third-order parameters) and variable sharpness (with fourth-order parameters).

As conventional signal processing algorithms based on the Second Order Statistics, optimized in presence of Gaussian noise, may decay in non-Gaussian noise, various works used HOS theory [2] as signal-processing basis for noise analysis and detection optimization; however, some methods work only with non-Gaussian signals [3][4][5] or only in Gaussian noise [5][6][7]; some can be applied only under certain assumptions of noise distributions [8][9]; some are not optimized in the case of weak signals [3]; finally some algorithms are complicated [8].

In order to overcome at least some of the aforesaid limitations and improve robustness, simplicity and generality of HOS-based detectors, the parametric *asymmetric generalized Gaussian* pdf model is introduced. It derives from the combination of the well-known generalized Gaussian pdf [10] and of the asymmetric Gaussian model presented in [11].

The first model is symmetric and depends on a real parameter, c, which is not easy to estimate from data. Nevertheless, c presents a physical meaning, as linked with the pdf *sharpness*. The HOS parameter which better describes sharpness variability is the fourth-order *kurtosis*, β_2. Hence the analytical relationship between c and β_2 is introduced (see [12] for details). The resulting symmetric function based on kurtosis has the same characteristics of the generalized Gaussian, and is a realistic noise-pdf model for $1.865<\beta_2\leq 30$.

In order to introduce into this variable-sharpness model also possible deviation from symmetry, the resulting kurtosis-based function is modified by taking into account

0-8186-7576-4/96 $5.00

the asymmetric Gaussian model [11]. It directly derives from the Gaussian shape, but is asymmetric and depends on two second-order parameters, the left and right variances [11]. By introducing these two parameters in the kurtosis-based generalized Gaussian function, the "asymmetric generalized Gaussian" model can be obtained.

The new model is compared with the generalized Gaussian and the asymmetric Gaussian pdfs, which result as its particular cases. It is applied in the design of a LOD test, used for detecting deterministic signals corrupted by real underwater acoustic noise radiated by ship traffic [13].

2. The asymmetric generalized Gaussian pdf

In the context of noise modelling, one of the most noticeable ways in which estimated noise distributions deviate from Gaussianity is in kurtosis β_2, i.e., the ratio of the fourth and the square of the second central moments. It is equal to 3 in the Gaussian case; the sharpness of the pdf shape is higher (lower) than the corresponding Gaussian function as β_2 is larger (smaller) than 3. A good model for general pdfs has variable sharpness.

One of the well-known symmetric pdf models is the generalized Gaussian, which depends on the parameter c:

$$p_{gG}(n) = \frac{\gamma c}{2\Gamma(1/c)} e^{-|\gamma(n-\mu)|^c} \tag{1}$$

where $\{n\}$ is generic noise with mean value μ and variance

$$\sigma^2,\ \gamma = \sqrt{\frac{\Gamma(3/c)}{\sigma^2\Gamma(1/c)}},\ \Gamma(k) = \int_0^{+\infty} e^{-x}x^{k-1}dx.$$

c cannot be directly estimated from data samples; hence the relationship between c and β_2 was found [12]. It derives from the β_2 definition and is expressed by the following formula:

$$\beta_2^x = \frac{m_4^x}{(m_2^x)^2} = \frac{E\{(n-\mu)^4\}}{\left(E\{(n-\mu)^2\}\right)^2} = \frac{\Gamma(5/c)\Gamma(1/c)}{(\Gamma(3/c))^2} \Rightarrow$$

$$c = c(\beta_2) \approx \sqrt{\frac{5}{\beta_2 - 1.865}} - 0.12 \text{ for } 1.865<\beta_2\leq 30. \tag{2}$$

This formula allows one to express $p_{gG}(n)$ in terms of β_2 [12]. Its validity is confirmed by observing that for $\beta_2>3$ the resulting pdf has *heavy tails*, as expected [10].

In order to generalize this model so that it can be also asymmetric, the asymmetric Gaussian model presented in [11] is taken into account. It depends on *two second-order parameters* (deriving from the definition of variance), σ_l^2 and σ_r^2, called respectively "left" and "right variances" and estimated from finite sequences of the process $\{n\}$ according to the following formulas:

$$\sigma_l^2 = \frac{1}{N_l - 1}\left(\sum_{k=1,\, n_k<\mu}^{N_l} (n_k - \mu)^2\right) \quad \text{and}$$

$$\sigma_r^2 = \frac{1}{N_r - 1}\left(\sum_{k=1,n_k>\mu}^{N_r} (n_k - \mu)^2\right) \tag{3}$$

where N_l (N_r) is the number of n_k samples $<\mu$ ($>\mu$). The model expression follows:

$$p_{aG}(n) = \begin{cases} \dfrac{2}{\sqrt{2\pi}(\sigma_l + \sigma_r)} e^{-\frac{(n-\mu)^2}{2\sigma_l^2}} & n < \mu \\ \dfrac{2}{\sqrt{2\pi}(\sigma_l + \sigma_r)} e^{-\frac{(n-\mu)^2}{2\sigma_r^2}} & n \geq \mu \end{cases} \tag{4}$$

As well as the kurtosis-based generalized Gaussian model, it is analytically simple and easy to be estimated if some data sequences are available (the model includes the Gaussian case for $\sigma_l^2 = \sigma_r^2$). The left and right variances are linked with the variance (the well-known second-order parameter) and with the skewness (the third-order parameter describing and quantifying pdf asymmetry) as follows [11]:

$$\sigma^2 = \sigma_l^2 + \sigma_r^2 - \sigma_l\sigma_r$$

$$m_3^n = \beta_1 = E\{(n-\mu)^3\} = \frac{4}{\sqrt{2\pi}(\sigma_l + \sigma_r)}(-\sigma_l^4 + \sigma_r^4) \tag{5}$$

(where E is the expectation value).

In a similar way, these two parameters are introduced in the kurtosis-based generalized Gaussian pdf in order to transform it into the following asymmetric generalized Gaussian model:

$$p_{agG}(n) = \begin{cases} \dfrac{c\gamma_a}{\Gamma(1/c)} e^{-\gamma_l^c[-(n-\mu)]^c} & n < \mu \\ \dfrac{c\gamma_a}{\Gamma(1/c)} e^{-\gamma_r^c[(n-\mu)]^c} & n \geq \mu \end{cases} \tag{6}$$

where $\gamma_a = \dfrac{1}{\sigma_l + \sigma_r}\left(\dfrac{\Gamma(3/c)}{\Gamma(1/c)}\right)^{1/2}$, $\gamma_l = \dfrac{1}{\sigma_l}\left(\dfrac{\Gamma(3/c)}{\Gamma(1/c)}\right)^{1/2}$,

$$\gamma_r = \frac{1}{\sigma_r}\left(\frac{\Gamma(3/c)}{\Gamma(1/c)}\right)^{1/2}.$$

It is easy to notice that if $\sigma_l^2=\sigma_r^2$, then the pdf coincides with the generalized Gaussian, hence it is symmetric; if $\sigma_l^2=\sigma_r^2$ and $\beta_2=3$, then it coincides with the Gaussian model. Figure 1 presents a family of the pdf as β_2 varies.

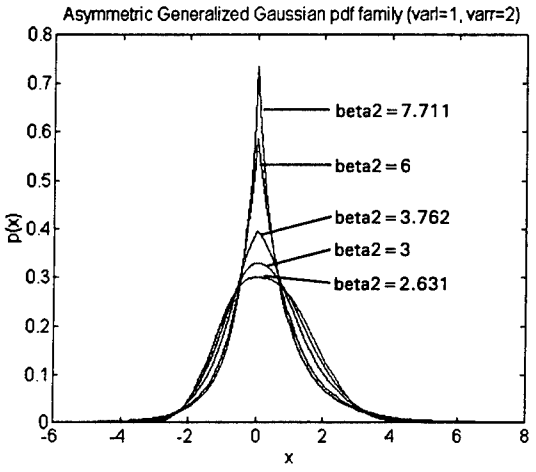


Fig. 1. Asymmetric generalized Gaussian family ($\mu=0$).

3. The LOD test designed on the basis of the new model

The model is suitable for the design of a LOD test [1], as the non-linearity $g_{LO}(.)$ and the maximum asymptotic relative efficiency ρ can be expressed in terms of elementary functions. In particular:

$$g_{LO}(y)=\begin{cases}-c\gamma_l[-(y-\mu)]^{c-1} & y<\mu\\ c\gamma_r(y-\mu)^{c-1} & y\geq\mu\end{cases} \quad (7)$$

$$\rho=\frac{\gamma_a c^2\Gamma(2-1/c)(\gamma_l+\gamma_r)}{\Gamma(1/c)}(\sigma_l^2+\sigma_r^2-\sigma_l\sigma_r) \quad (8)$$

The respective graphs are presented in Figs. 2 and 3. From their analysis it is easy to conclude that the test works better for values of the kurtosis larger than 3 (i.e., for super-Gaussian and, in particular, for impulsive noise pdfs), as expected; nevertheless it can reach good performances even in more critical conditions of sub-Gaussian noise.

In order to deduce test performances *from a theoretical point of view*, these graphs can be compared with similar graphs for other well-known pdfs: for example, LOD non linearities and maximum ARE curves computed in terms of the *c*-based generalized Gaussian, the generalized Cauchy, the generalized beta functions are presented in [10], those expressed in terms of the asymmetric Gaussian and the kurtosis-based generalized Gaussian pdfs are shown in [12][14].

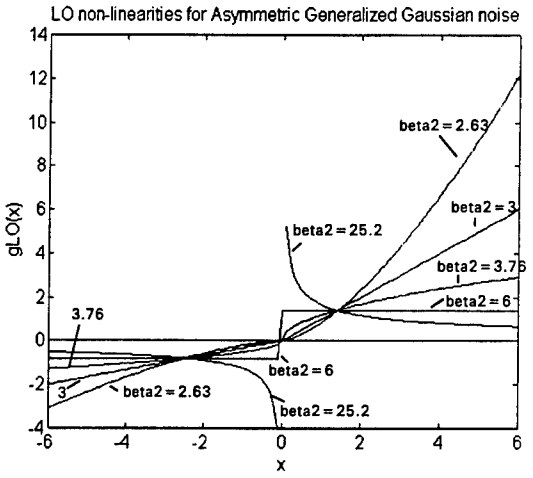


Fig. 2. $g_{LO}(.)$ graphs as β_2 varies -$\sigma_l^2=3$, $\sigma_r^2=1$.

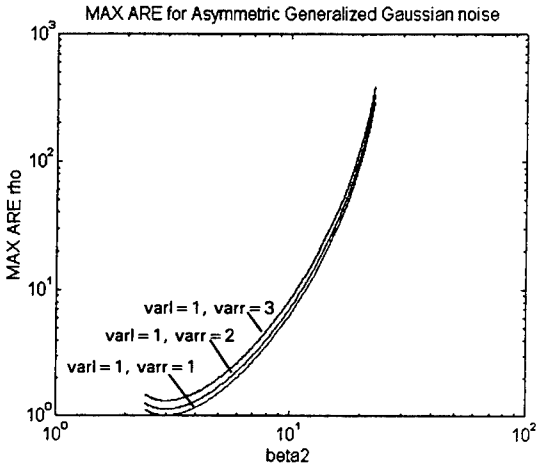


Fig. 3. ρ graphs as β_2 and σ_r^2 varies.

4. Experimental results on real data

From an experimental point of view, the capability of the proposed model of describing realistically generalized noise pdfs was evaluated by applying it to the problem of detecting known constant signals corrupted by underwater acoustic ship-traffic-radiated noise [13]. The noise data sequences were analysed and characterized at average by the estimated parameters μ=-17.5, β_2=2.51, σ_l= 1550 and σ_r=1350 [13]. A comparison between noise histogram (computed on 10 records of 10000 samples) and the new pdf model estimated on the basis of the aforesaid second- and fourth-order parameters can be deduced from Figs. 4 and 5: a good fitness between data and model is shown.

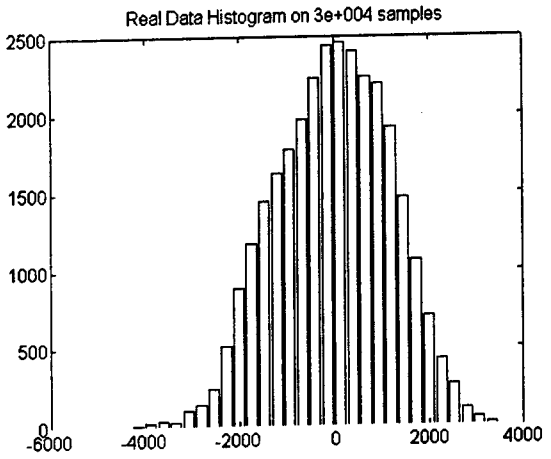


Fig. 4. The non-normalized data histogram.

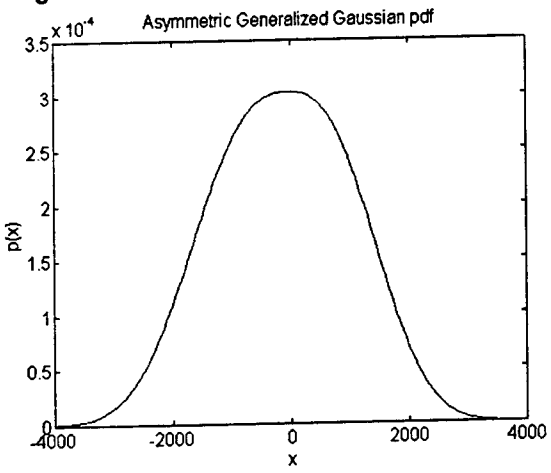


Fig. 5. The asymmetric generalized Gauss. pdf.

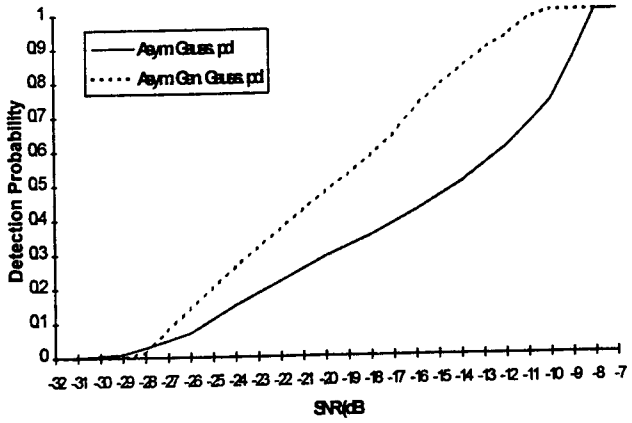


Fig. 6. Comparison of test results. $P_{FA} = 0.05$.

The detection performances obtained by applying the HOS-based model to a LOD test are presented by means of experimental curves of the Detection Probability P_D as SNR varies, given a certain value of the Probability of False Alarm, P_{FA}. In the diagram of Fig. 6 the performances obtained by using the new pdf model (depending on the left and right variances and on the kurtosis) are compared with those provided by an asymmetric Gaussian model (depending only on the left and right variances).

The performance improvement is mainly associated to the capability of the new pdf to model non-Gaussian noise in a more realistic way, so that it can be better filtered.

References

[1] S.A. Kassam, *Signal Detection in Non-Gaussian Noise*, Springer Verlag, Berlin, 1988.

[2] C. Nikias, J. Mendel, "Signal Processing with Higher-Order Spectra," *IEEE SP Magazine*, pp. 10-37, 1993.

[3] M.J. Hinich, G.R. Wilson, "Detection of non-Gaussian signals in non-Gaussian noise using the bispectrum," *IEEE Trans. on ASSP*, 38 (7), pp. 1126-1131, July 1990.

[4] D. Kletter and H. Messer, "Suboptimal detection of non-Gaussian signals by third-order spectral analysis," *IEEE Tr. on ASSP*, 38 (6), pp. 901-909, June 1990.

[5] Sangfelt E., Persson L., "Experimental performance of some Higher-Order cumulant detectors for hydroacoustic transients," *Proc. of IEEE HOS Workshop*, pp. 182-186, June 1993.

[6] Giannakis G.B., Tsatsanis M.K., "Signal detection and classification using matched filtering and Higher order Statistics," *IEEE Trans. on SP*, 38 (7), pp. 1284-1296, July 1990.

[7] R.F. Dwyer, "Use of the kurtosis statistics in the frequency domain as an aid in detecting random signals," *IEEE J. of Ocean. Eng.*, OE-9 (2), pp. 85-92, April 1984.

[8] R.J. Webster, "Ambient noise statistics," *IEEE Trans. on SP*, 41 (6), pp. 2249-2253, 1993.

[9] A. Tesei, C.S. Regazzoni, "Signal detection in non-Gaussian noise by a kurtosis-based probability density function model," *IEEE Workshop on HOS*, pp. 162-165, June 1995.

[10] J.H. Miller, J.B. Thomas "Detectors for discrete-time signals in non-Gaussian noise," *IEEE Trans. on IT*, 18 (2), pp. 241-250, 1972.

[11] A. Tesei, C.S. Regazzoni, "HOS-based noise models for signal-detection optimization in non-Gaussian environments", *Proc. of IEEE ISIT'95*, Whistler, Sept. 1995, p. 296.

[12] A. Tesei, C.S. Regazzoni, "Use of fourth-order statistics in non-Gaussian noise modelling for signal detection: the Generalized Gaussian pdf in terms of kurtosis," *Proc. of EUSIPCO-96*, Sept. 1996 (in press).

[13] C.S. Regazzoni, A. Tesei, G. Tacconi, "A comparison between spectral and bispectral analysis for ship detection from acoustical time series," *IEEE Proc. of ICASSP'94*, pp. 289-292, April 1994.

[14] A. Tesei, and C.S. Regazzoni, "Application to locally optimum detection of a new noise model," *IEEE Proc. of ICASSP'96*, May 1996 (in press).

Composite Cumulant-Property Mapping for MA Cumulant Matching

Achilleas G STOGIOGLOU and Steve McLAUGHLIN
Dept. of Electrical Eng., The University of Edinburgh, Edinburgh EH9 3JL,
Scotland, U.K., Tel/Fax: +44 [131] 650 5660 / 650 6554. E-Mail: ashil@ee.ed.ac.uk

Abstract

This paper considers the use of composite property mappings for MA cumulant matching. The algorithm makes use of two property mappings corresponding to rank and structure properties of a matrix consisting of MA cumulants. It is proved that these two properties are sufficient to characterise a matrix consisting of true MA cumulants. This result clearly implies that provided that convergence is achieved, the composite property mapping algorithm performs some kind of cumulant matching. The issue of convergence is also discussed in the paper. Numerical results are presented to show the performance of the algorithm. **Keywords:** *Higher Order Statistics, System Identification.*

1 Introduction

Consider the following finite impulse response (FIR) signal model: $x(t) = \sum_{i=0}^{q} h(i)w(t-i)$ where the system input is assumed to be non-Gaussian, independent identically distributed (IID) , random process with $E\{w(t)\} = 0$, $E\{w(t)w(t+n)\} = \beta_2\delta(n)$, and $E\{w(t)w(t+n_1)w(t+n_2)\} = \gamma_3\delta(n_1, n_2)$. We assume that $h(0), h(q) \neq 0$. A method for the enhancement of third order cumulants of MA models was presented in [1]. That method is based on the use of Composite Property Mapping Algorithms (CPMA) [3]. The reader is referred to [3] for further information on CPMA and to [5] for a general introduction to set theoretic estimation. Composite property mapping algorithms have originally been used within a HOS framework in [4] for blind array processing. This is a follow-up to the work described in [1]. Some of the material presented here is also included in [2].

2 Cumulant Enhancement

In the following we summarise the main steps involved in the method of [1]:

$$\widehat{C}_0 = \begin{bmatrix}
\hat{c}_{0,q} & \hat{c}_{1,q} & \cdots & \hat{c}_{q-1,q} & \hat{c}_{q,q} & 0 & 0 & \cdots & 0 \\
0 & \hat{c}_{0,q} & \hat{c}_{1,q} & \cdots & \hat{c}_{q-1,q} & \hat{c}_{q,q} & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \cdots & & \ddots & \ddots & & \vdots \\
0 & \cdots & 0 & \hat{c}_{0,q} & \hat{c}_{1,q} & \cdots & \hat{c}_{q-1,q} & \hat{c}_{q,q} & 0 \\
0 & \cdots & 0 & 0 & \hat{c}_{0,q} & \hat{c}_{1,q} & \cdots & \hat{c}_{q-1,q} & \hat{c}_{q,q} \\
\hat{c}_{1,q-1} & \hat{c}_{0,q-1} & \hat{c}_{1,q-1} & \cdots & \hat{c}_{q-1,q-1} & \hat{c}_{q,q-1} & 0 & \cdots & 0 \\
\hat{c}_{2,q-2} & \hat{c}_{1,q-2} & \hat{c}_{0,q-2} & \cdots & \hat{c}_{q-1,q-2} & \hat{c}_{q,q-2} & \cdots & & 0 \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
\hat{c}_{1-q,1} & \hat{c}_{2-q,1} & \hat{c}_{3-q,1} & \cdots & \hat{c}_{q-3,1} & \hat{c}_{q-2,1} & \hat{c}_{q-1,1} & \hat{c}_{q,1} & 0 \\
\hat{c}_{q,0} & \hat{c}_{q+1,0} & \hat{c}_{q+2,0} & \cdots & \hat{c}_{q-4,0} & \hat{c}_{q-3,0} & \hat{c}_{q-2,0} & \hat{c}_{q-1,0} & \hat{c}_{q,0}
\end{bmatrix}$$

Figure 1. *The cumulant matrix used in the CPMA algorithm. It is assumed that* $c_{\tau_1,\tau_2} = c_{3,x}(\tau_1, \tau_2)$.

1. Collect the sample cumulants corresponding to the minimal sufficient set of lags in the following vector:

$$\hat{\theta}_0 = [\hat{c}_{3,x}(0,0), \hat{c}_{3,x}(0,1), \hat{c}_{3,x}(1,1), \ldots, \hat{c}_{3,x}(0,q), \ldots, \hat{c}_{3,x}(q-1,q), \hat{c}_{3,x}(q,q)]^{\mathsf{T}}. \quad (1)$$

2. The elements of $\hat{\theta}_0$ are then used to build the matrix $\hat{\mathbf{C}}_0$ depicted in figure 1 $\hat{\mathbf{C}}_0$ is a $(2q+1) \times (2q+1)$ matrix. The matrix $\hat{\mathbf{C}}_0$ contains all the third order cumulants in the vector $\hat{\theta}_0$. It can be shown that a matrix C which results from matrix $\hat{\mathbf{C}}_0$ after replacing sample cumulants with true MA cumulants, possesses the following two theoretical properties:
   (a) $rank(\mathbf{C}) = q+1$
   (b) The structural composition of C is determined by a characteristic matrix A and we say that C is a linear structured matrix. (More details in [1])
3. We perform the following iteration: $\hat{\mathbf{C}}_{k+1} = \mathcal{F}(\hat{\mathbf{C}}_k) = \mathcal{F}_{\mathbf{A}}(\mathcal{F}_{q+1}(\hat{\mathbf{C}}_k))$. until $1 - r_{q+1}(\hat{\mathbf{C}}_{k+1}) \leq \epsilon$, where ϵ is a predefined small positive number. The

0-8186-7576-4/96 $5.00

mappings $\mathcal{F}_{\mathbf{A}}$, $\mathcal{F}_{q+1}$ and the quantity $r_{q+1}(\hat{\mathbf{C}}_{k+1})$ are defined as follows:

(a) The mapping $\mathcal{F}_{q+1}$ corresponds to the rank property of C. It is implemented using SVD (SVD-based rank reduction) $\mathcal{F}_{q+1}(\mathbf{X}) = \sum_{k=1}^{q+1} \sigma_k \mathbf{u}_k \mathbf{v}_k^\top$. where $\mathbf{X} \in \mathbf{R}^{(2q+1)\times(2q+1)}$ and $\mathbf{X} = \sum_{k=1}^{2q+1} \sigma_k \mathbf{u}_k \mathbf{v}_k^\top$.

(b) The mapping $\mathcal{F}_{\mathbf{A}}$ corresponds to the linear-structure property of C: $\mathcal{F}_{\mathbf{A}}(\mathbf{X}) = T^{-1}(\mathbf{A}[\mathbf{A}^\top\mathbf{A}]^{-1}\mathbf{A}^\top\ T(\mathbf{X}))$.

(c) It is possible to get an idea of how close a matrix $\mathbf{X}$ is to a $q+1$-rank matrix, by examining how close the following quantity is to 1, provided that $\sigma_{q+1} > \sigma_{q+2}$:

$$r_{q+1}(X) = \frac{\sum_{k=1}^{q+1} \sigma_k}{\sum_{k=1}^{2q+1} \sigma_k}. \tag{2}$$

3 CPMA for Cumulant Matching

In [1] this composite property mapping method was used as a preprocessing step before applying some linear methods for system identification. A question that rises naturally, is whether a matrix that possesses both properties defined in the previous section, contains true cumulants of some MA model.

To start with we assume that the matrix sequence generated by the iterative algorithm described earlier converges to a matrix S, which has both the desired structure and rank properties. It is interesting to examine whether this matrix consists of real cumulants of some MA model. Since the matrix S has the same structural characteristics as those of a matrix constructed of real cumulants , then if $s_{i,j}$ is a nonzero element of S, we denote $s_{i,j}$ as $s(\tau_1, \tau_2)$, where (τ_1, τ_2) are the *lags* we associate with the i,j-element of a structurally equivalent matrix which is constructed from real cumulants. Then because of the structure property, the same symmetries that apply to lags of cumulants will apply to these associated *lags* of $s(\tau_1, \tau_2)$. In the following it is assumed that $s(0,q), s(q,q) \neq 0$. The following Lemma is required:

Lemma 1 *Suppose that we are given a* $(2q+1)\times(2q+1)$ *matrix* S *, which has the two prescribed properties (structure and rank). Then the following equation holds for* $s(\tau_1, \tau_2)$*:*

$$s(j,n) = \sum_{i=0}^{q} \frac{s(i,q)s(i+n,q)}{s(0,q)s(q,q)} s(j+i,q) \tag{3}$$

for $n = 0, ..q-1$ *and* $j = n-q, ..., q$.

The proof of this Lemma is given in the Appendix.
Since we know that $s(0,q), s(q,q) \neq 0$ we can find a $\gamma_3 \neq 0$ such that

$$s(i,q) = \gamma_3 \frac{s(0,q)}{s(0,q)} \frac{s(q,q)}{s(0,q)} \frac{s(i,q)}{s(0,q)}. \tag{4}$$

If we combine equations (3) and (4), we obtain the following:

$$\begin{aligned} s(j,n) &= \sum_{i=0}^{q} \frac{s(i,q)s(i+n,q)}{s(0,q)s(q,q)} \gamma_3 \frac{s(0,q)}{s(0,q)} \frac{s(q,q)}{s(0,q)} \frac{s(i+j,q)}{s(0,q)} \\ &= \gamma_3 \sum_{i=0}^{q} \frac{s(i,q)}{s(0,q)} \frac{s(i+n,q)}{s(0,q)} \frac{s(i+j,q)}{s(0,q)}. \end{aligned} \tag{5}$$

Equation (5) shows that $s(j,n)$ is the third order cumulant of an MA model with parameters $h(i) = s(i,q)/s(0,q)$. Thus the following theorem holds:

Theorem 1 *Every* $(2q+1)\times(2q+1)$ *matrix* S *possessing the structure and rank properties defined in this section, consists of real cumulants of some MA(q) model.*

The above theorem implies, that the cumulant enhancement method summarised earlier, when it converges to a matrix with the prescribed properties, performs some kind of cumulant matching.

4 Convergence Properties

An important issue that needs to be addressed here is that of convergence. Let us consider a matrix sequence generated according to the algorithmic rule,

$$\mathbf{C}_k = \mathcal{F}(\mathbf{C}_{k-1}), \quad \text{for} \quad k \geq 1 \tag{6}$$

in which the initial matrix $\mathbf{C}_0$ is the experimentally generated matrix $\hat{\mathbf{C}}_0$. Then at every step, the matrix $\mathbf{C}_k$ has the right structure and is "nearer" to a matrix with rank $q+1$ than $\mathbf{C}_{k-1}$. In iterative mappings of this type, convergence is guaranteed only when all property sets are convex [5]. However, in our case, it is obvious that the set of matrices with rank $q+1$ is not convex, and this violates the assumptions required for Theorem (1) in [3]. In [7], Dologlou *et al* provide an interesting theorem which shows that the norm of the difference $\mathbf{C}_k - \mathbf{C}_{k+1}$ and the distance between $\mathbf{C}_k$ and the set of matrices with rank less than or equal to $q+1$ both converge to 0 when $k \to \infty$. The behavior predicted by this theorem is verified by our numerical simulations (summarised in figs.2 and 3). For system identification the following methods are used: The Least Squares method of [2] (LS), the Closed Formula (CF) of [6] and a nonlinear method for cumulant matching [8]. It has been observed though, that for some MA models convergence in the sense described in

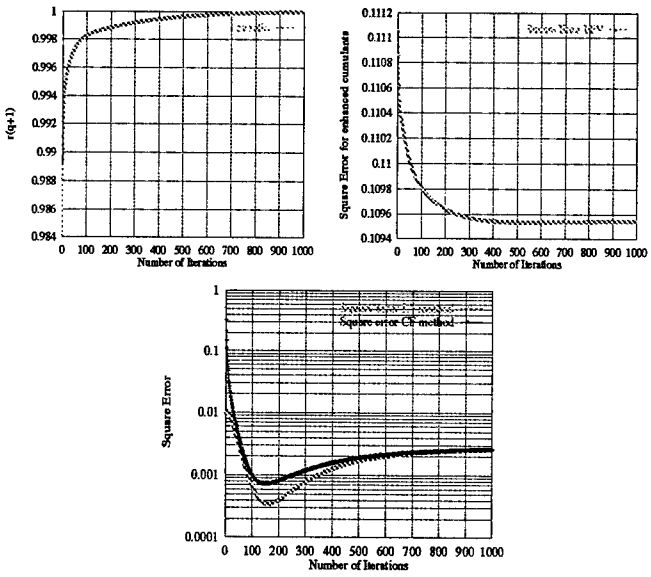


Figure 2. *The singular value ratio (2), the square error of the enhanced cumulants and the square error of the estimated parameters as a function of number of iterations. The dark curve represents the CF method [6] and the light curve represents the LS method [2].*

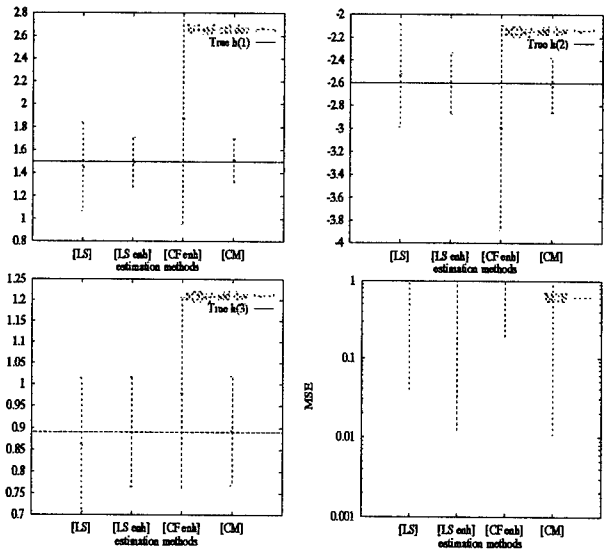


Figure 3. *System identification results after 100 Monte Carlo runs. Length of output sequence: 3000 samples. The horizontal lines represent the true values of the system parameters (model order=3). The vertical bars represent the average estimate +/- standard deviation. The last graph shows the MSE of the estimation.*

[7] can some times require thousands of iterations. Consequently there are cases where practically we cannot achieve convergence. In these cases, the algorithm can still be used for preprocessing the cumulants as proposed in [1]. Since convergence can sometimes be difficult to achieve, composite mapping algorithms cannot replace existing nonlinear methods for cumulant matching. The drawback of nonlinear cumulant matching methods is that there is always a danger of becoming trapped in a local minimum. However, in order to avoid this occurrence, good initial conditions are required and these are usually provided by linear methods. Because cumulant matching does not require any initial conditions, it can be applied[1] prior to linear methods in order to provide improved initial conditions for the nonlinear methods.

References

[1] A G Stogioglou , S McLaughlin. Third Order Cumulant Enhancement for MA Models. In *Proceedings of IEEE Workshop on HOS*, June 1995.

[2] A G Stogioglou, S McLaughlin. MA Cumulant Enhancement and Parameter Estimation. *IEEE Transaction on Signal Processing*, July 1996.

[3] J. Cadzow. Signal Enhancement- A Composite Property Mapping Algorithm. *IEEE Transactions on Acoustics, Speech and Signal Preocessing*, 36(1):49–62, 1988.

[4] J.-F. Cardoso. Fourth-Order Cumulant Structure Forcing. Application to Blind Array Processing. In *Proceedings IEEE SP Workshop on SSAP-92*, 1992.

[5] P. L. Combettes. The Foundations of set Theoretic Estimation. *Proceedings of the IEEE*, 81(2):181208, February 1993.

[6] G. Giannakis. Cumulants: A powerful Tool in Signal Processing. *Proceedings of the IEEE*, 75(9):1333–1334, September 1987.

[7] I Dologlou, J-C Pesquet and J Skowronski. Projection Based Rank Reduction Algorithms for Multichannel Modelling and Image Compression. *Subitted to Signal Processing*, 1995.

[8] K. Lii and M. Rosenblatt. Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes. *Ann. Statist*, 10:1195–1208, 1982.

[1] without necessarily achieving convergence

APPENDIX

Proof of Lemma 1: The vectors corresponding to the first $q+1$ rows of the matrix S are denoted as s_q^d where $d = 0, ..., q$ and the vectors corresponding to the last q rows of the matrix are denoted by $\mathrm{s}_{q-1}, ..., \mathrm{s}_0$. If we assume that $s(0,q), s(q,q) \neq 0$, then it is evident from their structure, that the $q+1$ vectors $\mathrm{s}_q^d, d = 0, ..., q$ are linearly independent. Given that the rank of the matrix is $q+1$, we can conclude that the vectors corresponding to the last q rows of the matrix, belong to the space spanned by the first $q+1$ rows. In particular since the last n elements of the vector s_n ($n = q-1, ..., 0$) are zero, it can easily be seen that they belong to the space spanned only by s_q^d for $d = 0, ..., q-n$. We can write this as

follows:

$$\mathbf{s}_n \in \text{span}\{\mathbf{s}_q^0, ..., \mathbf{s}_q^{q-n}\} \quad n = q-1, ..., 0. \tag{7}$$

Now if we take $n = q-1$, it is straight forward to prove that ,

$$\mathbf{s}_{q-1} = \frac{s(1,q)}{s(0,q)}\mathbf{s}_q^0 + \frac{s(q-1,q)}{s(q,q)}\mathbf{s}_q^1. \tag{8}$$

In scalar form, this translates to,

$$s(j,q-1) = \sum_{i=0}^{1} \frac{s(i,q)s(i+q-1)}{s(q,q)s(0,q)} s(j+i,q) \qquad j = -1, ..., q, \tag{9}$$

so equation (3) holds for $n = q-1$.

Assumption 1: *Let us suppose that equation (3) holds for every n such that $q-1 \geq n \geq k$, for some $k > 0$.*
We want to prove that it also holds for $n = k-1$.

$$\mathbf{s}_{k-1} = \sum_{i=0}^{q-k+1} \lambda_{k-1,i}\mathbf{s}_q^i. \tag{10}$$

Because of the cumulant like symmetries in the *lags* of $s(k-q-1, k-1)$, the first element of $\mathbf{s}_{k-1}$ is $s(k-q-1, k-1) = s(q-k+1, q)$. It is related with $s(0,q)$ as follows:

$$s(q-k+1, q) = \lambda_{k-1,0}\, s(0,q). \tag{11}$$

From equation (11) we can obtain the value of $\lambda_{k-1,0} = s(q-k+1,q)/s(0,q)$. Since $s(k-q-1, k-1) = (s(q-k+1,q)/s(0,q))s(0,q)$, equation (3) holds for $n = k-1$ and $j = k-q-1$.

Assumption 2: *Suppose that equation (3) holds for $n = k-1$ and $k-q-1 \leq j \leq m$ where $m \leq -2$.*
In other words we assume that we know that

$$s(m, k-1) = \sum_{i=-m}^{q-k+1} \frac{s(i,q)s(i+k-1,q)}{s(0,q)s(q,q)} s(i+m,q) = \sum_{i=-m}^{q-k+1} \lambda_{k-1,q-k+1-i}\, s(i+m,q). \tag{12}$$

We want to obtain the value of $\lambda_{k-1,m-k+q+2}$ and use this to show that equation (3) is valid for $n = k-1$ and $j = m+1$. So,

$$s(m+1, k-1) = \sum_{i=-m}^{q-k+1} \frac{s(i,q)s(i+k-1,q)}{s(0,q)s(q,q)} s(i+m+1,q) + (\lambda_{k-1,m-k+q+2})s(0,q), \tag{13}$$

but $s(m+1, k-1) = s(-m-1, k-m-2)$, where $k-m-2 \geq k$. Then according to Assumption 1 we have,

$$s(m+1, k-1) = s(-m-1, k-m-2) = \sum_{j=0}^{q-k+m+2} \frac{s(j,q)s(j+k-m-2,q)}{s(0,q)s(q,q)} s(j-m-1,q). \tag{14}$$

Consequently, the previous equation can now be rewritten as follows,

$$s(m+1, k-1) = \sum_{j=1}^{q-k+m+2} \frac{s(j,q)s(j+k-m-2,q)}{s(0,q)s(q,q)} s(j-m-1,q) + \frac{s(0,q)s(k-m-2,q)}{s(0,q)s(q,q)} s(-m-1,q). \tag{15}$$

If we make the transformation $j = i+m+1$ in equation (15) we obtain ,

$$s(m+1, k-1) = \sum_{i=-m}^{q-k+1} \frac{s(i+m+1,q)s(i+k-1,q)}{s(0,q)s(q,q)} s(i,q) + \frac{s(0,q)s(k-m-2,q)}{s(0,q)s(q,q)} s(-m-1,q). \tag{16}$$

Now observe that the summations in equations (16) and (13) are equal , thus we can deduce that

$$\lambda_{k-1,m-k+q+2} = \frac{s(k-m-2,q)s(-m-1,q)}{s(0,q)s(q,q)},$$

and consequently equation (13) can be rewritten as

$$s(m+1, k-1) = \sum_{i=-m-1}^{q-k+1} \frac{s(i,q)s(i+k-1,q)}{s(0,q)s(q,q)} s(i+m+1,q). \tag{17}$$

Equation (17) demonstrates that equation (3) is valid for $j = m+1$ and $n = k-1$. Now, knowing that the initial equation corresponding to $n = k-1$ and $j = k-q-1$ holds, we have demonstrated that we can prove equation (3) to be valid for $n = k-1$ and $k-q-1 \leq j \leq -1$. From expression (7) we know that,

$$\mathbf{s}_{k-1} = \sum_{i=0}^{q-k+1} \lambda_{k-1,i}\mathbf{s}_q^i. \tag{18}$$

We have already obtained the values of $\lambda_{k-1,0}$ to $\lambda_{k-1,q-k}$, but we still need to find $\lambda_{k-1,q-k+1}$. This is easily obtained if we consider the following expression for the last non-identically zero element of $\mathbf{s}_{k-1}$:

$$s(q, k-1) = \lambda_{k-1,q-k+1}\, s(q,q), \tag{19}$$

$$\lambda_{k-1,q-k+1} = \frac{s(q,k-1)}{s(q,q)}. \tag{20}$$

Since we know all the λ's in (18), we can now write (18) in scalar form:

$$s(j, k-1) = \sum_{i=0}^{q-k+1} \frac{s(i,q)s(i+k-1,q)}{s(0,q)s(q,q)} s(i+j,q) \tag{21}$$

where j=k-1-q,...,q. Given that equation (3) is valid for $n = q-1$ we have shown that it is valid for every n such that $q-1 \geq n \geq 0$. ■

Efficient ARMA Parameter Estimation of Non Gaussian Processes by Minimization of the Fisher Information under Cumulant Constraints

Jean-Luc Vuattoux and Eric Le Carpentier
Laboratoire d'Automatique de Nantes, URA C.N.R.S. 823
Ecole Centrale de Nantes/Universite de Nantes,
1 rue de la Noë, 44072 Nantes cedex 03, France
vuattoux@lan.ec-nantes.fr

Abstract

The problem of estimating the parameters of a non causal ARMA system, driven by an unknown input noise with unknown probability density function (PDF) is addressed. A maximum likelihood approach is proposed in this paper. The main idea of our approach is that the assumed PDF of the input noise is the PDF minimizing the Fisher information (FI) among PDFs matching the estimated cumulants up to 4th order. This minimization problem is hard to solve, so we use an over-parameterized PDF model, which is a gaussian mixture, and minimize the FI in this set. A new parameter estimation method is given and its robustness properties are detailed. The performances of the resulting identification scheme are compared to those of another higher order method.

1. Introduction

The identification of the parameters of a discrete linear shift-invariant system by observation of its output is of considerable interest in time series and spectral analysis, filtering and prediction. In non gaussian case, numerous methods based on higher order statistics (HOS) have been introduced due to the fact that the output of these systems carries phase information. Their main disadvantage is that they do not provide any information about the theoretical performances of the estimator and its optimality in the sense of the covariance matrix of the estimated parameters.

To obtain an optimal estimator, it is necessary to know the exact probability density function (PDF) of the input noise in order to calculate the maximum likelihood (ML) estimator. If the PDF is unknown, we can assume a certain class of PDFs for the input and obtain the optimality in the minimax sense by using a ML approach with the PDF which minimizes the Fisher information (FI) in this class and provides the most robust (in Huber's sense) [3, 6] parameter estimates.

In connection with higher order statistics, we consider the class of cumulants constrained PDFs and determine the PDF which minimizes the FI under cumulant constraints. This optimization problem was partially solved in [9] but the results are limited to the symmetrical sub-gaussian PDFs. So this problem is always open for super-gaussian and non symmetrical PDFs.

In this paper, we propose a new parameter estimation method based on the prediction error method (PEM) using cumulants of second, third and fourth order and the minimization of the FI. We use a model of PDF, appropriate for non gaussian processes with heavy tails, which is a gaussian mixture (GM) PDF:

$$f_M(u) = p\Phi_1(u) + (1-p)\Phi_2(u), \quad 0 \leq p \leq 1 \qquad (1)$$

where $\Phi_1(u)$ and $\Phi_2(u)$ are Gaussian PDFs. We consider the case of a non symmetrical PDF (see the symmetrical case in [4]) constrained by the second, third and fourth order cumulants, used in practice.

In section 2, we present the procedure for the determination of the set of centered GM PDFs having the same variance, skewness and kurtosis (second, third and fourth order cumulants), and the element of this set minimizing the FI. This solution is given analytically for the null skewness case. In section 3, the proposed parameter estimation scheme is explained and the robustness properties of our estimator are given. The estimation algorithm and simulation results are presented in section 4. In section 5, a conclusion is given.

2. Model for the input PDF

Here we introduce the cumulant matched GM as the model for the input distribution. Despite the fact that it does not result from any constrained mini- or maximization of

0-8186-7576-4/96 $5.00

PDF measure, it has very useful characteristics and interesting properties (see [4]).

So, let C_2, C_3 and C_4, respectively the variance, the skewness and the kurtosis of any PDF ($C_2 > 0$, $C_4 \geq C_3^2/C_2 - 2C_2^2$). We will show that there exists always a non empty set F_M of centered GMs (1) having these cumulants. The problem is which mixture to choose in this set, when it contains more than one element, as model for the input PDF. So we decide to take the mixture model of F_M minimizing the FI, i.e.

$$f_M^0 = \arg \min_{f \in F_M} I_f \tag{2}$$

where I_f is the FI defined as

$$I_f = \int_{-\infty}^{\infty} \frac{(f')^2}{f}\, du \tag{3}$$

With the GM, this integral can be evaluated only numerically.

Consider m_1, m_2 and V_1, V_2, respectively the means and variances of $\Phi_1(u)$ and $\Phi_2(u)$ in the GM given by (1). To obtain a centered GM with given variance C_2, skewness C_3 and kurtosis C_4, we must have:

$$\begin{cases} pm_1 + (1-p)m_2 = 0 & (a) \\ p(V_1 + m_1^2) + (1-p)(V_2 + m_2^2) = C_2 & (b) \\ pm_1(3V_1 + m_1^2) + (1-p)m_2(3V_2 + m_2^2) = C_3 & (c) \\ p(3V_1^2 + 6m_1^2V_1 + m_1^4) + \\ \quad (1-p)(3V_2^2 + 6m_2^2V_2 + m_2^4) - 3C_2^2 = C_4 & (d) \end{cases} \tag{4}$$

We can see that if $p = 0$ or $p = 1$, the PDF $f_M(u)$ (1) is Gaussian with variance C_2 and it is possible only if $C_3 = C_4 = 0$. So, if $C_3 \neq 0$ or $C_4 \neq 0$ then $0 < p < 1$.

Now, the problem is to determine the set of solutions of the system (4). To do this, we use a new parameterization:

$$\begin{cases} m_1 = k(1-p) \\ m_2 = -kp \end{cases} \tag{5}$$

where k is real. So, the relation (4a) is always verified and m_1 and m_2 can take all values.

For convenience, we consider now that $k \neq 0$. $k = 0$ case will be precised later. Then in using (4b) and (4c), we get the formula of the variances V_1 and V_2 in function of p and k:

$$\begin{cases} V_1 = C_2 + \frac{C_3}{3kp} - \frac{k^2(1-p^2)}{3} & (a) \\ V_2 = C_2 - \frac{C_3}{3k(1-p)} - \frac{k^2p(2-p)}{3} & (b) \end{cases} \tag{6}$$

Next, by replacing (6) in (4b), (4c), (4d) and by combination of these, we obtain the equation linking p and k:

$$\begin{aligned} p^2(1-p)^2(p^2 - p + 1)k^6 + p(1-p)(4p-2)\times \\ C_3k^3 + \tfrac{3}{2}p(1-p)C_4k^2 - \tfrac{C_3^2}{2} = 0 \end{aligned} \tag{7}$$

This equation is analytically solvable only if $C_3 = 0$ and $k \neq 0$, so if $C_4 \leq 0$. In the general case, we solve numerically (7) in order to determine the pairs (p, k), with $0 < p < 1$ and $k \neq 0$, solutions of (7). We can remark immediately that, if the pair (p, k) is solution of (7), then the pair $(1-p, -k)$ too. And the two resulting mixtures are identical.

If $k = 0$, then we cover the class of symmetrical super-gaussian PDFs studied in [4]. So, it ensues the following proposition:

Proposition 1: Let C_2, C_3 and C_4, respectively the variance, the skewness and the kurtosis of any PDF. Then there exists a non empty set F_M of centered GMs having these cumulants. We have to distinguish three cases:

(i) $C_3 \neq 0$: then F_M is characterized by the pairs (p, k), where $0 < p < 1$ and $k > 0$, solutions of the equation (7) and for which the variances V_1 and V_2 of the two Gaussian PDFs of the mixture are non negative, m_1 and m_2 being given by (5). F_M is noted $GM_{p,k}$.

(ii) $C_3 = 0$ and $C_4 \leq 0$: then $k \neq 0$. F_M is characterized by the pairs (p_1, k) and $(1 - p_1, k)$, where

$$p_1 = \frac{1}{2} - \sqrt{-1 + 2\sqrt{1 + \frac{6C_4}{k^4}}} \tag{8}$$

and $(-8C_4)^{\frac{1}{4}} \leq k \leq \sqrt{\frac{3C_2}{1-p_1^2}}$ to have $0 < p_1 < 1$ and V_1 and V_2 non negative. F_M is noted $SBGM_k$.

(iii) $C_3 = 0$ and $C_4 \geq 0$ (super-gaussian): then $k = 0$. F_M is characterized by

$$\begin{cases} m_1 = m_2 = 0 \\ V_1 = C_2 - \alpha\sqrt{\frac{C_4}{3}} \\ V_2 = C_2 + \frac{1}{\alpha}\sqrt{\frac{C_4}{3}} \\ p = \frac{1}{1+\alpha^2} \end{cases} \quad \forall \alpha \in]0, \sqrt{\tfrac{3C_2^2}{C_4}}] \tag{9}$$

F_M is noted SGM_α.

Remark 1: For the cases (i) and (ii), we show easily that in the boundary case $C_4 = C_3^2/C_2 - 2C_2^2$, the sets $GM_{p,k}$ and $SBGM_k$ contain an unique PDF which corresponds to the Bernoulli distribution.

Now the determination of the PDF f_M^0 (2) of F_M which minimizes the FI leads to the following results:

Proposition 2: The GM of $SBGM_k$ minimizing the Fisher information (3) is f_M^0, where $F_M = SBGM_k$ in (2), with

$$\begin{cases} m_1 = -m_2 = (\frac{-C_4}{2})^{\frac{1}{4}} \\ V_1 = V_2 = C_2 - \sqrt{\frac{-C_4}{2}} \\ p = \frac{1}{2} \end{cases} \tag{10}$$

It is quite natural that the solution be the symmetrical PDF (sub-gaussian PDFs class) of the set SGM_k because it is the more Gaussian of this set (all its cumulants of odd order are null) in the sense that the minimum value of the FI, for any PDF, is reached for the Gaussian PDF [7].

Proposition 3: The mixture PDF model of SGM_α minimizing I_{f_M} is obtained for $\alpha \longrightarrow 0$ and then $C_2 I_{f_M} \longrightarrow 1$.

Due to the Proposition 3, it seems that the model of Proposition 1 (*iii*), when α tends to 0 but is not 0, is an ϵ-approximation of the solution of the FI minimization under constraints of C_2 and C_4 for the class of super-gaussian PDFs since the absolute minimum of $C_2 I_f$ is 1, obtained for the Gaussian PDF. In practice, α is taken small enough (see [3]).

3. ARMA parameters estimation

Let the observed process $\{y_t\}$ be modeled as the output of a discrete stable linear shift-invariant system $H_{\theta_0}(z)$ with input $\{e_t\}$:

$$y_t = H_{\theta_0}(z) e_t \tag{11}$$

where

$$H_{\theta_0}(z) = \frac{A(z)C(z^{-1})}{B(z)D(z^{-1})} = \frac{(\sum_{i=0}^{n_A} a_i z^{-i})(\sum_{i=0}^{n_C} c_i z)}{(\sum_{i=0}^{n_B} b_i z^{-i})(\sum_{i=0}^{n_D} d_i z)} \tag{12}$$

with $a_0 = b_0 = c_0 = d_0 = 1$ and

$$\theta_0 = [a_1 \dots a_{n_A}\, b_1 \dots b_{n_B}\, c_1 \dots c_{n_C}\, d_1 \dots d_{n_D}]^T \tag{13}$$

We assume that all the roots of $A(z)$ and $B(z)$ are inside the unit-circle (causal minimum phase part) and all the roots of $C(z)$ and $D(z)$ are outside the unit-circle (anti-causal maximum phase part). The input $\{e_t\}$ is an independent identically distributed (i.i.d.) sequence with unknown PDF $f_e(u)$.

Given N consecutive samples of the system output y_t, $t = 1, \dots, N$, we want to estimate the actual parameter θ_0. The prediction error sequence $\{w_t(\theta)\}$ [5, 7] is related to the data through

$$w_t(\theta) = H_\theta^{-1}(z) y_t \tag{14}$$

With PEMs, the estimate $\widehat{\theta}$ of θ_0 is equal to θ which minimizes some criterion depending on the sequence of prediction errors

$$J(\theta) = \frac{1}{N} \sum_{t=1}^{N} \ell(w_t(\theta)) \tag{15}$$

where $\ell(.)$ is a scalar-valued norm, i.e.

$$\widehat{\theta} = \arg\min_{\theta} J(\theta) \tag{16}$$

Like $f_e(u)$ is unknown, there are two possibilities: either choosing a norm giving satisfying results for a broad class of input PDFs (robustification), or estimating $f_e(u)$ from the available data. Our approach is nearest of the robust identification in the sense where we take the PDF f_M^0 (2) of the class F_M, which is $GM_{p,k}$ or $SBGM_k$ or SGM_α, depending on the values of the second, third and fourth order cumulants of the PDF $f_e(u)$ for the choice of this class. The criterion to minimize is $J(\theta)$ with the norm $\ell(w) = \ell^0(w) = -\log[f_M^0(w)]$ (ML approach).

In Huber's sense [3, 6], if the true PDF f_e belongs to the class F_M, we obtain the most robust estimator (16) of the estimator class generated by F_M with the norm $\ell^0(.)$. It is possible to show that the proposed estimator $\widehat{\theta}$ (16) is asymptotically optimal in the minimax sense for the particular class of PDFs F_M. Under some assumptions (see [7]), the estimate (16) is consistent and the following expressions hold

$$\begin{cases} \sqrt{N}(\widehat{\theta} - \theta_0) \sim \mathcal{N}(0, V(\ell^0, f_e)) & (a) \\ V(\ell^0, f_e)) \leq V(\ell^0, f_M^0) = V^* & (b) \end{cases} \tag{17}$$

On the other hand, we find that

$$\mathrm{cov}(\widehat{\theta} - \theta_0) \geq N^{-1} V^* \tag{18}$$

Thus, for $f_e = f_M^0$, the asymptotic covariance $V(\ell^0, f_e)$ of the proposed estimate (16) reaches the lower possible boundary V^*, which depends on the FI of f_M^0 and on θ_0. Its calculus is detailed in [7]. For other $f_e \in F_M$, the asymptotic covariance does not exceed V^*. If $f_e \notin F_M$, only the relation (17a) holds. In all the cases, $V(\ell^0, f_e)$ is obtained theoretically with the results of [7] and [5].

4. Algorithm and simulation results

Each step of the algorithm consists of the three parts:

1) Estimate the cumulants of the prediction error process w_t (14).

2) Calculate the model f_M^0 for the input PDF by (2), based on the estimated cumulants of w_t. Following the values of these cumulants, we choose between the numerical procedure given in Proposition 1 and the models (10) or (9) for the calculus of f_M^0.

3) Find the minimum of the criterion (15) (with $\ell(w) = \ell^0(w)$) in the search direction of a quasi-Newton algorithm, calculated with the input model f_M^0. This calculus is not detailed here but it is similar to the one presented in [7].

In the initialization phase of our ML approach, any $4th$-order methods can be used, for example, the W-slice algorithm [8], to avoid convergence to false local minima.

To demonstrate the asymptotic efficiency of our ML approach, we made many simulations with a non-causal ARMA model driven by different symmetrical (belonging to sub- or super-gaussian class of PDFs) or non symmetrical input noises. We took a model in [2] and inversed its causal real pole to obtain our non-causal model. It has poles at 5 and $0.6179 \pm j0.5077$ and zeros at -0.7 and 2, so with transfer function given by:

$$H(z) = \frac{(1 + 0.7z^{-1}).(1 - 0.5z)}{(1 - 1.2358z^{-1} + 0.6396z^{-2}).(1 - 0.2z)}$$

The algorithm was tested by simulation on the presented model considered as unknown and driven by three differ-

Table 1. ARMA parameter estimates (N=2000, 100 Monte-Carlo runs).

Input	True parameter	ML		LS+max$\|K\|$	
		Mean	Std	Mean	Std
I	a_1= 0.7000	0.7018	0.0246	0.7003	0.0267
	b_1=-1.2358	-1.2243	0.0295	-1.2266	0.0412
	b_2= 0.6396	0.6329	0.0270	0.6358	0.0436
	c_1=-0.5000	-0.4756	0.1329	-0.4838	0.1485
	d_1=-0.2000	-0.1865	0.1525	-0.1969	0.1687
II	a_1= 0.7000	0.7046	0.0231	0.6996	0.0255
	b_1=-1.2358	-1.2299	0.0306	-1.2267	0.0408
	b_2= 0.6396	0.6337	0.0276	0.6356	0.0448
	c_1=-0.5000	-0.4846	0.1308	-0.4932	0.1361
	d_1=-0.2000	-0.1862	0.1443	-0.2061	0.1585
III	a_1= 0.7000	0.7002	0.0136	0.7038	0.0266
	b_1=-1.2358	-1.2336	0.0162	-1.2262	0.0440
	b_2= 0.6396	0.6369	0.0175	0.6360	0.0430
	c_1=-0.5000	-0.4998	0.0560	-0.4839	0.1514
	d_1=-0.2000	-0.2030	0.0624	-0.1934	0.1719

ent i.i.d. input noises: laplacian (I)(super-gaussian), uniform (II)(sub-gaussian) or exponential (III). 100 independant Monte-Carlo runs were performed for each simulation. The signal's length used is $N = 2000$ samples. We compared this results with a method [1] (noted LS+max$|K|$), where the spectrally equivalent minimum phase system is primarily identified using least squares method (LS). Then, among all the spectrally equivalent systems, we choose the model which maximizes the absolute value of the estimated normalized kurtosis of the innovation process.

In Table 1, the mean and the standard deviation (Std) of the parameter estimates are summarized for the different input noises. The presented results show the good behaviour of our method compare to the LS+max$|K|$ method with smaller bias and Std.

5. Conclusion

A possible way to obtain an efficient parameter estimates in case of unknown non gaussian input is presented. The innovation of the proposed PDF model is that it is the element, minimizing the FI, of a set of GMs having the same four first cumulants than the true input PDF. This PDF model is parameterized by its second and fourth order cumulants for the classes of null skewness PDFs, and determined numerically for the non zero skewness PDFs class. An interesting result has been obtained in the super-gaussian case for which the PDF model (9) seems to be an ϵ-approximation of the solution of the more general problem of FI minimization under constraints of C_2 and C_4.

Simulation results seem to confirm the good behaviour and robustness of our method compared to other methods based on higher-order statistics. Future works will address means to find the analytical form of the PDF f_M^0 in the case of non symmetrical PDF.

References

[1] M. Boumahdi and J.-L. Lacoume. Blind identification of non-minimum phase filters by maximising the kurtosis. *Proc. of EUSIPCO 94, Edinburgh, Scotland, UK*, pages 195–198, September 1994.

[2] G. Giannakis and J. Mendel. Identification of non-minimum phase systems using higher-order statistics. *IEEE Trans. on A.S.S.P.*, 37(3):360–377, March 1989.

[3] P. Huber. Robust estimation of a location parameter. *Ann. Math. Stat.*, 35:73–104, 1964.

[4] E. Le Carpentier and J.-L. Vuattoux. ARMA model identification using higher order statistics and Fisher information concepts. *Proc. of EUSIPCO 96, Trieste, Italy, to appear*, September 1996.

[5] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[6] B. Poljak and J. Tsypkin. Robust identification. *Automatica*, 16:53–63, 1980.

[7] O. Shalvi and E. Weinstein. Maximum likelihood and lower bounds in system identification with non-gaussian inputs. *IEEE Trans. Inform. Theory*, 40(2):328–339, March 1994.

[8] J.-L. Vuattoux, D. Zazula, and E. Le Carpentier. Causal ARMA parameter estimation using a linear combination of third-order cumulant slices. *Proc. of ATHOS, Begur, Girona, Spain*, pages 464–468, June 1995.

[9] V. Živojnović. Higher-order statistics and Huber's robustness. *Proc. of ATHOS, Begur, Girona, Spain*, pages 236–240, June 1995.

Minimum Entropy Filtering for Improving Nonstationary Sonar Signal Classification

Michael K. Broadhead, Lisa A. Pflug, and Robert L. Field
Naval Research Laboratory, Stennis Space Center, MS 39529 USA

Abstract

The Minimum Entropy Method is studied with regard to its performance in removing multipath distortion from passive transients, to improve the performance of classifiers. It was found that the method often works well if the kurtosis of the associated multipath Green's function is high enough, and that signal stationarity is not required. We also found that, while there are usually a few filter lengths at which the best solutions are obtained with conventional convergence criteria, good solutions exist across a much broader range of filter lengths if the iterations are not allowed to proceed to convergence. That is, kurtosis needs to be increased, but not maximized. In many cases, two or three iterations is sufficient.

1. Introduction

The passive sonar classification problem can be decomposed into two stages: 1) recovering the source time signature of a transient event from a set of received signals by accounting for environmental distortion effects, and 2) applying a pattern recognition algorithm to the estimated source signature for final classification. By environmental distortion, we refer to effects present in the received data at the sensor array that are not present in the source signature. In our case, environmental effects consist primarily of multipath and low level ambient noise. For a spatial point source, if we incorporate the environmental effects into a Green's function, and assume time-invariance, the received pressure time series at a desired location can be modeled as the convolution of the transient source signature with the Green's function. A term representing additive noise effects can be added to the convolution. The Green's function, of course, depends on the environmental acoustic parameters and the source and receiver location.

When a Green's function has been determined by numerical solution of the wave equation, it can be used to deconvolve the measured time series for an estimate of the source signature, which is referred to as the deterministic approach. Broadhead *et al.* [1] reviewed this approach, and performed an additional study in a bottom-limited propagation environment, showing that there was extreme sensitivity to inaccuracy in the bottom geoacoustic parameters.

Broadhead [2] used a statistical source estimation approach to address the problem of recovering a source signature without specific knowledge of its location, or the environmental parameters necessary to accurately compute the Green's functions. He gave examples, for the single channel case, where this can be done if the Green's functions representing environmental distortion are lepto-kurtic (a specific type of non-Gaussianity). The method used, called the minimum entropy deconvolution method (MED), was introduced by Wiggins in 1977 [3]. This method was further refined and interpreted by various researchers (see bibliography in Ref. [4]). The goal of this method is to produce a filter that drives the output of the system to lower entropy (greater order), or equivalently, drive the governing distribution more towards non-Gaussianity (higher kurtosis). The success of MED depends on the non-Gaussianity of the input random process, but apparently does not require stationarity (examples are given in [2]).

In this paper we continue the work begun in [2] with a more thorough and systematic exploration of the solution space provided by the MED parametric method. The results in this paper show that exploitation of higher order parametric methods to achieve classification performance gains for nonstationary sonar signals appears promising.

2. MED Algorithm

The MED algorithm has been thoroughly described in the literature, and will not be repeated here, but a minimum of terminology must be defined. We seek the MED filter f of length N that is a stationary point of the functional

$$V = \sum_j \hat{g}_j^4 \Big/ \left(\sum_j \hat{g}_j^2 \right)^2 , \qquad (1)$$

where

$$\hat{g}_j = \sum_l f_l x_{j-l} , \qquad (2)$$

0-8186-7576-4/96 $5.00

$\hat{g}$ is the Green's function estimate, x is the input signal and V is the Varimax norm (essentially kurtosis). The resulting nonlinear system of equations is solved iteratively. A starting point is given by taking f as a delta function. After iterating to some stopping criterion (to be discussed), we are left with the filter f and the Green's function estimate $\hat{g}$. To obtain an estimate of the source signature $\hat{s}$, we calculate the inverse of f.

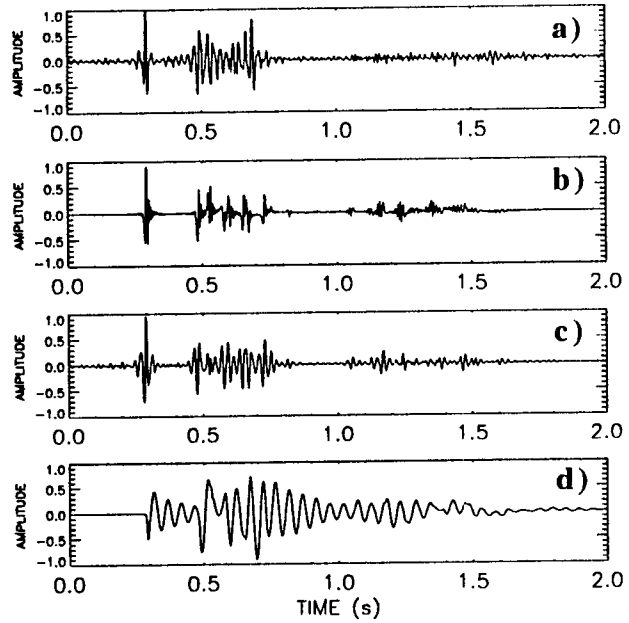


Fig. 1. Signal type examples for the 4300 m range. (a) Data, (b) PE Green's function, (c) Short pulse simulation, (d) Long pulse simulation.

3. Signal Description

We have three types of input signals: 1) data, 2) short pulse simulations (SPSIMUL), and 3) long pulse simulations (LPSIMUL). The data analyzed was obtained in an experiment conducted in the Atlantic Ocean, in the vicinity of Blake Plateau. For details, refer to Refs. [1], [2], and [5]. A typical time series is shown in Fig. 1 (a) (250 m receiver depth, 4.3 km source-to-receiver range). The bottom interacting events occur after about 0.4 seconds.

In Fig. 1(b) we show the corresponding calculated (PE model) Green's function. We used this and the two pulses shown in Fig. 2 to create (by convolution) two types of simulations: 1) a short simulation representative of the data, as shown in Fig. 1(c), and 2) a long pulse simulation, shown in Fig. 1(d), that creates more overlap between the various arrivals. We have displayed only the 4.3 km range, but have also processed the 600 meter and 7.9 km ranges. Refer again to the above references for more examples of time series.

In Fig. 2, as mentioned, we display the two pulse types. In 2(a), the short pulse is our best estimate from measurements from a source array mounted hydrophone of the true source pulse on the data. The longer pulse in 2(b) is an exponentially damped sinusoid.

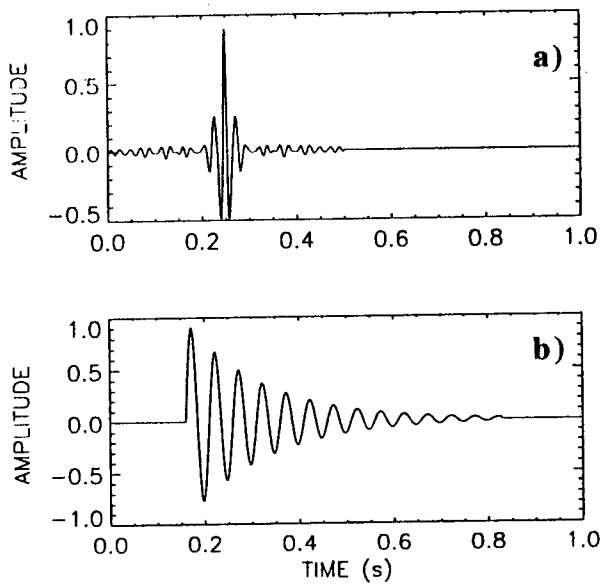


Fig. 2. Source types.

4. Processing Methodology

We developed two basic processing methodologies, which will be referred to as CONVRG and BEST. CONVRG uses a conventional convergence criterion, and the output is a correlation coefficient between the source estimate and the known source at each filter length from 1 to 50. (The correlation coefficient, given by the symbol γ is standard except that we always report it as the absolute value). The convergence criterion was as follows: the correlation coefficient is calculated between the current MED filter iterate and the previous. When this value exceeds the specified tolerance, the iteration is stopped. We used a tolerance level of 0.9999.

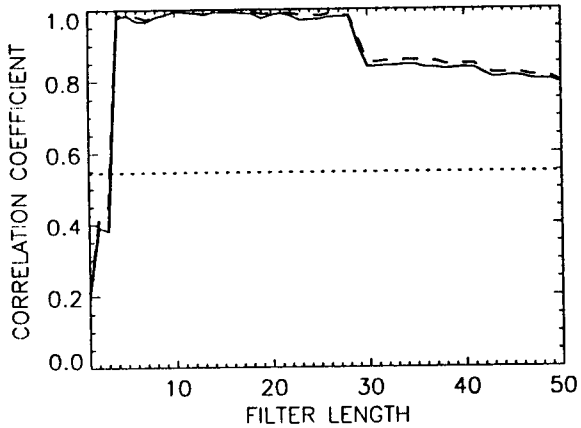


Fig. 3. Results for LPSIMUL, 600 m case.

We mimic the case of doing no preprocessing before classification, that is, just correlating the received signal with the source signature. This value gives a measure of how much distortion was introduced by the multipath, and will be indicated by short dashes in the figures. The output of CONVRG will be indicated by a solid curve.

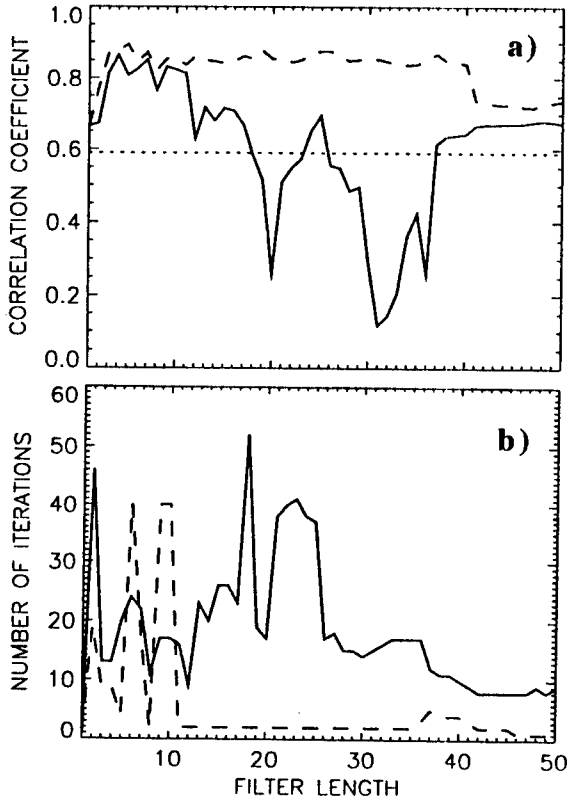


Fig. 4. (a) CONVRG/BEST (Solid/Dash) results for SPSIMUL, 7900 meter case. (b) Number of iterations for results in (a).

BEST outputs the correlation coefficient between the best possible source estimate and the known signature at each N out of a possible *itermax* iterations, without regard to actually trying to maximize V. *itermax* was variously either 30 or 40 iterations. This curve will be represented by long dashes in the figures. In both cases, the number of iterations actually performed at each filter length, and the estimated Green's function kurtosis were also output.

There were two stages in both algorithms where some regularization could be required: 1) on a given iteration, the Toeplitz coefficient matrix, which may become ill-conditioned, and 2) the calculation of the inverse of the MED filter, which may have spectral zeros (a frequency domain method was used). The data and SPSIMUL cases used a pre-whitening value of 0.01% for stage 1). The 4300 m range of LPSIMUL also used this value. No pre-whitening was used for the other two ranges. In no case was pre-whitening used for the inverse filter. A definite sensitivity to the amount of pre-whitening was noticed.

5. Results

In Table I we summarize the results in the form of correlation coefficients between the estimated and known pulses. For the different signal type and processing methodology combinations, only the highest coefficient

Range (m)	600	4300	7900
DATA/ CONVRG	0.896	0.802	0.822
DATA/ BEST	0.902	0.865	0.858
SPSIMUL/ CONVRG	0.975	0.946	0.864
SPSIMUL/ BEST	0.975	0.958	0.893
LPSIMUL/ CONVRG	0.995	0.963	0.992
LPSIMUL/ BEST	0.995	0.965	0.993

TABLE I Summary of highest correlation coefficients for all cases.

obtained is reported in each case. In most cases the highest coefficient obtained for BEST and CONVRG were comparable. They were significantly different at other filter lengths, for some cases, however, which we will discuss later. LPSIMUL results were typically better than SPSIMUL. We will speculate as to why, also later.

In all cases, the best results were a significant improvement over doing no preprocessing of this kind, where the γs then are less than 0.7. As would be expected, the results were better for the simulations than for the real measurements. Not only did the data have some noise, but the "true" source is no known pulse with complete accuracy.

In Fig. 3, we display results for LPSIMUL, 600 m. This figure demonstrates that conventional convergence criteria can work very well. This also happens to be a case where the associated Green's function kurtosis was very high (124.6). The resulting simulated data kurtosis (after convolution with the long pulse) was 26.2.

We consider a short pulse simulation example in Fig. 4(a) for the 7900 m range. This figure shows that for some cases, beyond a certain value of N, the algorithm performance drops off significantly (often worse than doing nothing) for conventional convergence criteria. However, the long dashed line (BEST) shows that good solutions are available, only at many less iterations (often, only 2) than needed to maximize V. This is shown in Fig. 4(b), where the solid curve represents the number of iterations required to satisfy the conventional convergence criterion (CONVRG), and the long-dash curve represents the number of iterations associated with the best possible source estimate in 40 or less iterations.

In contrast to the previous case, the initial Green's function kurtosis was much lower (27.3). Also, the change in kurtosis after convolution with the short pulse was less (kurtosis of simulated data = 14.8). Note that the starting kurtosis for this signal is comparable to the final kurtosis of the convolved signal in the example in Fig. 3. These factors are probably significant for determining when conventional convergence criteria will or will not work well.

In Fig. 5, we display the kurtosis of the MED Green's function estimates produced by both CONVRG and BEST, as a function of filter length. This figure is fairly typical of the results. It shows a steady increase in the final (maximized) kurtosis value for CONVRG, but a level average value for BEST, indicating that the most accurate estimate of the Green's function is not necessarily the estimate with the highest kurtosis.

In Fig 6, we show pulse estimates for the highest and lowest correlation coefficients for SPSIMUL/CONVRG. This gives a rough idea of the visual quality of match expected for the range of correlation coefficients relevant to this study. In Fig. 6(a), for the 600 m case, $\gamma = 0.975$. In Fig. 6(b), $\gamma = 0.864$, and the range is 7900 m.

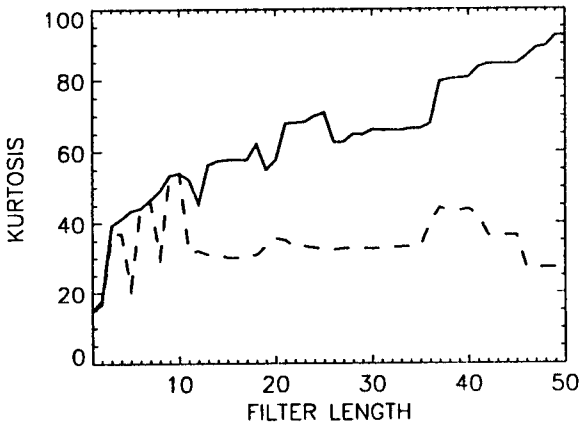


Fig. 5. Green's function estimate kurtosis for SPSIMUL, 7900 meter case.

6. Discussion and Conclusions

We have found that for the cases studied the processor is always capable of providing better results than doing no preprocessing at all. In most cases, the best results were very good ($\gamma > 0.9$). Even when the processor gave good results only over a narrow range of filter lengths, many good solutions were still available at other filter lengths, only at a smaller number of iterations than maximizing the V norm would require. Since our goal is to produce a class of candidate solution signals for classification, this is useful in that it may be exploitable in increasing the probability of having a "good" solution in the class of signal candidates, albeit at the expense of adding a dimension to the search space.

The degree of kurtosis possessed by the Green's function representing the multipath distortion appears to be an important factor in the quality of the results. The final kurtosis value of the convolved signal may also be important, which remains to be determined.

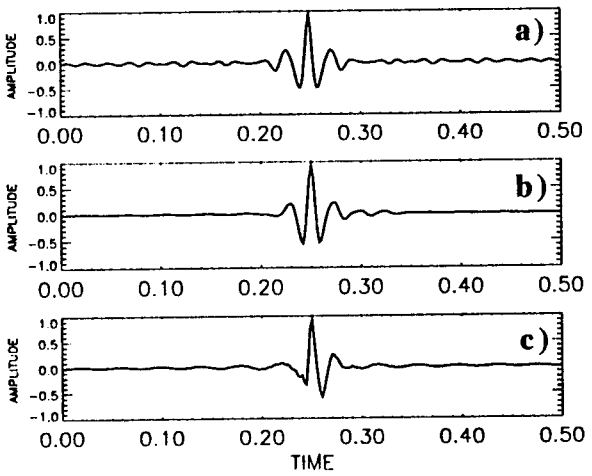


Fig. 6. Best and worst source estimates for SPSIMUL/CONVRG a) 600 meter range, γ = 0.975, b) 7900 meter range, γ = 0.864.

Acknowledgments

NRL contribution number NRL/PP/7173--95-0125.

References

[1] M. K. Broadhead, R. L. Field, and J. H. Leclere, "Sensitivity of the deconvolution of acoustic transients to Green's function mismatch, " *J.* Acoust. Soc. Am., Vol. 94, pp. 994-1002, 1993.

[2] M. K. Broadhead, "Broadband Source Signature Extraction from Underwater Acoustics Data with Sparse Environmental Information," *J.* Acoust. Soc. Am., Vol. 97, No. 2, pp. 1322-1325, 1995.

[3] R. A. Wiggins, "Minimum entropy deconvolution," Proc. Int. Symp. Computer Aided Seismic Analysis and Discrimination, June 9, 10, Falmouth, Mass., IEEE Computer Society, pp. 7-14, 1977, and Geoexploration, Vol. 16, pp. 21-35, 1978.

[4] A. T. Walden, "Non-Gaussian reflectivity, entropy, and deconvolution," Geophysics, Vol. 50, pp. 2862-2888, 1985.

[5] R. L. Field and J. H. Leclere, "Measurements of bottom-limited ocean impulse responses and comparisons with the time domain parabolic equation," J. Acoust. Soc. Am., Vol. 93, pp. 2599-2616, 1993.

On Coherent Modelling of Non-Gaussian Radar Clutter

D. Robert Iskander and Abdelhak M. Zoubir
Signal Processing Research Centre, QUT
GPO box 2434, Brisbane, Q. 4001, Australia
d.iskander@qut.edu.au

Abstract

In this paper we introduce a general distribution, called the Generalised Bessel K (GBK) distribution, that involves the modified Bessel function of the second kind. The statistical properties of the proposed distribution as well as its application to coherent modelling of radar clutter are investigated. It is shown that the GBK-distribution includes a large number of the well known clutter models and is still mathematically tractable.

1. Introduction

Detecting targets embedded in clutter is one of the important tasks for a radar signal processing practitioner. In parametric detection it is essential that the clutter be understood and properly modelled.

In many applications clutter cannot be assumed to be Gaussian. This has motivated work on modelling clutter by non-Gaussian probability distributions. In practice, two main problems are apparent. Firstly, it is not unusual to encounter the data which is incompatible with a given distribution at hand. Secondly, for optimal (in the Neyman-Pearson sense) detection of signals in coherent and correlated clutter, multivariate probability density functions are required [2, 7].

To overcome these problems one may consider modelling clutter by a number of different probability models and use classification techniques for determining which probability model fits the given data most closely [10]. In practice, however, this technique is difficult and involves a considerable amount of computation. Thus, more general models for clutter that lead to optimal detection are sought.

In this contribution we introduce a new clutter model that involves modified Bessel functions of the second kind. We call it the Generalised Bessel K function (GBK) distribution unlike the other generalisations of the K-distribution [5] which involve modified Bessel functions of the first and second kind. The GBK-distribution includes a large number of the well known clutter models. At the same time it is mathematically tractable. Application of the GBK-distribution reduces the complexity in an adaptive radar system such as the one proposed in [10], since only one clutter model needs to be employed.

2. Genesis of the GBK-distribution

A K-distributed random variable can be obtained from the multiplication of a Rayleigh and a Gamma variate [6]. Another K-distribution can be obtained by multiplying an Exponential and a Gamma variate [9]. Taking into account the fact that the Exponential distribution is included by the Gamma distribution, the K-distribution can be generalised by considering a distribution that originates from compounding two Gamma distributions. Teich and Diament call such a distribution the K'-distribution [9].

The K'-distribution can be generalised further by noting that the Gamma distribution is included by the generalised Gamma distribution [8]. Thus we consider a distribution which originates from compounding two generalised Gamma distributions. The two component distributions may be then given as

$$f_{X|Y}(x \mid y) = \frac{c\, x^{c\alpha_1 - 1}}{y^{c\alpha_1}\Gamma(\alpha_1)} \exp\left[-\left(\frac{x}{y}\right)^c\right] \qquad (1)$$

and

$$f_Y(y) = \frac{c\, y^{c\alpha_2 - 1}}{\beta^{c\alpha_2}\Gamma(\alpha_2)} \exp\left[-\left(\frac{y}{\beta}\right)^c\right], \qquad (2)$$

where all parameters are assumed to be positive, $f(\cdot \mid \cdot)$ denotes the conditional pdf, and $\Gamma(\cdot)$, is the standard Gamma function. The pdf of the GBK-distributed random variable X is derived using the integral formula [3, p. 313, Eq. 17] as

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{X|Y}(x \mid y) f_Y(y) dy \\ &= \frac{2c\left(\frac{x}{\beta}\right)^{\frac{c}{2}(\alpha_1+\alpha_2)-1}}{\beta\Gamma(\alpha_1)\Gamma(\alpha_2)} K_{(\alpha_2-\alpha_1)}\left[2\left(\frac{x}{\beta}\right)^{\frac{c}{2}}\right], \\ &= f_X(x; \alpha_1, \alpha_2, \beta, c) \qquad (3) \end{aligned}$$

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν.

0-8186-7576-4/96 $5.00

3. Properties of the GBK-distribution

The GBK-distribution is fully characterised by the four parameters, α_1, α_2, β, and c. Note that α_1 and α_2 are interchangeable due to the symmetry property of the modified Bessel function. The parameters α_1, α_2, and c control the shape of the probability density function while the parameter β controls its scale. In Figures 1–3 the probability density curves are shown where one parameter at a time is varied.

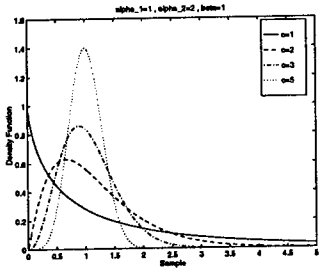


Figure 1. Pdfs for a GBK-distributed variate with $\alpha_1 = 1, \alpha_2 = 2$, and $\beta = 1$ for four different values of c.

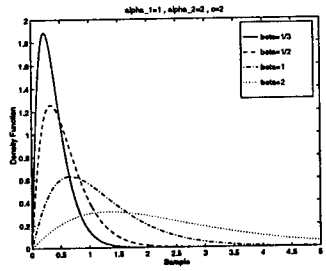


Figure 2. Pdfs for a GBK-distributed variate with $\alpha_1 = 1, \alpha_2 = 2$, and $c = 2$ for four different values of β.

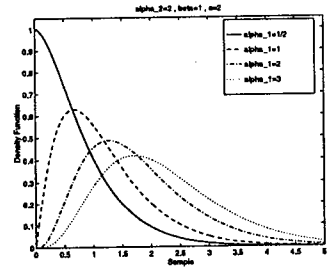


Figure 3. Pdfs for a GBK-distributed variate with $\alpha_2 = 2, \beta = 1$, and $c = 2$ for four different values of α_1.

The GBK-distribution includes a large number of distributions which occur frequently in data modelling. In Table 1 some of these distributions are reproduced with the appropriate parameter selection. The notation used for each of the distributions corresponds to the given reference in the third column of Table 1. In particular, the GBK-distribution includes the generalised Gamma distribution, both types of aforementioned K-distributions, the K'-distribution, and the Jakeman and Tough's generalisation of the K-distribution derived from a random walk in other than 2 dimensions [5, Eq. 2.11].

In radar applications it is desirable that the Rayleigh distribution (the first order amplitude distribution of a complex Gaussian process) be included in the general model [2]. For the GBK-distribution the Rayleigh distribution is included for two sets of parameters, $f_X(x; \frac{1}{2}, 1, 2\sigma, 4)$ and $f_X(x; 1, \nu \to \infty, 2a, 2)$. Note that the second set of parameters has a limit as in the case of the K-distribution [6]. On the other hand the first set of parameters is finite. Thus, the reduction of the GBK-distribution to the Rayleigh distribution is numerically more stable than the K-distribution.

Moments. The kth-order moment of the GBK-distribution is derived using [3, p. 313, Eq. 15] as

$$\begin{aligned} \mathsf{E}[X^k] &= \int_0^\infty x^k f_X(x)\, dx \\ &= \frac{\Gamma\left(\alpha_1 + \frac{k}{c}\right)\Gamma\left(\alpha_2 + \frac{k}{c}\right)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\beta^k, \end{aligned} \tag{4}$$

where $\mathsf{E}[\,\cdot\,]$ denotes the expectation operator.

Spherical Invariance. In order to design optimal detection schemes one needs to show that the first order envelope pdf given by (3) fulfils the requirements of spherically invariant random process (SIRP). The clutter process is spherically invariant if its Nth-order pdf can be given as

$$f_X(\boldsymbol{x}) = (2\pi)^{-N}|M|^{-1/2} h_{2N}(\boldsymbol{x}^T M^{-1}\boldsymbol{x}), \tag{5}$$

where $\boldsymbol{x} = [x_{c1}\, x_{c2} \cdots x_{cN}\, x_{s1}\, x_{s2} \cdots x_{sN}]^T$, where x_{ci} and x_{si}, $i = 1, \ldots, N$ are the in-phase and quadrature components of the radar clutter process, respectively, M is the covariance matrix of $\boldsymbol{x}$, and $h_{2N}(\,\cdot\,)$ is a suitable function. Following [2, 7], we derive

$$\begin{aligned} h_{2N}(q) &= \frac{4a^N c^{N-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}(\sqrt{aq})^{\frac{c}{2}(\alpha_1+\alpha_2+N-1)-2N} \\ &\quad \times \sum_{k=1}^{N} \frac{(-1)^{m+k} P_{(N,k)}}{\left(\frac{c}{2}\sqrt{aq}\right)^{N-k}} K_{(\alpha_2-\alpha_1-k+1)}\left[2(\sqrt{aq})^{\frac{c}{2}}\right], \end{aligned} \tag{6}$$

where $q = \boldsymbol{x}^T M^{-1}\boldsymbol{x}$,

$$a = \frac{\Gamma\left(\alpha_1 + \frac{2}{c}\right)\Gamma\left(\alpha_2 + \frac{2}{c}\right)}{2\Gamma(\alpha_1)\Gamma(\alpha_2)}, \quad m = \begin{cases} 1, & N\text{-odd} \\ 0, & N\text{-even} \end{cases},$$

and the coefficients $P_{(N,k)}$ are calculated recurrently

$$P_{(N,k)} = P_{(N-1,k)}\, C_{(N,k)} + P_{(N-1,k-1)},$$

with

$$C_{(N,k)} = \begin{cases} 0, & k > N \\ 1, & k = N \\ \frac{c}{2}\alpha_1 - (N-1) + \frac{c(k-1)}{2}, & k < N, \end{cases}$$

$P_{(0,0)} = 1$, $P_{(N,0)} = 0$, and $P_{(0,k)} = 0$. The result given in Eq. (6) satisfies the monotonicity condition for $\alpha_1 c \leq 2$. Thus, the GBK-distribution, can be used for coherent modelling of clutter when $\alpha_1 c \leq 2$. Such a result enables us to determine which of the nested distributions can represent a first order amplitude distribution of an SIRP. The conditions for this representation are given in the last column of Table 1.

4. Parameter Estimation

The practicality of the GBK-distribution requires the estimation of its parameters. The derivation of maximum likelihood estimates of the parameters of the GBK-distribution is cumbersome [4]. Therefore, a feasible alternative for estimating the parameters c, α_1, α_2, and β in (3) is the one based on higher order moments. Using (4) one can estimate the required parameters using any four sample moments of the data including fractional moments. In general, the estimation procedure has to be performed numerically in a four dimensional parameter space.

Estimation is greatly simplified if the parameter c is known. In such a case, the parameters α_1, α_2, and β can be obtained using the set of three different moment ratios

$$R_{k_i} = \frac{\mathsf{E}[X^{k_i c}]}{\mathsf{E}[X^{(k_i-1)c}]} = (\alpha_1+k_i-1)(\alpha_2+k_i-1)\beta^c, \quad i = 1,2,3. \tag{7}$$

Letting $k_1 = p$, $k_2 = q$, $k_3 = r$, and replacing the moments by their sample counterparts one obtains

$$\hat{\beta} = \left[\frac{p(\hat{R}_r - \hat{R}_q) + q(\hat{R}_p - \hat{R}_r) + r(\hat{R}_q - \hat{R}_p)}{(q-p)(p-r)(r-q)}\right]^{\frac{1}{c}}, \tag{8}$$

and

$$\hat{\alpha}_{1,2} = \frac{\hat{R}_q[(r-1)^2 - (p-1)^2] + \hat{R}_p[(q-1)^2 + (r-1)^2]}{2[\hat{R}_p(r-q) + \hat{R}_q(p-r) + \hat{R}_r(q-p)]} + \frac{\hat{R}_r[(p-1)^2 + (q-1)^2] \pm \sqrt{\hat{Z}}}{2[\hat{R}_p(r-q) + \hat{R}_q(p-r) + \hat{R}_r(q-p)]}, \tag{9}$$

where

$$\begin{aligned} \hat{Z} = {} & \hat{R}_p^2(q-r)^4 + \hat{R}_q^2(p-r)^4 + \hat{R}_r^2(p-q)^4 \\ & - 2\hat{R}_p\hat{R}_q[r^4 - 2r^3(p+q) + r^2(p^2+4pq+q^2) \\ & - 2r(pq^2+p^2q) + p^2q^2] \\ & - 2\hat{R}_q\hat{R}_r[p^4 - 2p^3(q+r) + p^2(q^2+4qr+r^2) \\ & - 2p(qr^2+q^2r) + q^2r^2] \\ & - 2\hat{R}_p\hat{R}_r[q^4 - 2q^3(p+r) + q^2(p^2+4pr+r^2) \\ & - 2q(pr^2+p^2r) + p^2r^2]. \end{aligned}$$

Taking into consideration the fact that estimates based on higher order moments show large variability, it is of interest to estimate the parameters from lower order moments by appropriately selecting the parameters k_1, k_2, and k_3. Thus, one can use fractional moments, i.e., of order other than a positive integer.

This estimation technique can be extended to the case where c is unknown as follows.

STEP 1. Set the parameter c to $c_0 > c$ (via initial guess).
STEP 2. Calculate the estimates of α_1, α_2, and β.
STEP 3. Repeat STEP 2 with $c_1 = c_0 - s$, where s is a step until α_1, α_2, and β are positive.

In the algorithm we take into consideration the fact that for c_0 greater than the true value of c the obtained estimates of α_1, α_2, and β may be negative or complex. Simulation results for $k_1 = 1.001$, $k_2 = 1.05$, $k_3 = 1$, and $c_0 = 6$ are given in Tables 2–4 where the parameters selections are equivalent to the ones in Figures 1–3, respectively. Averages were computed over 500 independent trials in each case.

From the results one concludes that there exists a strong relationship between the estimates of β and c, a weak relationship between the estimates of α_2 and c and practically no relationship between the estimates of α_1 and c. Also, there exists a relationship between the estimates of α_1 and α_2 which, as mentioned earlier, are interchangeable. An attempt has been made to improve upon β when $c < 2$ by using the maximum likelihood based result

$$\hat{\beta} = \exp\left[\frac{1}{N}\sum_{i=1}^{N}\log(x_i) - \frac{1}{\hat{c}}\left(\Psi(\hat{\alpha}_1) + \Psi(\hat{\alpha}_2)\right)\right],$$

where $\Psi(\cdot)$ is the digamma function, but no significant changes in the estimates were observed when compared to the ones obtained using moments.

5. Conclusions

A general distribution, called GBK-distribution, was derived for coherent modelling of radar clutter. The GBK-distribution is completely characterised by the set of four parameters. It includes a large number of popular distributions, being at the same time mathematically tractable. The estimation of the parameters of the GBK-distribution was investigated and an estimation method based on moments was proposed.

References

[1] E. Conte, M. Di Bisceglie, M. Longo, and M. Lops. Canonical Detection in Spherically Invariant Noise. *IEEE Transactions on Communications*, 43:347–353, 1995.

[2] E. Conte and M. Longo. Characterisation of Radar Clutter as a Spherically Invariant Random Process. *IEE Proceedings-F*, 134:191–197, 1987.

[3] A. Erdélyi *et. al. Tables of Integral Transforms*, volume I. McGraw-Hill, 1954.

[4] D. R. Iskander, A. M. Zoubir, and B. Boashash. Advances in Detection of Signals in Non-Gaussian Noise. Technical report, SPRC, QUT, December 1995. ISBN 1 86435 150 0.

[5] E. Jakeman and R. J. A. Tough. Generalized K distribution: a statistical model for weak scattering. *J. Opt. Soc. Am.*, 4:1764–1772, 1987.

[6] R. S. Raghavan. A Method for Estimating Parameters of K-Distributed Radar Clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 27:238–246, 1991.

[7] M. Rangaswamy, D. Weiner, and A. Öztürk. Non-Gaussian Vector Identification Using Spherically Invariant Random Processes. *IEEE Transactions on Aerospace and Electronic Systems*, 29:111–124, 1993.

[8] E. W. Stacy and G. A. Mihram. Parameter Estimation for a Generalized Gamma Distribution. *Technometrics*, 7:349–358, 1965.

[9] M. C. Teich and P. Diament. Multiply Stochastic Representations for K distributions and their Poisson Transforms. *J. Opt. Soc. Am.*, 6:80–91, 1989.

[10] A. M. Zoubir. Statistical Signal Processing for Application to Over-The-Horizon Radar. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 94*, volume VI, pages 113–116, Adelaide, Australia, 1994.

Table 1. Special Cases of the GBK-distribution, $f_X(x;\alpha_1,\alpha_2,\beta,c)$.

Type[1]	pdf	Ref.	SIRP
χ	$f_X(x;\frac{n}{4},\frac{n+2}{4},2,4)$	[8]	$n\leq 2$
χ^2	$f_X(x;\frac{n}{4},\frac{n+2}{4},2,2)$	[8]	$n\leq 4$
CG	$f_X(x;\frac{1}{2},1,2,4)$	[8]	Yes
exp	$f_X(x;\frac{1}{2},1,2a,2)$	[8]	Yes
Γ	$f_X(x;\frac{\nu}{2},\frac{\nu+1}{2},2a,2)$	[8]	$\nu\leq 2$
GΓ	$f_X(x;\frac{\nu}{2},\frac{\nu+1}{2},a2^{1/p},2p)$	[8]	$p\nu\leq 2$
GHG	$f_X(x;\frac{1}{2\nu},\frac{1+\nu}{2\nu},a2^{1/\nu},2\nu)$	[1]	Yes
GHL	$f_X(x;\frac{1}{2},\nu,\frac{2}{a},2)$	[1]	Yes
HG	$f_X(x;\frac{1}{4},\frac{3}{4},2,4)$	[8]	Yes
J&T	$f_X(x;\frac{n}{2},\alpha,\frac{2}{\beta},2)$	[5]	$n\leq 2$
K	$f_X(x;1,\nu+1,2a,2)$	[6]	Yes
K_0	$f_X(x;1,1,N,1)$	[9]	Yes
K	$f_X(x;1,\alpha,\frac{N}{\alpha_N},1)$	[9]	Yes
K'	$f_X(x;\beta,\alpha,\frac{N}{\alpha\beta},1)$	[9]	$\beta\leq 2$
R	$f_X(x;\frac{1}{2},1,2\sigma,4)$	[8]	Yes
R	$f_X(x;1,\nu\to\infty,2a,2)$	[6]	Yes
SG	$f_X(x;\frac{3}{4},\frac{5}{4},2,4)$	[8]	No
W	$f_X(x;\frac{1}{2},1,a2^{1/p},2p)$	[8]	$p\leq 2$

[1] CG–Circular Gaussian, Γ–Gamma, GΓ–Generalised Gamma, GHG–Generalised Half Gaussian, GHL–Generalised Half Laplace, HG–Half Gaussian, J&T–Jakeman & Tough's model, R–Rayleigh, SG–Spherical Gaussian, W–Weibull.

Table 2. Sample mean and variance of the parameter estimates of a GBK variate with $\alpha_1=1$, $\alpha_2=2$, $\beta=1$, and varying c for $N=500$.

	$\hat{\alpha}_1$		$\hat{\alpha}_2$	
c	$\widehat{\mathrm{E}[\hat{\alpha}_1]}$	$\widehat{\mathrm{Var}[\hat{\alpha}_1]}$	$\widehat{\mathrm{E}[\hat{\alpha}_2]}$	$\widehat{\mathrm{Var}[\hat{\alpha}_2]}$
1	1.0938	0.1694	1.2391	0.2409
2	1.1080	0.1617	1.2575	0.1904
3	1.0979	0.1939	1.3342	0.1892
5	1.1160	0.2025	1.5599	0.3208
	$\hat{\beta}$		$\hat{c}$	
c	$\widehat{\mathrm{E}[\hat{\beta}]}$	$\widehat{\mathrm{Var}[\hat{\beta}]}$	$\widehat{\mathrm{E}[\hat{c}]}$	$\widehat{\mathrm{Var}[\hat{c}]}$
1	2.1744	1.9842	1.1617	0.0723
2	1.3421	0.1959	2.2838	0.1976
3	1.1844	0.0643	3.3874	0.3285
5	1.0613	0.0189	5.3943	0.5649

Table 3. Sample mean and variance of the parameter estimates of a GBK variate with $\alpha_1=1$, $\alpha_2=2$, $c=2$, and varying β for $N=500$.

	$\hat{\alpha}_1$		$\hat{\alpha}_2$	
β	$\widehat{\mathrm{E}[\hat{\alpha}_1]}$	$\widehat{\mathrm{Var}[\hat{\alpha}_1]}$	$\widehat{\mathrm{E}[\hat{\alpha}_2]}$	$\widehat{\mathrm{Var}[\hat{\alpha}_2]}$
$\frac{1}{3}$	1.0941	0.1623	1.2381	0.1881
$\frac{1}{2}$	1.1115	0.1690	1.2514	0.1974
1	1.1262	0.1785	1.2836	0.1967
2	1.1366	0.1646	1.2813	0.1843
	$\hat{\beta}$		$\hat{c}$	
β	$\widehat{\mathrm{E}[\hat{\beta}]}$	$\widehat{\mathrm{Var}[\hat{\beta}]}$	$\widehat{\mathrm{E}[\hat{c}]}$	$\widehat{\mathrm{Var}[\hat{c}]}$
$\frac{1}{3}$	0.4549	0.0204	2.3038	0.1903
$\frac{1}{2}$	0.6756	0.0484	2.2894	0.1927
1	1.3234	0.2123	2.2705	0.2158
2	2.6448	0.7808	2.2592	0.1938

Table 4. Sample mean and variance of the parameter estimates of a GBK variate with $\alpha_2=2$, $\beta=1$, $c=2$, and varying α_1 for $N=500$.

	$\hat{\alpha}_1$		$\hat{\alpha}_2$	
α_1	$\widehat{\mathrm{E}[\hat{\alpha}_1]}$	$\widehat{\mathrm{Var}[\hat{\alpha}_1]}$	$\widehat{\mathrm{E}[\hat{\alpha}_2]}$	$\widehat{\mathrm{Var}[\hat{\alpha}_2]}$
$\frac{1}{2}$	0.4421	0.0178	0.6801	0.0515
1	1.0533	0.1260	1.1983	0.1358
2	2.0831	1.5862	2.3401	2.1653
3	2.4552	3.0989	2.7858	4.0916
	$\hat{\beta}$		$\hat{c}$	
α_1	$\widehat{\mathrm{E}[\hat{\beta}]}$	$\widehat{\mathrm{Var}[\hat{\beta}]}$	$\widehat{\mathrm{E}[\hat{c}]}$	$\widehat{\mathrm{Var}[\hat{c}]}$
$\frac{1}{2}$	1.8653	0.0670	2.7534	0.1080
1	1.4043	0.1748	2.3418	0.1853
2	1.1210	0.3175	2.0974	0.2458
3	1.2372	0.4245	2.1504	0.2786

Robust Processing of Heavy Tails Signals - Comparison of Approaches

Liam Galin and Hagit Messer
Department of Electrical Engineering - Systems, Tel Aviv University
Tel-Aviv 69978, Israel
Tel. +972-36408119, Fax. +972-36407095
E-mail: messer@eng.tau.ac.il

Abstract

This paper deals with robust estimation of AR parameters. We compare the performance of the LMS algorithm to the performance of two robust, adaptive algorithms: the LMAD algorithm of Shao and Nikias in which the error signal in the LMS algorithm is hard-limited before used to control the weights, and the LLMS algorithm in which the input process is soft-limited before the LMS algorithm is applied. The comparison is done in terms of rate of convergence and stability (steady state variance). We show that with a proper choice of limiting level, the LLMS algorithm outperforms the LMAD algorithms when applied to symmetric, α stable processes of $1 \leq \alpha \leq 2$.

1. Introduction and background

Recently, there is an increasing interest in signal processing of non-Gaussian processes. In general, a zero-mean symmetric distribution deviates from the normal distribution either by its local properties (about the origin) or by the fact that the tails of the probability density function (*pdf*) decay slower than the tails of the Gaussian *pdf*. While local features of the *pdf* are very sensitive to the presence of additive, Gaussian noise, the heavy tails signals preserve their nature even in the presence of such noise. In this paper we are focused on heavy-tails non-Gaussian processes, which reflect rare but strong values in the signal, i.e., impulsive-like processes.

The traditional approach to robust signal processing in the presence of impulsive noise involves passing the input through a non-linear device (such as a limiter) prior to the conventional, second order processing (e.g., [1]). Alternatively, Shao and Nikias [3] suggest to model a zero-mean, symmetric heavy tails process as an α-stable process [2] and to match an optimal procedure to the underline distribution. In this paper we study the differences between the two approaches by comparing the performance of the algorithms based on two approaches for adaptive estimation of AR parameters.

Consider the first order AR process:

$$x(n) = ax(n-1) + v(n) \tag{1}$$

where a is a constant to be estimated and $v(n)$ is a heavy-tails, zero-mean, symmetric process. We compare the performance of 3 adaptive algorithms for estimating a:

- **The LMS algorithm [4]:**
  In this algorithm, which is optimal for Gaussian processes, the estimate of the AR parameter a, $\hat{a}_{LMS}$, is the steady state solution of the difference equation:

  $$w_{LMS}(n+1) = w_{LMS}(n) + \mu x(n-1)e(n) \tag{2}$$

  where $e(n) = x(n) - w_{LMS}(n)x(n-1)$ and $w_{LMS}(0) = 0$.

- **The $LLMS$ algorithm:**
  Here the estimate of a, $\hat{a}_{LLMS}$, is the steady state solution of the difference equation:

  $$w_{LLMS}(n+1) = w_{LLMS}(n) + \mu x_L(n-1)e(n) \tag{3}$$

  where $e(n) = x_L(n) - w_{LLMS}(n)x_L(n-1)$ and $w_{LLMS}(0) = 0$. $x_L(n)$ is the limited input signal, i.e., $x(n)$ after passing through a limiter.

- **The $LMAD$ algorithm [3]:**
  In this algorithm, the estimate of a, $\hat{a}_{LMAD}$, is the steady state solution of the difference equation:

  $$w_{LMAD}(n+1) = w_{LMAD}(n) + \mu x(n-1)sign\{e(n)\} \tag{4}$$

  where $e(n) = x(n) - w_{LMAD}(n)x(n-1)$ and $w_{LMAD}(0) = 0$. This algorithm is based on modeling the signal as a symmetric α-stable signal and matching an algorithm to the assumed distribution. Note that the $LMAD$ algorithm of [3] suggests hard-limiting the error signal in the usual LMS algorithm (2).

0-8186-7576-4/96 $5.00

Table 1. Percentage of samples truncated by a limiter for various α and limit values

alpha	c=1	3.16	10	31.6	100	316.2	1000	3162.2
1.0	50.0248	6.3325	0.6288	0.0623	0.0060	0.0005	0.0001	0
1.1	49.6025	4.8290	0.3753	0.0297	0.0024	0.0002	0	0
1.2	49.2903	3.5778	0.2190	0.0142	0.0012	0.0001	0	0
1.3	49.1153	2.6348	0.1299	0.0070	0.0004	0	0	0
1.4	48.9086	1.8972	0.0720	0.0025	0	0	0	0
1.5	48.6815	1.3296	0.0411	0.0016	0	0	0	0
1.6	48.5489	0.8810	0.0219	0.0005	0	0	0	0
1.7	48.4386	0.5553	0.0114	0.0002	0	0	0	0
1.8	48.2056	0.3037	0.0042	0	0	0	0	0
1.9	48.0656	0.1295	0.0017	0	0	0	0	0
2.0	47.9901	0	0	0	0	0	0	0

The LMS algorithm is well known. The $LMAD$ algorithm is deeply discussed in [3]. We discuss the $LLMS$ algorithm in Section 2. In Section 3 we present the results of the comparison between the 3 algorithms and discuss them.

2. The $LLMS$ algorithm

The $LLMS$ algorithm presents the traditional approach for robust signal processing of impulsive-like signals, where rare, strong values are replaced by a pre-fixed value. In this paper we suggest the use of a soft limiter, so

$$\begin{aligned} x_L(t) &= x(t) \quad ; \quad |t| \le c \\ x_L(t) &= c_L \quad ; \quad |t| > c \end{aligned} \tag{5}$$

where c determines the limiting range and c_L determines the limiting value. The LMS algorithm is then applied to the limited input, $x_L(n)$. The main question is, how to choose the values of c and of c_L? To study the effect of a limiter, we have simulated zero-mean, symeteric α-stable processes of unit covariation.

Fig. 1 presents the average of the learning curves (with the same μ) of 200 runs of a processes with $\alpha = 1$ for which ω_{LLMS} of (3) as a function of n is plotted. For all different values of c, μ is chosen such that the rate of convergence of the algorithm is maximized while keeping its steady state variance smaller than a given value. It shows that the rate of convergence of the $LLMS$ algorithm increases as c decreases. Table 1 presents the percentage of samples truncated by a limiter for various values of α and of c, where $c_L = 0$. It shows that for $c > 300$, the number of truncated samples is practically negligible for any $1 \le \alpha \le 2$. The translation of this observation to the design of the $LLMS$ algorithm is not straight-forward since the effect of truncation on a sufficient statistic for estimating the AR parameters of a sequence is not simple. In Table 2 we present the mean and the variance of the steady-state estimate of the AR parameter $a = 0.99$ using the $LLMS$ algorithm for different values of c (here $c_L = 0$). We derive the statistics of the estimate based on 100 Monte-Carlo runs, at each of them 3000 samples were used, which are significantly more

Table 2. Steady state value and variance of the $LLMS$ for various α and c

c	$\alpha = 1.0$	$\alpha = 1.3$	$\alpha = 1.5$	$\alpha = 1.7$	$\alpha = 2.0$
1.0	0.968614	0.947430	0.941695	0.932215	0.888754
	(0.003705)	(0.004636)	(0.005341)	(0.006189)	(0.007360)
3.2	0.965060	0.939276	0.926273	0.919618	0.866334
	(0.003293)	(0.003610)	(0.003585)	(0.003737)	(0.004694)
10.0	0.957781	0.931475	0.917954	0.919880	0.895272
	(0.002835)	(0.002261)	(0.002955)	(0.002124)	(0.002302)
31.6	0.943861	0.930466	0.933893	0.945618	0.945512
	(0.002202)	(0.001670)	(0.000897)	(0.000408)	(0.000160)
100.	0.942769	0.955895	0.962565	0.973128	0.945521
	(0.001475)	(0.000315)	(0.000326)	(0.000035)	(0.000160)
316.2	0.961561	0.976227	0.981633	0.977140	0.945521
	(0.000371)	(0.000418)	(0.000007)	(0.000018)	(0.000160)
1000.0	0.979454	0.969481	0.982322	0.977140	0.945521
	(0.000047)	(0.003460)	(0.002821)	(0.000018)	(0.000160)
3162.3	0.986577	0.956853	0.982322	0.977140	0.945521
	(0.000778)	(0.028933)	(0.002821)	(0.000018)	(0.000160)
10000.0	0.943689	0.946341	0.982322	0.977140	0.945521
	(0.017883)	(0.018018)	(0.002821)	(0.000018)	(0.000160)

than the acquisition time of the algorithm, as shown in Fig. 1. It shows that at all values of $1 \le \alpha \le 2$, the steady state performance is the best for $300 > c > 100$.

Therefore, the well-known trade-off between speed of reaction and stability is also preserved in the $LLMS$ algorithm. The best choice of the parameter c which controls this trade-off by maximizing the rate of convergence without hurting much the steady-state performance of the algorithm, seems to be $c = 100$. Note, however, that this value is applicable to processes of unit covariation. For processes with different covaration, the value of c should be scaled accordingly.

The choice of c_L depends on the application. In our case, we try to match an AR model to the data. Since a first order AR process is unlikely to have very strong values, it is reasonable to suppress them by letting c_L be the median of the data (in our case, $c_L = 0$). Indeed, our simulation study shows that as c_L decreases, the steady state variance of the $LLMS$ algorithm decreases (see Fig. 2). In other applications, where the rare, strong values may better fit the problem, it may be more reasonable to assume $c_L = c$ or an in-between value as $c_L = 0.5c$.

3. Comparison of the 3 algorithms and discussion

We have simulated the 3 algorithms for $a = 0.99$ where $x(n)$ is a symmetric, normalized (to unit covariation) α-stable signal of 3000 samples. For the $LLMS$ algorithm we have determined $c_L = 0$ and $c = 100$, following the discussion in Section 2. The performance of the algorithms is evaluated from 100 Monte-Carlo runs. In Fig. 3 we present

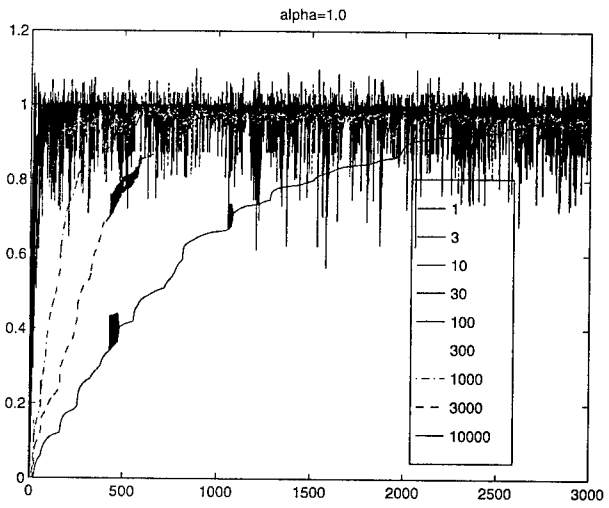


Figure 1. The learning curves (averaged over 200 runs with the same μ) of the $LLMS$ algorithm with different limiting levels ($c = 1, 3, 10, 30, 100, 300, 1000, 3000, 10000$, $c_L = 0$). The input is a symmetric α-stable process of $\alpha = 1$.

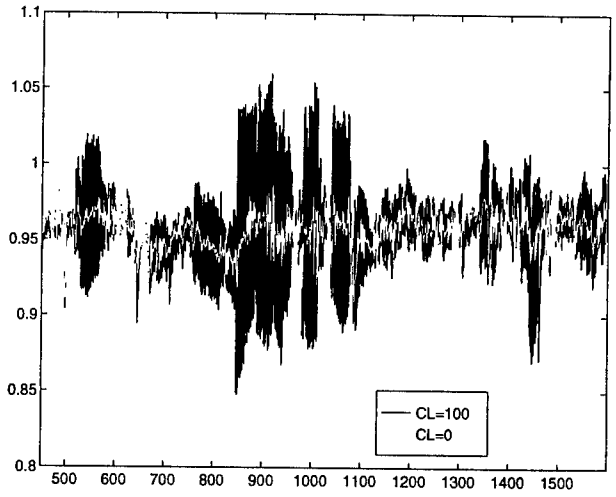


Figure 2. The steady state learning curves (averaged over 200 runs with the same μ) of the $LLMS$ algorithm with different limiting levels: $c_L = 0, 100$ ($c = 100$). The input is a symmetric α-stable process of $\alpha = 1.3$.

the average of 100 learning curves of the 3 algorithms for an α-stable process with $\alpha = 1.5$, built as for the experiment of Fig. 1. It confirms that with a heavy-tails signal, the $LMAD$ outperforms the LMS significantly, as first suggested in [3]. Note, however, that if μ is chosen under the same criterion (to maximize the rate of convergence of the algorithm while keeping its steady state variance smaller than a given value) **for each run separately** and then the learning curves of the 100 runs are averaged, the advantage of the $LMAD$ over the LMS is dramatically less significant. (See Fig. 4).

The difference between Fig. 3 and Fig. 4 is explained by the effect of a rare, large value of $x(n)$ on the LMS algorithm. For a given μ, the presence of such large value slows down the convergence of the algorithm dramatically. When μ is adjusted to each sequence separately, the spread of the μs over the 100 α-stable sequences can be shown to be large, but the averaged learning curve converges much faster than that of the averaged algorithm (with the same μ) and is much smoother.

In Fig. 5 we present the equivalent of Fig. 3 for the case where the input happens to be Gaussian ($\alpha = 2$). It shows that while the $LMAD$ algorithm is worse than the LMS algorithm, the $LLMS$ algorithm works similarly well to the LMS for Gaussian data.

That is, with heavy-tails signals, where the LMS algorithm fails, the $LLMS$ algorithm **outperforms** the $LMAD$ algorithm (but not significantly) in both rate of convergence and stability, independent of the presentations used (Fig. 3 and Fig. 4). With Gaussian data, the LMS algorithm outperforms the $LMAD$ but not the $LLMS$. Therefore, the $LLMS$ algorithm, when properly designed, is better than the $LMAD$ in handling heavy-tails signals in terms of robustness: its performance is as good as that of the $LMAD$ when applying to heavy-tails signal while performing as good as the LMS (and better than the $LMAD$) with Gaussian signals.

The advantage of the $LLMS$ over the $LMAD$ algorithm can be explained by comparing equations (3) and (4) to (2). In the LMS algorithm, the adaptation is controlled by the error signal $e(n)$ - if the error is large the adaptation step is larger, so it converges faster to the region of small errors, where the fine adaptation is done. By hard-limiting the error signal in the $LMAD$ algorithm one looses this automatic weighting of the adaptation step. The $LLMS$ algorithm, on the other hand, keeps this feature while handling spikes by limiting the dynamic range of the input.

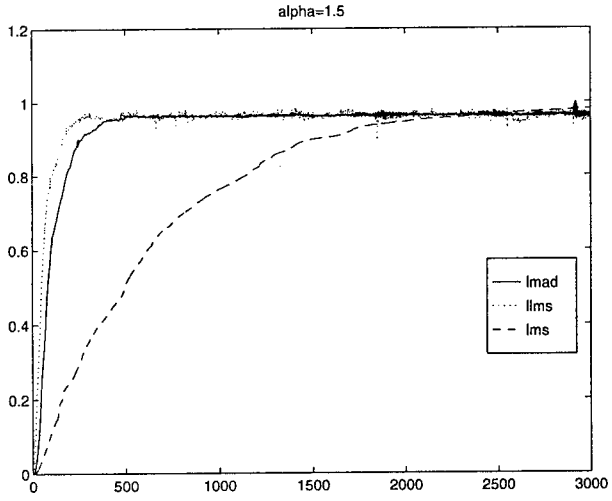


Figure 3. The learning curves (averaged over 100 runs with the same μ) of the LMS, the $LMAD$ and the $LLMS$ algorithms (with $c = 100$, $C_L = 0$). The input is a symmetric α-stable process of $\alpha = 1.5$.

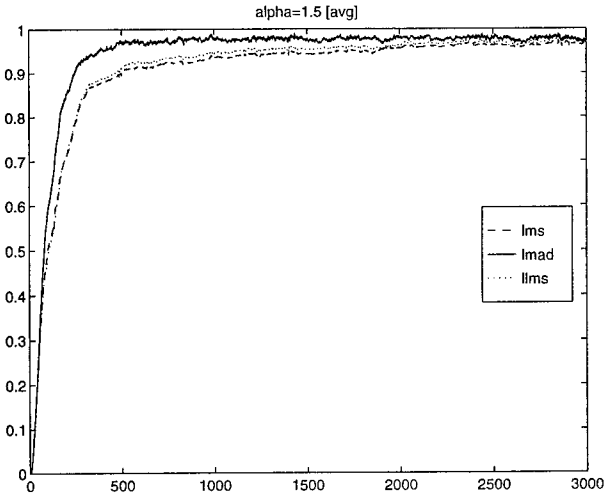


Figure 4. The averaged learning curves (averaged over 100 runs with optimizing μ for each run) of the LMS, the $LMAD$ and the $LLMS$ algorithms (with $c = 100$, $C_L = 0$). The input is a symmetric α-stable process ($\alpha = 1.5$).

References

[1] S. Kassam. *Signal Detection in Non-Gaussian Noise.* Springer, New York, 1988.

[2] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* Chapman and Hall, New York, 1994.

[3] M. Shao and C. Nikias. Signal processing with fractional lower-order moments: Stable processes and their applications. *Proceedings of the IEEE*, 81:986–1010, July 1993.

[4] B. Widrow and S. Stearns. *Adaptive Signal Processing.* Prentice Hall, Englewood Cliff, New Jersey, 1985.

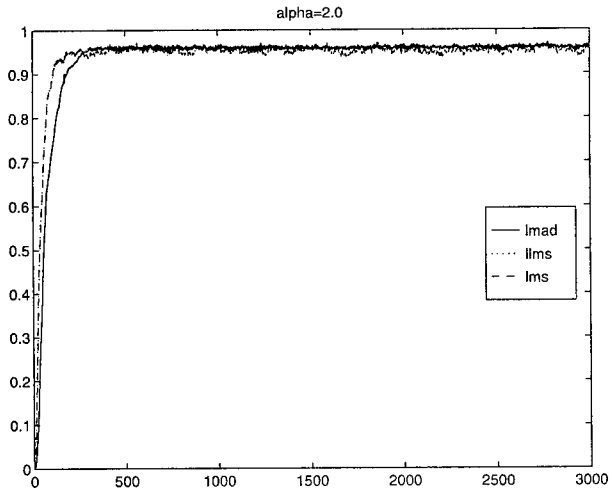


Figure 5. The averaged learning curves (averaged over 100 runs with the same μ) of the LMS, the $LMAD$ and the $LLMS$ algorithms (with $c = 100$, $C_L = 0$). The input is a symmetric α-stable process of $\alpha = 2$, i.e., a Gaussian process.

Frequency Estimation of Sinusoidal Signals in Alpha-Stable Noise Using Subspace Techniques*

Mustafa A. Altınkaya, Hakan Deliç†, Bülent Sankur, Emin Anarım
Signal and Image Processing Laboratory
Department of Electrical and Electronics Engineering
Boğaziçi University
Bebek 80815 Istanbul, Turkey
{altink, delic, sankur, anarim}@busim.ee.boun.edu.tr

Abstract

In the frequency estimation of sinusoidal signals observed in impulsive noise environments, techniques based on Gaussian noise assumption are unsuccessful. One possible way to find better estimates is to model the noise as an alpha-stable process and to use the fractional lower order statistics of data to estimate the signal parameters. In this work noise and signal subspace methods, namely MUSIC and Principal Component-Bartlett, are applied to fractional lower order statistics of sinusoids embedded in alpha-stable noise. The simulation results show that techniques based on lower order statistics are superior to their second order statistics-based counterparts, especially when the noise exhibits a strong impulsive attitude.

1. Introduction

Most of the work on the frequency estimation problem assumes that the additive noise has Gaussian distribution. This is partly because of the nice properties of the Gaussian model which allows for simplification of the theoretical work and decreases the computational complexity in signal parameter estimation. As long as the noise distribution can fit approximately to a Gaussian model, in particular for the tails of the distribution, one can obtain good estimators with the Gaussian noise assumption. But if the noise process belongs to a non-Gaussian, especially a heavily-tailed, distribution class or when the noise is of impulsive nature, parameter estimators which are based on Gaussian noise assumption break down.

Impulsive noise processes can be modeled using stable distributions. If a signal can be thought of as the sum of a large number of independent and identically distributed random variables, the limiting distribution will be in the class of stable distributions according to Generalized Central Limit Theorem [5], and stable distributions cover Gaussian distribution in the limit.

If the additive noise has a heavily-tailed distribution which is successfully modeled by alpha-stable distributions, the performance of covariation-based frequency estimators is better than that of the traditional estimators which are based on second order statistics.

In this work subspace-based estimation methods using covariations are considered. In Section 2, the $S\alpha S$ distributions are briefly discussed. In Section 3, the application of fractional lower order moments (FLOM) to frequency estimation problem is presented. Section 4 covers the results of the simulation experiments. Finally conclusions are in Section 5.

2. $S\alpha S$ Distributions

An important sub-class of stable distributions are symmetric alpha-stable ($S\alpha S$) distributions. The characteristic function of $S\alpha S$ variables is given by:

$$\phi(\omega) = \exp\{j\delta\omega - \gamma|\omega|^{\alpha}\} \qquad (1)$$

where α is the characteristic exponent ($0 < \alpha \leq 2$), δ is the location parameter ($-\infty < \delta < \infty$) and γ is the dispersion ($\gamma > 0$). Without losing generality we may take the location parameter $\delta = 0$ as in the zero mean Gaussian noise assumption case. This assumption will

*This work was supported by TÜBİTAK under contracts EEEAG-83 and EEEAG-139.

†On leave from the Department of Electrical and Computer Engineering, University of Southwestern Louisiana, Lafayette, LA 70504-3890, USA.

lead to the characteristic function:

$$\phi(\omega) = \exp\{-\gamma|\omega|^{\alpha}\}. \quad (2)$$

For $S\alpha S$ processes only the moments of order $p < \alpha$ exist. So the estimation methods based on second order statistics of the data cannot be applied. One solution is to use FLOM of the process [5]. The so-called covariations [4] of two random variables are used instead of second order moments in the analysis. The covariation of two jointly $S\alpha S$ real random variables with dispersions γ_x and γ_y are given as:

$$[X,Y]_{\alpha} = \frac{E[XY^{<p-1>}]}{E[|Y|^p]}\gamma_y \quad (3)$$

where $\gamma_y = [Y,Y]_{\alpha}$ is the dispersion of random variable Y and $Y^{<p-1>} = |Y|^{p-2}Y$.

3. Frequency Estimation Problem

In the frequency estimation problem the signal model assumed consists of multiple sinusoids

$$s_n = \sum_{k=1}^{K} A_k \sin\{\omega_k n + \theta_k\} \quad (4)$$

observed in additive $S\alpha S$ noise

$$x_n = s_n + z_n, \qquad n = 1, \cdots, N. \quad (5)$$

where A_k is the amplitude, ω_k is the angular frequency, and θ_k is the phase of the kth real sinusoid. K is the number of sinusoids and N is the sample size. x_n and z_n are realizations of observation sequence X_n and $S\alpha S$ noise sequence Z_n, respectively.

When the noise samples are independent and identically distributed, the observation sequence can be modeled as a stable AR-process:

$$X_n = a_1 X_{n-1} + \cdots + a_M X_{n-M} + b_0 Z_n. \quad (6)$$

This leads to the Generalized Yule-Walker Equation when X_{n-m} is given as [5]:

$$\begin{aligned} E[X_n|X_{n-m}] &= a_1 E[X_{n-1}|X_{n-m}] + \cdots \\ &\quad + a_M E[X_{n-M}|X_{n-m}], \quad (7) \\ E[X_{n+l}|X_n] &= \lambda(l) X_n \quad (8) \end{aligned}$$

where $m = 1, \cdots, M$. If $\lambda(l)$ denotes the covariation coefficient of X_{n+l} with X_n, one can find the AR-parameters by solving the following linear set of equations:

$$\mathbf{Ca} = \boldsymbol{\lambda} \quad (9)$$

with

$$\mathbf{C} = \begin{bmatrix} \lambda(0) & \lambda(-1) & \cdots & \lambda(1-M) \\ \lambda(1) & \lambda(0) & \cdots & \lambda(2-M) \\ \vdots & \vdots & \vdots & \vdots \\ \lambda(M-1) & \lambda(M-2) & \cdots & \lambda(0) \end{bmatrix},$$

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix}, \qquad \boldsymbol{\lambda} = \begin{bmatrix} \lambda(1) \\ \lambda(2) \\ \vdots \\ \lambda(M) \end{bmatrix}.$$

In the frequency estimation of sinusoids given by the Equations 4 and 5 the sinusoidal signal component can be assumed to be a stable AR process of order $2K$. As in the Gaussian additive noise case, the model order M of the AR model for the noisy signal should be selected higher than $2K$ in order to allow sufficient additional subspace dimension for the noise component. Assuming that the signal and the noise components are stable processes with the same characteristic exponent, their covariation can be calculated as follows:

$$\begin{aligned} [x_j, x_k]_{\alpha} &= [s_j + e_j, s_k + e_k]_{\alpha} \quad (10) \\ &= [s_j, s_k]_{\alpha} + [s_j, e_k]_{\alpha} \\ &\quad + [e_j, s_k]_{\alpha} + [e_j, e_k]_{\alpha} \end{aligned}$$

where $j, k = 1, \ldots, N$. Since the signal and additive noise are assumed to be independent, the cross-covariation of noise and signal components with each other is

$$\begin{aligned} [s_j, e_k]_{\alpha} &= 0 \quad (11) \\ [e_j, s_k]_{\alpha} &= 0. \end{aligned}$$

On the other hand the covariations of the signal component and noise component with themselves are found as:

$$\begin{aligned} [s_j, s_k]_{\alpha} &= \lambda(j-k)\gamma_{s_k} \quad (12) \\ [e_j, e_k]_{\alpha} &= \delta_{j,k}\gamma_{e_k} \quad (13) \end{aligned}$$

where $\delta_{j,k}$ is the Kronecker delta.

The covariation matrix for alpha-stable processes has the same meaning as that of the covariance matrix for Gaussian processes. As one performs eigen-decomposition of the covariation matrix, the larger eigenvalues will correspond to signal subspace eigenvectors and the remaining eigenvectors will constitute the noise subspace. So one can perform eigen-analysis on the covariation matrix and then apply a suitable

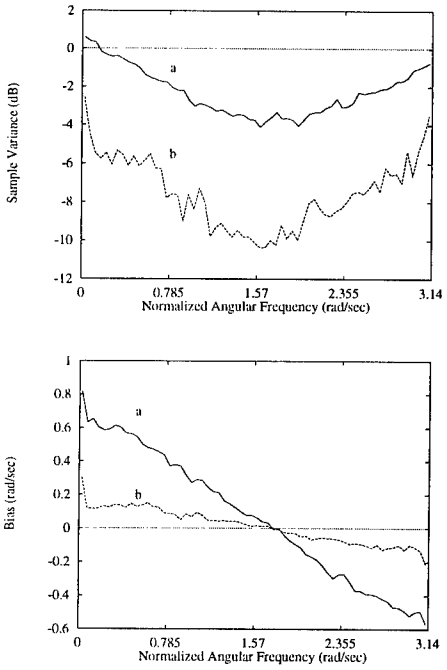


Figure 1. Sample variance and bias of PC-Bartlett and ROC-Bartlett frequency estimators versus normalized angular frequency, a) PC-Bartlett, b) ROC-Bartlett ($\alpha = 1.0$, $p = 0.8$ (ROC-Bartlett), $M = 20$, GSNR $= 5$ dB, $N = 50$, 100 noise realizations, 20 phase realizations).

noise subspace or a signal subspace technique to estimate the parameters. Note that the covariation matrix is not symmetric. This makes the eigen-analysis more complicated and renders many of the subspace-based parameter estimation techniques developed for Gaussian processes unsuitable for the general alpha-stable processes.

One such technique applied to direction of arrival estimation problem is the Robust Covariation-Based MUSIC (ROC-MUSIC) [6]. In this work, we first apply ROC-MUSIC which is a noise subspace method to frequency estimation in alpha-stable environments problem and then we also apply Robust Covariation-Based-Bartlett (ROC-Bartlett) which is a signal subspace method, to the problem.

The second order statistics-based principal component Bartlett frequency estimate is obtained by the peaks of the spectrum estimator [3]:

$$\text{PC-Bartlett}(\omega) = \frac{1}{M}\sum_{i=1}^{2K} \lambda_i \left|\mathbf{d}^H \mathbf{v}_i\right|^2 \qquad (14)$$

where $\mathbf{d}$ is the complex sinusoidal vector $\mathbf{d} = [1 \;\; \exp\{j\omega\} \;\; \cdots \;\; \exp\{j\omega(M-1)\}]$, and λ_i and $\mathbf{v}_i$ are

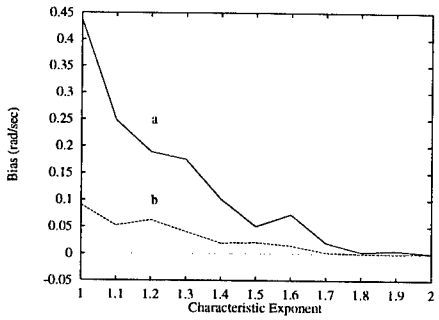


Figure 2. Bias of MUSIC and ROC-MUSIC frequency estimators versus characteristic exponent of alpha-stable noise, a) PC-Bartlett and MUSIC, b) ROC-Bartlett and ROC-MUSIC ($\omega = 0.76$ rad/sec, $M = 20$, GSNR $= 5$ dB, $N = 50$, 100 noise realizations, 20 phase realizations).

ordered eigenvalues such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$, and the corresponding eigenvectors of $M \times M$ autocorrelation matrix. ROC-Bartlett is obtained by substituting the covariation matrix for the autocorrelation matrix.

4. Simulation Experiments

We have used ROC-MUSIC and ROC-Bartlett methods to estimate the frequency of a single real sinusoid. The modified FLOM (MFLOM) estimator given by [6]

$$\hat{\mathbf{C}}(k,l) = \frac{\sum_{i=1}^{N-M+1} X_{k+i-1}|X_{l+i-1}|^{p-2}X_{l+i-1}}{\sum_{i=1}^{N-M+1}|X_{l+i-1}|^p}, \quad k,l = 1,\cdots,M, \qquad (15)$$

is defined for moment order $p \in [0,2]$ and it is used to estimate the (k,l)th element of the sample covariation matrix $\hat{\mathbf{C}}$. M denotes the order of AR-model. We have applied $S\alpha S$ noise sequences with varying α and γ parameters. To generate the $S\alpha S$ noise process we used the method described by Tsihrintzis and Nikias [7] which is a special case of the more general method including the non-symmetric alpha-stable random variable generation given by Chambers, Mallows and Stuck [2]. The moment order p and the sample size N were equal to 0.8 and 50, respectively. The AR-model order was chosen as 20 in the simulations. The *generalized* SNR, GSNR $= 10\log\left(\frac{1}{\gamma N}\sum_{n=1}^{N}|s(n)|^2\right)$ is equal to 5 dB.

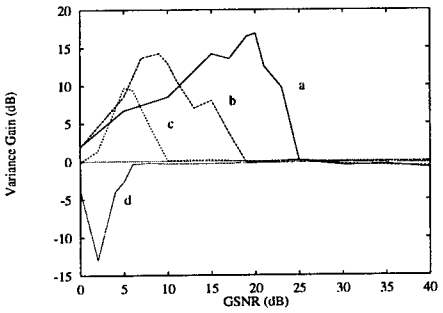


Figure 3. Variance reduction of ROC-Bartlett with respect to PC-Bartlett frequency estimator versus GSNR averaged on the frequency axis, a) $\alpha = 1.0$, b) $\alpha = 1.4$, c) $\alpha = 1.8$, d: $\alpha = 2.0$ ($M = 20$, $N = 50$, 100 noise and phase realizations).

4.1. Frequency Dependence of Bias and Variance

In Figure 1 the sample variance and the bias of PC-Bartlett and ROC-Bartlett frequency estimators are plotted against the angular frequency for $\alpha = 1.0$ (Cauchy distribution) and GSNR = 5 dB. The number of noise realizations and phase realizations are 100 and 20, respectively, making a total of 2000 Monte Carlo runs. The ROC-Bartlett has approximately 5 dB lower sample variance than the PC-Bartlett.

The bias curves depict a symmetry around approximately $\omega = 1.7$ rad/sec. The ROC-Bartlett performs much better than the PC-Bartlett. The difference of their bias value is more than 0.4 rad/sec around $\omega = 0.2$ rad/sec.

4.2. Dependence of Bias upon α

The bias behaviour of the estimators for $\omega = 0.76$ rad/sec as a function of the characteristic exponent α of the noise is shown in Figure 2. The figure indicates that the bias gets smaller as α increases. When $\alpha = 1$ the bias values are 0.45 rad/sec for PC-Bartlett and MUSIC and it is less than 0.1 rad/sec for their ROC versions. As this figure depicts for the single tone case as in our experiments, MUSIC and Bartlett estimators show exactly the same performance.

4.3. Dependence of Variance Reduction upon the GSNR

In Figure 3, the variance reduction achieved by ROC-Bartlett with respect to PC-Bartlett is plotted against GSNR for different values of α. The number of Monte Carlo runs is 100, each with a different noise and phase realization. The curve exhibiting the highest gain belongs to $\alpha = 1.0$ (Cauchy noise). This gain is approximately 17 dB when GSNR = 20 dB. The curves show that the variance increase introduced by the ROC-Bartlett against PC-Bartlett is negligible with the exception of Gaussian noise case where the GSNR threshold of ROC estimator is higher with respect to that of the second order statistics-based estimator. This behaviour validates the robustness of FLOM-based subspace techniques and it is also shared by the noise subspace technique ROC-MUSIC.

5. Conclusion

When the additive noise in the frequency estimation problem can be modeled as an alpha-stable process, the FLOM-based subspace techniques perform better than their second order statistics-based counterparts. Both ROC-MUSIC and ROC-Bartlett methods showed superior performance with respect to MUSIC and PC-Bartlett methods in our experiments, especially for low α values.

References

[1] M. A. Altınkaya. *Frequency Estimation of Sinusoidal Signals.* PhD thesis, Boğaziçi University, Istanbul, Turkey, 1996.

[2] J. M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. *Journal of Amererican Statistical Association*, 71:340–346, 1976.

[3] S. M. Kay. *Modern Spectal Estimation: Theory and Application.* Prentice Hall, Englewood Cliffs, New Jersey 07632, 1988.

[4] G. Miller. Properties of certain symmetric stable distribution. *Journal of Multivariate Analysis*, 8:346–360, 1978.

[5] M. Shao and C. L. Nikias. Signal processing with fractional lower order moments: Stable processes and their applications. *Proceedings of the IEEE*, 81:986–1010, 1993.

[6] P. Tsakalides and C. L. Nikias. Subspace-based direction finding in alpha-stable noise environments. In *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing*, Detroit, Michigan, U.S.A., May 8-12 1995.

[7] G. A. Tsihrintzis and C. L. Nikias. Performance of optimum and suboptimum receivers in the precence of impulsive noise modeled as an alpha-stable process. *IEEE Transactions on Communications*, 43(2/3/4):904–914, February/March/April 1995.

DATA-ADAPTIVE ALGORITHMS FOR SIGNAL DETECTION IN IMPULSIVE NOISE MODELED AS A SUBGAUSSIAN, ALPHA-STABLE PROCESS

George A. Tsihrintzis

Communication Systems Lab
Department of Electrical Engineering
University of Virginia
Charlottesville, VA 22903-2442

Chrysostomos L. Nikias

Signal and Image Processing Institute
Department of Electrical Engineering-Systems
University of Southern California
Los Angeles, CA 90089-2564

ABSTRACT

We address the problem of coherent detection of a signal embedded in heavy-tailed noise modeled as a subGaussian, alpha-stable process. We assume that the signal is a complex-valued vector of length L, known only within a multiplicative constant. The dependence structure of the noise, i.e., the underlying matrix of the subGaussian process, is not known. The intent is to implement a generalized likelihood ratio detector which employs robust estimates of the unknown noise underlying matrix and the unknown signal strength. The performance of the proposed adaptive detector is compared to that of an adaptive matched filter that uses Gaussian estimates of the noise underlying matrix and the signal strength and is found to be clearly superior. The proposed new algorithms are evaluated via Monte-Carlo simulation.

Key words - *Signal detection, subGaussian process, adaptive matched filter*

1. INTRODUCTION

The design of modern signal processing systems includes the design of signal detectors that will operate in noise/interference that is inherently nonGaussian and rather follows some distribution with tails that are significantly heavier than the tails of the Gaussian distribution. Such interference is termed "impulsive" and is characterized by a significant probability of its attaining high values. In an impulsive operational environment, traditional Gaussian receivers will perform very poorly and exhibit a number of false alarms or misses that is unacceptably high. Thus, a need arises to design receivers that maintain high performance when operating in the radar environment and are robust to fluctuations in the characteristics of the interference. This task can be achieved only if good statistical models are available to quantify the interference.

Classical statistical-physical models for impulsive interference have been proposed by Middleton [2, 4, 3, 5] and are based on the filtered-impulse mechanism. These models include three different classes of interference, namely A, B, and C. Interference in class A is "coherent" in narrowband receivers, causing a negligible amount of transients. Interference in class B, however, is "impulsive," consisting of a large number of overlapping transients. Finally, interference in class C is the sum of the other two interferences. The Middleton model has been shown to describe real impulsive interferences with high fidelity; however, it is mathematically involved for signal processing applications. This is particularly true of the class B model, which contains seven parameters, one of which is purely empirical and in no way relates to the underlying physical model. Moreover, mathematical approximations need to be used in the derivation of the Middleton model, which are equivalent to changes in the assumed physics of the noise and lead to ambiguities in the relation between the mathematical formulae and the physical scenario [1]. Very recently, an alternative to the Middleton model was proposed, which was based on the theory of symmetric, α-stable (SαS) distributions [8, 6].

In particular, it was shown in [9, 6] that, under very general assumptions, the first order distribution of impulsive interference follows a SαS law. The stable model was then tested with a variety of real data and found, in all cases examined, to match the data with high fidelity [9]. The performance of optimum and suboptimum receivers in the presence of SαS impulsive interference was examined in [12], both theoretically and via Monte-Carlo simulation, and a method was presented for the real time implementation of the optimum nonlinearities. From this study, it was found that the corresponding optimum receivers perform in the presence of SαS impulsive interference quite well, while the performance of Gaussian and other suboptimum receivers is unacceptably low. It was also shown that a receiver designed on a Cauchy assumption for the first order distribution of the impulsive interference performed only slightly below the corresponding optimum receiver, provided that a reasonable estimate of the noise dispersion was available, which for real-time signal processing purposes could be obtained via the fast algorithms in [11].

The study in [12] was later extended to include the optimum demodulation algorithm for reception of signals with random phase in impulsive intereference [13], as well in the direction of asymptotically optimum, multichannel detection structures for reception of amplitude-fluctuating

This work was supported by the Office of Naval Research under contract N00014-92-J-1034.

bandpass signals [14]. In all cases, the key finding has been the same robustness result for Cauchy-based algorithms as opposed to their Gaussian counterparts.

In this paper, we look at the problem of coherent detection of a signal embedded in heavy-tailed noise modeled as a subGaussian, alpha-stable process. SubGaussian processes are a special class of multidimensional alpha-stable processes which can efficiently model the presence of outliers, as well as a wide range of dependence structures in time series. We assume that the signal is a complex-valued vector of length L, known only within a multiplicative constant. The dependence structure of the noise, i.e., the underlying matrix of the subGaussian process, is not known. The intent is to implement an adaptive detector in which robust estimates of the noise underlying matrix and the signal strength are obtained from independent, multiple observations. The performance of the proposed adaptive detector is compared to that of an adaptive matched filter that employs Gaussian estimates of the noise covariance matrix and the signal strength [15]. More specifically, the paper is organized as follows: Section 2 provides a brief review of the basic definitions and properties of *subGaussian* SαS processes. In Section 3, we derive adaptive algorithms for detection of a (within a multiplicative constant) known signal in subGaussian noise of unknown underlying matrix. In Section 4, we illustrate the performance of the proposed detector in a computer simulation study in which we also compare it to the adaptive matched filter performance. We summarize the paper, draw conclusions, and suggest possible future research topics in Section 5.

2. SUBGAUSSIAN SYMMETRIC, ALPHA-STABLE PROCESSES

A subGaussian random vector $\underline{X}$ can be defined as a random vector with characteristic function of the general form

$$\phi(\underline{\omega}) = \exp[-\frac{1}{2}(\underline{\omega}^T \underline{\underline{R}}\underline{\omega})^{\alpha/2}], \qquad (1)$$

where $\underline{\underline{R}}$ is a positive-definite matrix. Unfortunately, closed-form expressions for the joint pdf of subGaussian random vectors are known only for the Gaussian ($\alpha = 2$) and Cauchy ($\alpha = 1$) cases:

$$f_G(\underline{X}) = \frac{1}{\sqrt{(2\pi)^L ||\underline{\underline{R}}||}} \exp(-\underline{X}^T \underline{\underline{R}}^{-1} \underline{X}) \text{ (Gaussian)} \qquad (2)$$

$$f_C(\underline{X}) = \frac{c||\underline{\underline{R}}||^{-1/2}}{[1 + \underline{X}^T \underline{\underline{R}}^{-1} \underline{X}]^{(L+1)/2}} \text{ (Cauchy)} \qquad (3)$$

where L is the length of the random vector, $||\underline{\underline{R}}||$ is the determinant of $\underline{\underline{R}}$, and $c = \frac{1}{\pi^{(L+1)/2}} \Gamma(\frac{L+1}{2})$.

The following proposition relates Gaussian and subGaussian random vectors and can, in fact, be used to generate subGaussian random deviates [7, pp. 77-84]:

Theorem 1 *Any subGaussian random vector is a SαS random vector. In addition, any subGaussian random vector can be expressed in the form*

$$\underline{X} = w^{\frac{1}{2}} \underline{G}, \qquad (4)$$

where w is a positive $\frac{\alpha}{2}$-stable random variable [7] and $\underline{G}$ is a Gaussian random vector of mean zero and covariance matrix $\underline{\underline{R}}$.

SubGaussian SαS processes combine the capability to model statistical dependence with the capability to model the presence of outliers in observed time series of various degrees of severity. The example in Fig.1 is indicative of the concept. Consider a subGaussian vector of length $L = 100$ and diagonal underlying covariance matrix $\underline{\underline{R}} = \text{diag}\{1, 1, \ldots, 1\}$. Typical realizations of the vector are shown in Figs. 1(a) and 1(b) for characteristic exponents $\alpha = 2$ and $\alpha = 1.5$, respectively. Clearly, it is difficult to distinguish one vector from the other visually. However, if we look over 1000 independent realizations of the first component of the vector, we obtain Figs. 1(c) and 1(d), respectively, in which a clear difference is observed.

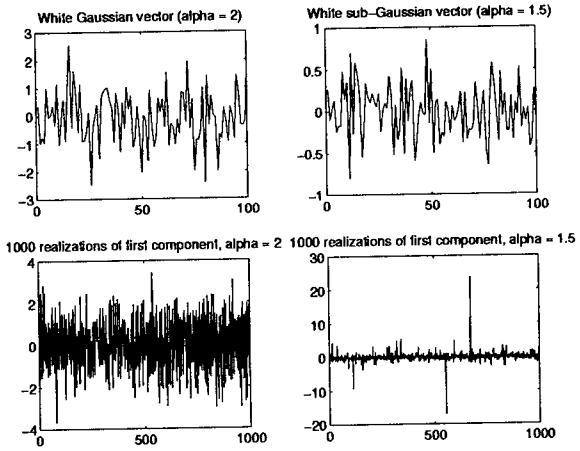


Figure 1: Typical realizations of subGaussian random vectors

The following proposition expresses the underlying matrix of a subGaussian vector in terms of its covariation matrix and can, therefore, be used to obtain high quality estimates of the underlying matrix of the vector from independent observations [7, pp. 89].

Theorem 2 *Let $\underline{X} = [X_1, X_2, \ldots, X_L]^T$ be a subGaussian random vector with underlying matrix $\underline{\underline{R}}$. Then, its covariation matrix $\underline{\underline{C}}$ will consist of the elements*

$$C_{ij} \equiv [X_i, X_j]_\alpha = 2^{-\frac{2}{\alpha}} R_{ij} R_{jj}^{\frac{\alpha-2}{2}} \qquad (5)$$

Eq.(5) can now be used to compute an estimate of the underlying matrix $\underline{\underline{R}}$ from the estimate of the covariation matrix $\underline{\underline{C}}$.

Theorem 3 *Let*

$$\hat{C}_{ij} = C(p, \alpha)[\frac{1}{K} \sum_{k=1}^{K} X_i^k (X_j^k)^{<p-1>}]^{\alpha/p} [\frac{1}{K} \sum_{k=1}^{K} |X_j^k|^p]^{\alpha/p-1}$$

be the estimator of the covariation matrix elements, where $p < \alpha/2$. *The estimates*

$$\begin{aligned} \hat{R}_{jj} &= [2^{\frac{\alpha}{2}}\hat{C}_{jj}]^{\frac{2}{\alpha}} \\ \hat{R}_{ij} &= 2^{\frac{\alpha}{2}}\hat{C}_{ij}/\hat{R}_{jj}^{\frac{\alpha-2}{2}} \end{aligned} \tag{6}$$

are consistent and asymptotically normal with means R_{jj} *and* R_{ij}, *respectively, and variances as in [10].*

The procedure is illustrated with the following simulation study: Consider a subGaussian random vector of length $L = 32$ and underlying matrix $\underline{\underline{R}} = \text{diag}\,\{1, 1, \ldots, 1\}$. We assume that $K = 500$ independent realizations of the vector are available and plot the 16^{nd} row of the mean over 100 Monte-Carlo simulations of the following two estimates:

$$\hat{\underline{\underline{R}}} = \frac{1}{K}\sum_{k=1}^{K}\underline{X}^k\underline{X}^{kT} \tag{7}$$

$$\hat{\underline{\underline{R}}} = \text{as obtained from covariation matrix estimate} \tag{8}$$

We examined the cases of $\alpha = 2$ and $\alpha = 1.5$. Clearly, the Gaussian estimate fails when $\alpha = 2$, while the covariation-based estimate maintains high performance in both the cases of $\alpha = 2$ and $\alpha = 1.5$.

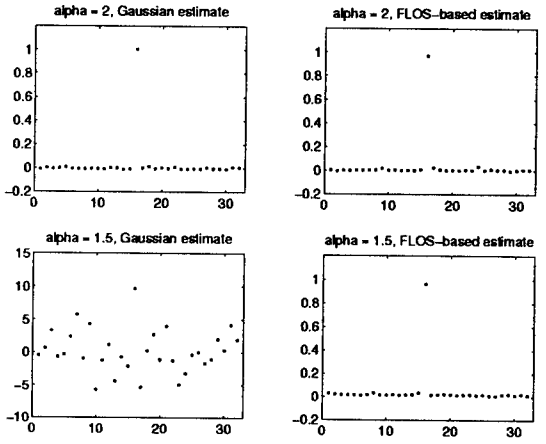


Figure 2: Illustration of the performance of estimators of the underlying matrix of a subGaussian vector

Next, we consider the estimation of the amplitude of a signal of known shape embedded in subGaussian noise from a number of independent observations. The following Proposition outlines the procedure and states its performance.

Theorem 4 *Consider the collection of* K *vectors* $\underline{X}^k = A\underline{s} + \underline{N}^k$, $k = 1, 2, \ldots, K$, *where* $\underline{s}^T\underline{s} = 1$. *Form the least-squares estimates* $\hat{A}_k = \underline{s}^T\underline{X}^k = \underline{s}^TA\underline{s} + \underline{s}^T\underline{N}^k = A + \underline{s}^T\underline{N}^k$, $k = 1, 2, \ldots, K$. *Define* $\hat{A} = sm\,\{\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_K\}$, *where* $sm\{\cdots\}$ *indicates the sample median of its arguments. The estimate* $\hat{A}$ *is consistent and asymptotically normal with mean equal to the true signal amplitude* A *and variance* $\frac{1}{K}[\frac{\pi\alpha\gamma^{1/\alpha}}{2\Gamma(1/\alpha)}]^2$, *where* $\gamma = 2^{-\frac{2}{\alpha}}(\underline{s}^T\underline{\underline{R}}\underline{s})^{\frac{\alpha}{2}}$.

3. DATA-ADAPTIVE ALGORITHMS FOR COHERENT SIGNAL DETECTION

We consider the hypothesis testing problem

$$\begin{aligned} H_0 &: \underline{X}^k = \underline{N}^k \\ H_1 &: \underline{X} = \underline{S} + \underline{N}^k, \end{aligned} \qquad k = 1, 2, \ldots, K$$

where all the vectors have dimension (length) L and $k = 1, 2, \ldots, K$ indexes independent, identically distributed realizations.

We make the following assumptions:

1. The noise vectors $\underline{N}^k$ have a sub-Gaussian distribution, i.e.,
$$\underline{N}^k = w_k^{\frac{1}{2}}\underline{G}^k,$$
where w_k is a positive $(\alpha/2)$-stable random variable of unit dispersion, $\underline{G}^k$ is a Gaussian random vector of covariance matrix $\underline{\underline{R}}$, and w_k and $\underline{G}^k$ are independent.
2. The signal vector $\underline{S} = A\underline{s}$ consists of a known shape $\underline{s}$ (for which $\underline{s}^T\underline{s} = 1$) and an unknown amplitude A.

The proposed test statistic is a generalized likelihood ratio test that makes use of the multidimensional Cauchy pdf defined in Eq.(2):

$$t_C = \sum_{k=1}^{K}\log\left[\frac{1 + \underline{X}^T\hat{\underline{\underline{R}}}^{-1}\underline{X}}{1 + (\underline{X} - \hat{A}\underline{s})^T\hat{\underline{\underline{R}}}^{-1}(\underline{X} - \hat{A}\underline{s})}\right] \tag{9}$$

For the estimates $\hat{\underline{\underline{R}}}$ and $\hat{A}$, we choose the estimates proposed in Eq.(6) and Proposition 4, respectively.

Assuming Gaussian noise of unknown covariance matrix $\underline{\underline{R}}$ and unknown signal amplitude, the data-adaptive detector attains the form of an adaptive matched filter [15], i.e., it computes the test statistic

$$t_G = (2/K)\sum_{k=1}^{K}(\hat{A}\underline{s})^T\hat{\underline{\underline{R}}}^{-1}\underline{X} - |\hat{A}|^2\underline{s}^T\hat{\underline{\underline{R}}}^{-1}\underline{s}, \tag{10}$$

where $\hat{A} = (1/K)\sum_{k=1}^{K}\frac{\underline{s}^T\underline{\underline{R}}^{-1}\underline{X}^k}{\underline{s}^T\underline{\underline{R}}^{-1}\underline{s}}$ and $\hat{\underline{\underline{R}}} = (1/K)\sum_{k=1}^{K}(\underline{X} - \hat{A}\underline{s})(\underline{X} - \hat{A}\underline{s})^T$.

4. COMPUTER ILLUSTRATION

The small sample performance of both the Gaussian and the proposed Cauchy detectors can be accurately assessed only via Monte-Carlo simulation. To this end, we chose an observation vector of length $L = 8$ and $K = 10$ independent copies of it, while for the signal we chose a shape of a square pulse of height $1/\sqrt{L}$ and an amplitude of $A = 1$. The sub-Gaussian interference was assumed to be of characteristic exponent $\alpha = 2, 1.75, 1.5, 1.25, 1$, and 0.75 and underlying matrix $\underline{\underline{R}} = \text{diag}\,\{1, 1, \ldots, 1\}$. The performance of the Gaussian and the Cauchy detectors was assessed via 10,000 Monte-Carlo runs.

In Fig. 3, we compare the performance of the Gaussian and the Cauchy detectors for different values of the characteristic exponent α. We see that, for $\alpha = 2$, the Gaussian detector, as expected, outperforms the Cauchy detector; however, for all other values of α, the Cauchy detector maintains a high performance level, while the performance of the Gaussian detector deteriorates down to unacceptably low levels.

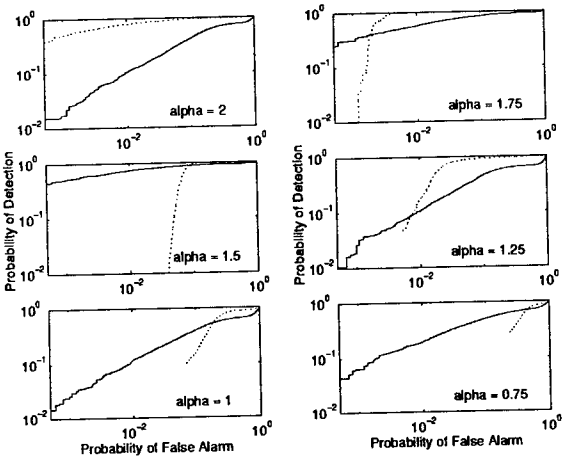


Figure 3: Comparison of the small sample performance of the Gaussian (dotted line) and the Cauchy (solid line) detector.

5. SUMMARY, CONCLUSIONS, AND FUTURE WORK

In this paper, we addressed the problem of detection of a signal, known within a multiplicative constant, in subGaussian impulsive interference of unknown underlying matrix. From this study, we found that the Gaussian detectors for the same problem deteriorate in performance when required to operate in subGaussian interference. On the other hand, a detector based on the multidimensional Cauchy distribution exhibited resistance to the presence of the subGaussian interference and high performance, comparable to the performance of the Gaussian detector in Gaussian interference. Future research in this area seems to indicate the need for evaluation of both the proposed subGaussian interference model and the corresponding detectors on real data sets. Such a process in underway and its results are expected to be announced soon.

6. REFERENCES

[1] L A Berry. Understanding Middleton's canonical formula for class A noise. *IEEE Transactions on Electromagnetic Compatibility*, EMC-23:337–344, 1981.

[2] D. Middleton. Statistical-physical models of electromagnetic interference. *IEEE Trans. Electromagnetic Compatibility*, EMC-19(3):106–127, 1977.

[3] D. Middleton. Canonical non-Gaussian noise models: Their implications for measurement and for prediction of receiver performance. *IEEE Transactions on Electromagnetic Compatibility*, EMC-21(3), 1979.

[4] D. Middleton. Procedures for determining the parameters of the first-order canonical models of class A and class B electromagnetic interference. *IEEE Trans. Electromagnetic Compatibility*, EMC-21(3):190–208, 1979.

[5] D. Middleton. Threshold detection in non-Gaussian interference environments: Exposition and interpretation of new results for EMC applications. *IEEE Transactions on Electromagnetic Compatibility*, EMC-26(1), 1984.

[6] C L Nikias and M Shao. *Signal Processing with Alpha-Stable Distributions and Applications*. John Wiley & Sons, Inc., New York, NY, 1995.

[7] G Samorodnitsky and M S Taqqu. *Stable, Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, New York, NY, 1994.

[8] M Shao and C L Nikias. Signal processing with fractional lower-order moments: Stable processes and their applications. *Proc. IEEE*, 81:986–1010, 1993.

[9] M Shao and C L Nikias. On symmetric stable models for impulsive noise. Report 231, USC-SIPI, February, 1993.

[10] G A Tsihrintzis and C L Nikias. Data-adaptive algorithms for signal detection in sub-gaussian impulsive interference. *IEEE Trans. Signal Processing*. (submitted on Jan. 20, 1996, pp. 25).

[11] G A Tsihrintzis and C L Nikias. Fast estimation of the parameters of alpha-stable impulsive interference. *IEEE Trans. Signal Processing*, June 1996. to appear.

[12] G A Tsihrintzis and C L Nikias. Performance of optimum and suboptimum receivers in the presence of impulsive noise modeled as an α–stable process. *IEEE Trans. Comm.*, COM-43:904–914, March 1995.

[13] G A Tsihrintzis and C L Nikias. Incoherent receivers in alpha-stable impulsive noise. *IEEE Trans. Signal Processing*, SP-43:2225–2229, September 1995.

[14] G A Tsihrintzis and C L Nikias. Asymptotically optimum multichannel detection of fluctuating targets in alpha-stable impulsive interference. *Signal Processing*, (submitted on July 19, 1995, pp. 20).

[15] H L Van Trees. *Detection, Estimation, and Modulation Theory, Part I.* Wiley, New York, 1968.

Plenary

Wavelets, Statistics, and Biomedical Applications

M. Unser

Wavelets, Statistics, and Biomedical Applications

Michael Unser

Biomedical Engineering and Instrumentation Program, NCRR, Bldg. 13, Room 3N17
National Institutes of Health, Bethesda, MD 20892-5766 USA
Email : unser@helix.nih.gov

SUMMARY

This paper emphasizes the statistical properties of the wavelet transform (WT) and discusses some recent examples of applications in medicine and biology.

The redundant forms of the transform (continuous wavelet transform (CWT) and wavelet frames) are well suited for detection tasks (e.g., spikes in EEG, or microcalcifications in mammograms). The CWT, in particular, can be interpreted as a prewhitening multi-scale matched filter. Redundant wavelet decompositions are also very useful for the characterization of singularities, as well as for the time-frequency analysis of non-stationary signals. We briefly discuss some examples of applications in phonocardiography, electrocardiography (ECG), and electroencephalography (EEG).

Wavelet bases (WB) provide a similar, non-redundant decomposition of a signal in terms of the shifts and dilations of a wavelet (hierarchical or pyramidal transform). There are also non-hierarchical versions that constitute a direct extension of the traditional block transforms (Fourier, DCT, etc..). This makes WB well suited for any of the tasks for which block transforms have been used traditionally: data compression, data analysis (decorrelation), and data processing (generalized filtering). Wavelets, however, may present certain advantages because they can improve the signal-to-noise ratio, while retaining a certain degree of localization in the time (or space) domain. We present three illustrative examples. The first is a straightforward denoising technique that applies a soft threshold in the wavelet domain. The second is a more refined version that uses generalized Wiener filtering; it was initially proposed for reducing noise in evoked response potentials. The third is a statistical method for detecting and locating patterns of brain activity in functional images acquired using magnetic resonance imaging (fMRI).

Finally, we conclude by describing a wavelet generalization of the classical Karhunen-Loeve transform. In particular, we provide the solution for the optimal decomposition of a wide sense stationary process (unconstrained case).

1. THREE TYPES OF WAVELET TRANSFORMS

The wavelet transform is a linear signal transformation that uses templates $\psi_{(a,b)} = a^{-1/2}\psi((x-b)/a)$, which are shifted (index b) and dilated versions (index a) of a given wavelet function $\psi(x)$ [11, 53]. The wavelet transform of the signal $f \in H$ is parameterized by the scale and shift parameters a and b; it is typically written as

$$T_\psi f(a,b) = \langle f, \psi_{(a,b)} \rangle, \qquad (1)$$

where $\langle \cdot, \cdot \rangle$ is the inner product associated with the Hilbert space H (l_2 or L_2 depending on whether the signal f is discrete or continuous). A basic requirement is that the transform is reversible, that is, that the signal f can be reconstructed from its wavelet coefficients $T_\psi f(a,b)$. The distinction between the various types of wavelet transforms depends on the way in which the scale and shift parameters are discretized.

At the most redundant end, one has the continuous wavelet transform (CWT) for which these parameters vary in a continuous fashion [20]. This representation offers the maximum freedom in the choice of the analysis wavelet. The only requirement is that the wavelet satisfies an admissibility condition; in particular, it must have zero mean.

In practice, it is often more convenient to consider the WT for some discretized values of a and b (e.g., the dyadic scales $a = 2^i$ and integer shifts $b = k$ with $(i,k) \in Z^2$). The transform will be reversible if and only if the corresponding (countable) set of templates defines a wavelet frame (WF) [10, 19, 1]. In other words, the wavelet must be designed such that

$$\forall f \in H, \quad A \cdot \|f\|^2 \le \sum_{a,b} \left| \langle f, \psi_{(a,b)} \rangle \right|^2 \le B \cdot \|f\|^2 \qquad (2)$$

where A and B are two positive constants (framebounds).

A WF is just a redundant version of a wavelet basis (WB) which can be obtained for the critical sampling rate: $a = 2^i$, $b = 2^i \cdot k$ with $(i,k) \in Z^2$. In this case, the templates must also be linearly independent, which imposes even stronger constraints on the choice of ψ. If the framebounds in (2) are such that $A = B = 1$, then the transformation is orthogonal. Such wavelets can be constructed by starting from a

0-8186-7576-4/96 $5.00

multiresolution analysis of L_2 [26, 27]. The better known examples are the Daubechies wavelets [9], which are orthogonal and compactly supported; and the Battle-Lemarié wavelets which are splines with exponential decay [24, 27]. In the case of semi- and bi-orthogonal wavelet bases [8, 49, 2], one has the following signal representation

$$f = \sum_i \sum_{k \in Z} \langle f, \tilde{\psi}_{i,k} \rangle \cdot \psi_{i,k} \quad , \tag{3}$$

with the short form convention $\psi_{i,k} = 2^{-i/2}\psi(x/2^i - k)$. The analysis wavelet $\tilde{\psi}$ is the dual of ψ (the synthesis wavelet); in the orthogonal case, both wavelets are identical.

Basic texbook references on the wavelet transform are [11, 29, 53]. For computational issues, we refer the reader to [46]. An extensive review of its various uses in medicine and biology is given in [47]; specific biomedical applications are also described in [3].

2. WAVELET ANALYSIS AND FEATURE DETECTION

The redundant forms of the transform (CWT and WF) are usually preferable for signal analyses, feature extraction, and detection tasks for they provides a description that is truly shift-invariant. Next, we discuss some wavelet properties that are of special interest for this class of applications.

A. Wavelets and time-frequency analysis

An analysis wavelet ψ is typically a well localized bandpass function with a central frequency at ω_0; a standard requirement is that its time-frequency bandwitdth product is close to the limit specified by the uncertainty principle: $\Delta_\psi \cdot \Delta_{\hat{\psi}} \geq 1/2$. Thus, each analysis template $\psi_{(a,b)}$ tends to be predominantly located in a certain elliptical region of the time-frequency plane centered at $t = b$ and $\omega = \omega_0 / a$. The area of these localization regions is the same for all templates $((a \cdot \Delta_\psi) \times (\Delta_{\hat{\psi}} / a))$ and is constrained by the uncertainty principle. Thus, by measuring the correlation between the signal and each wavelet template, we obtain a characterization of its time-frequency content (scalogram). The main difference with the short-time Fourier transform is that the size of the analysis window is not constant for it varies in inverse proportion to the frequency. This property enables the wavelet transform to zoom in on details, but at the expense of a corresponding loss in spectral resolution. In this respect, we should note that most biomedical signals of interest include a combination of impulse-like events (spikes and transients) and more diffuse oscillations (murmurs, EEG waveforms) which may all convey important information for the clinician. The short-time Fourier transform or other conventional time-frequency methods are well adapted for the latter type of events but are much less suited for the analysis of short duration pulsations. When both types of events are present in the data, the wavelet transform can offer a better compromise in terms of localization. This may explain its recent success in biomedical signal processing. Recent examples of applications where time-frequency wavelet analysis appears to be particularly appropriate are the characterization of heart beat sounds [22, 21, 31], the analysis of ECG signals including the detection of late ventricular potentials [21, 16, 28, 39] , the analysis of EEGs [38, 37, 50], as well as a variety of other physiological signals [36].

B. Wavelets as a multi-scale matched filter

In essence, the continuous wavelet transform performs a correlation analysis, so that we can expect its output to be maximum when the input signal most resembles the analysis template $\psi_{(a,b)}$. Consider the measurement model $f(x) = \varphi_a(x - \Delta x) + n(x)$ where $\varphi_a(x) = \varphi(x/a)$ is a known deterministic signal at scale a, Δx an unknown location parameter, and $n(x)$ an additive white Gaussian noise component. Classical detection theory tells us that the optimal procedure for estimating Δx is to perform the correlation with all possible shifts of our reference template and to select the position that corresponds to the maximum output (matched filter). Therefore, it makes sense to use a wavelet transform-like detector whenever the pattern φ that we are looking for can appear at various scales.

If the noise is correlated instead of white, then we can get back to the previous case by applying a whitening filter. Interestingly, the wavelet-like structure of the detector is preserved exactly if the noise has a fractional brownian motion structure. Specifically, when the noise average spectrum has the form $\phi_n(\omega) = \sigma^2 / |\omega|^\alpha$ with $\alpha = 2H+1$ where H is the Hurst exponent, we can show that the optimum detector $\psi(x)$ is proportional to the αth fractional derivative of the pattern φ that we want to detect. Consequently, for $H>0$, the optimal detector is an admissible wavelet even if the initial template $\varphi(x)$ is not (e.g. it is a lowpass function). For example, the optimal detector for finding a Gaussian in $O(\omega^{-2})$ noise is the Mexican hat wavelet (2nd derivative of a Gaussian). As suggested by Strickland, this is perhaps one of the main reasons why the wavelet transform works well for detecting microcalcifications in mammograms [7, 32, 41].

3. WAVELET BASES

Wavelet bases provide a non-redundant decomposition of a signal in terms of the shifts and dilations of ψ (hierarchical or

pyramidal transform). Hence, it is possible to represent a signal through its wavelet expansion

$$f = \sum_i \sum_{k \in Z} c_{i,k} \psi_{i,k} \qquad (4)$$

where the $c_{i,k} = \langle f, \psi_{i,k} \rangle$ are the wavelet coefficients (scale index i, and position index k). There are also non-hierarchical versions (wavelet packets, M-band perfect reconstruction filterbanks) that constitute a direct extension of the traditional block transforms (Fourier, DCT, etc..). The important point for our purpose is that, in the discrete case, the decomposition formula (4) provides a one-to-one representation of the signal in terms of its wavelet coefficients (reversible linear transformation). This makes WB well suited for any of the tasks for which block transforms have been used traditionally: data compression, data analysis (decorrelation), and data processing (generalized filtering). Wavelets, however, may present certain advantages because they can improve the signal-to-noise ratio, while retaining a certain degree of localization in the time (or space) domain.

A. Data Compression

Data compression can be achieved by quantization in the wavelet domain, or by simply
discarding certain coefficients that are insignificant. This form of orthogonal (or close-to-orthogonal) decomposition has been used effectively for image compression [25, 4, 14, 40]. Traditionally, this has been one of the primary applications of wavelets.

B. Data Processing: wavelet denoising

One of the first application of the wavelet transform in medical imaging was for noise reduction in MR images [54]. The approach proposed by Weaver et *al.* was to compute an orthogonal wavelet decomposition of the image and apply the following soft thresholding rule on the coefficients $c_{i,k} = \langle f, \psi_{i,k} \rangle$:

$$\tilde{c}_{i,k} = \begin{cases} c_{i,k} - t_i & c_{i,k} \geq t_i \\ 0 & |c_{i,k}| \leq t_i \\ c_{i,k} + t_i & c_{i,k} \leq -t_i \end{cases} \qquad (5)$$

where t_i is a threshold that depends on the noise level at the ith scale; the image is then reconstructed by the inverse wavelet transform of the $\tilde{c}_{i,k}$'s. This is essentially the wavelet shrinkage denoising method later systematized by Donoho and Johnston [18, 17], as well as DeVore and Lucier [15]. This algorithm is extremely simple to implement and works well for moderate levels of noise. Asymptotically (as the scale goes to zero and as the noise energy gets distributed over more and more sample values), it has some interesting min-max optimality properties for a relatively large class of signals [17].

The approach can easily be taken one step further by considering more general pointwise non-linear transformations $\tilde{c}_{i,k} = F(c_{i,k})$. Consider the measurement model $c_{i,k} = c^s_{i,k} + n_{i,k}$ where $c^s_{i,k}$ denotes the wavelet coefficient of the noise-free signal and $n_{i,k}$ is an independent noise component. In principle at least, one could apply the optimal Bayesian estimation rule : $\tilde{c}_{i,k} = E[c^s_{i,k} \mid c_{i,k}]$, which minimizes the mean square error. This of course requires the knowledge of the *a posteriori* probability density function $p(c^s \mid c)$, which depends on our *a priori* knowledge on $c^s_{i,k}$ ($p(c^s)$), and on the noise distribution ($p(n) = p(c \mid c^s)$). We can also constrain ourselves to the class of linear estimators, and derive the optimal linear estimate

$$\tilde{c}_{i,k} = \left(\frac{E[(c^s_{i,k})^2]}{E[(c^s_{i,k} + n_{i,k})^2]} \right) \cdot c_{i,k}, \qquad (6)$$

which has the form of a generalized Wiener filter. This particular algorithm was first proposed by Bertrand et al. for the processing of evoked response potentials (ERPs) [5]. These are very noisy signals with a strong deterministic component. Because ERPs are usually acquired using multiple trials (typ. 100-600 repetitions), the optimal weighting factors in (6) can be estimated on a coefficient-by-coefficient basis in an initial training phase, or even updated recursively. In this particular application, the wavelet transform appears to be superior to the Fourier transform, the latter being optimal only when both the signal and noise are stationary (conventional Wiener filter).

C. Data Analysis: detecting changes in fMRI

Functional neuroimaging is a fast developing area aimed at investigating the neuronal activity of the brain in vivo. The data for those studies is provided by positron emission tomography (PET), and functional magnetic resonance imaging (fMRI). PET measures the spatial distribution of certain function-specific radiotracers injected into the bloodstream prior to imaging. A typical example is the measurement of cerebral glucose utilization with the tracer [^{18}F]2-fluoro-2-deoxy-D-glucose (FDG). fMRI, which is a more recent technique, allows for a visualization of local changes in blood oxygenation believed to be induced by neuronal activation. It is substantially faster than PET and also offers better spatial resolution. Yet, there is still disagreement among specialists concerning the exact nature of the biological processes that produce the observed changes in the MR signal.

The functional images obtained with those two modalities are extremely noisy and variable, and their interpretation requires the use of statistical analysis methods [51]. What is

typically of interest is the detection of the differences of activity between different groups of subjects (e.g. normal versus diseased) or between different experimental conditions with the same subject (e.g. rest versus word generation). In either case, the variability of the signal is such that multiple subjects or repeated trials are required in each subgroup.

The first step in this analysis is to register the various images so that they can be compared on a pixel-by-pixel basis [42]. The second step is to compute the difference between the aligned group averages and perform the statistical analysis. Testing in the image domain directly is difficult because of the amount of residual noise and the necessity to use a very conservative significance level to compensate for multiple testing (one test per pixel!). A better solution is to perform the testing in the wavelet domain [35, 33, 51]. The main advantage is that the discriminative information, which is smooth and well localized spatially, becomes concentrated into a relatively small number of coefficients, while the noise remains evenly divided among all coefficients. In addition, the number of statistical tests can be reduced considerably by first identifying the few wavelet channels that contain significant differences. A recent application of this technique to fMRI is presented in [34].

4. EXTENSION OF THE KARHUNEN-LOÈVE TRANSFORM

One stage of the fast wavelet transform algorithm can be conveniently described as a multivariate filtering operation using the so-called *polyphase* representation [53]. The corresponding filterbank system is shown in Fig. 1.

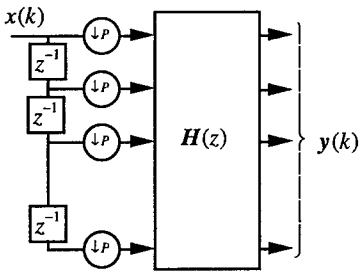


Fig. 1: Polyphase representation of a P-band wavelet analysis filterbank.

In this diagram, $x(k)$ represents the input signal and the y's are the various wavelet channels after one level of decomposition. In the standard dyadic case, there are only two channels (P=2), but the concept is also valid for larger values of P (P-band perfect reconstruction filterbank) [52, 53]. It turns out that the transformation is orthogonal if and only if the $P\times P$ transfer function matrix $\boldsymbol{H}(z)$ satisfies the paraunitary condition:

$$\boldsymbol{H}(z)\cdot\boldsymbol{H}(1/z)=\boldsymbol{I}_P, \qquad (7)$$

where $\boldsymbol{I}_P$ is the $P\times P$ identity matrix. Note that for traditional block transforms, the matrix $\boldsymbol{H}(z)$ does not depend on z (i.e., the various blocks are processed independently of each other). In order to design the optimal wavelet transform for a given class of input signals, it is therefore natural to seek the paraunitary matrix $\boldsymbol{H}(z)$ that provides the maximum energy compaction in the wavelet domain [44]. If the matrix $\boldsymbol{H}$ is constrained to be real (no delays), the solution corresponds to the classical Karhunen-Loève transform (KLT). If we allow for more general structures (for example, $\boldsymbol{H}(z)$ is an N-point FIR transfer function), we can get better results but the filter optimization subject to constraint (7) is a rather difficult task [44, 30, 6, 13]. One interesting property of the optimal solution is that the transformed components are uncorrelated; however, this is not a sufficient condition for optimality, in contrast with the standard KLT [44].

If we do not impose any order constraint on $\boldsymbol{H}(z)$, it is possible to derive the optimal solution analytically for any given wide sense stationary process with spectral power density $S_x(\omega)$. The two channel case is considered in [45]; the more general P-band case is treated in [43] using an elegant principal component formulation in the frequency domain. In each case, the solution depends on the spectral characteristics of the input signal and has the form of an ideal filter with pure "on" and "off" frequency bands. If the power spectral density is non-increasing, then the optimal solution is the ideal filterbank with P uniformly-spaced subbands. Interestingly, there are a number of wavelet transform constructions that converge asymptotically to this limit. The better known example is the family of Battle-Lemarié spline wavelets which converge to an ideal bandpass filter as the order of the spline goes to infinity [24, 48, 2]. Daubechies wavelets also exhibit similar convergence properties [23]. This partially explains why higher order wavelets usually result in smaller approximation errors.

These unconstrained solutions are primarily of interest from a theoretical point of view. For example, they can be very useful for deriving asymptotic bounds on the best performance achievable (e.g. coding gain over PCM) [12]. They are less relevant for implementation purposes because of the disadvantages of ideal filterbanks (slowly decaying impulse responses, Gibbs oscillations). This provides a good motivation for investigating more constrained solutions. As far as we know, there is not yet any general procedure for designing optimal FIR wavelets that is entirely satisfactory; this is currently an active area of research.

References

[1] A. Aldroubi, "Portraits of frames," *Proc. Amer. Math. Soc.*, vol. 123, no. 6, pp. 1661-1668, 1995.

[2] A. Aldroubi and M. Unser, "Families of multiresolution and wavelet spaces with optimal properties," *Numerical Functional Analysis and Optimization*, vol. 14, no. 5-6, pp. 417-446, 1993.

[3] A. Aldroubi and M. Unser, *Wavelets in Medicine and Biology*. Boca Raton, FL: CRC Press, 1996.

[4] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205-220, 1992.

[5] O. Bertrand, J. Bohorquez and J. Pernier, "Time-frequency digital filtering based on an invertible wavelet transform: an application to evoked potentials," *IEEE Transactions on Biomedical Engineering*, vol. 41, no. 1, pp. 77-88, 1994.

[6] H. Caglar, Y. Liu and A.N. Akansu, "Statistically optimized PR-QMF design," in Proc. *SPIE Visual Communication and Image Processing*, 1991, 86-94.

[7] L.P. Clarke, M. Kallergi, W. Qian, H.D. Li, R.A. Clark and M.L. Silbiger, "Tree-structured non-linear filter and wavelet transform for microcalcification segmentation in digital mammography," *Cancer Lett*, vol. 77, no. 2-3, pp. 173-81, 1994.

[8] A. Cohen, I. Daubechies and J.C. Feauveau, "Bi-orthogonal bases of compactly supported wavelets," *Comm. Pure Appl. Math.*, vol. 45, pp. 485-560, 1992.

[9] I. Daubechies, "Orthogonal bases of compactly supported wavelets," *Comm. Pure Appl. Math.*, vol. 41, pp. 909-996, 1988.

[10] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, no. 5, pp. 961-1005, 1990.

[11] I. Daubechies, *Ten lectures on wavelets*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.

[12] R.L. de Queiroz and H.S. Malvar, "On the asymptotic performance of hierarchical transforms," *IEEE Trans. Signal Processing*, vol. 40, no. 10, pp. 2620-2622, 1992.

[13] P. Desarte, B. Macq and D.T.M. Slock, "Signal-adapted multiresolution transform for image coding," *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 897-904, 1992.

[14] R.A. DeVore, B. Jawerth and B.J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 719-746, 1992.

[15] R.A. DeVore and B.J. Lucier, "Fast wavelet techniques for near-optimal image processing," in Proc. *IEEE Military Communications Conference*, New York, NY, 1992, 48.3.1-48.3.7.

[16] H. Dickhaus, L. Khadra and J. Brachmann, "Time-frequency analysis of ventricular late potentials," *Methods of Information in Medicine*, vol. 33, no. 2, pp. 187-95, 1994.

[17] D.L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Information Theory*, vol. 41, no. 3, pp. 613-627, 1995.

[18] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425-455, 1994.

[19] R.J. Duffin and A.C. Schaeffer, "A class of nonharmonic Fourier series," *Trans. Americ. Math. Soc.*, vol. 72, pp. 314-366, 1952.

[20] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math. Anal.*, vol. 15, no. 4, pp. 723-736, 1984.

[21] L. Khadra, H. Dickhaus and A. Lipp, "Representations of ECG-late potentials in the time frequency plane," *Journal of Medical Engineering & Technology*, vol. 17, no. 6, pp. 228-31, 1993.

[22] L. Khadra, M. Matalgah, B. El-Asir and S. Mawagdeh, "The wavelet transform and its applications to phonocardiogram signal analysis," *Medical Informatics*, vol. 16, no. 3, pp. 271-7, 1991.

[23] M.-J. Lai, "On the digital filter associated with Daubechies' wavelets," *IEEE Trans. Signal Processing*, vol. 43, no. 9, pp. 2203-2205, 1995.

[24] P.-G. Lemarié, "Ondelettes à localisation exponentielles," *J. Math. pures et appl.*, vol. 67, no. 3, pp. 227-236, 1988.

[25] A.S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 244-50, 1992.

[26] S.G. Mallat, "Multiresolution approximations and wavelet orthogonal bases of $L^2(R)$," *Trans. Amer. Math. Soc.*, vol. 315, no. 1, pp. 69-87, 1989.

[27] S.G. Mallat, "A theory of multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-11, no. 7, pp. 674-693, 1989.

[28] O. Meste, H. Rix, P. Caminal and N.V. Thakor, "Ventricular late potentials characterization in time-frequency domain by means of a wavelet transform," *IEEE Transactions on Biomedical Engineering*, vol. 41, no. 7, pp. 625-34, 1994.

[29] Y. Meyer, *Ondelettes et Opérateurs I : Ondelettes*. Paris, France: Hermann, 1990.

[30] P. Moulin, "A new look at signal-adapted QMF bank design," in Proc. *IEEE Int. Conference on Acoustics, Speech, and Signal Processing*, Detroit, MI, May 9-12, 1995, vol. II, pp. 1312-1315.

[31] M.S. Obaidat, "Phonocardiogram signal analysis: techniques and performance," *Journal of Medical Engineering & Technology*, vol. 17, pp. 221-227, 1993.

[32] W. Qian, L.P. Clarke, H.D. Li, R. Clark and M.L. Silbiger, "Digital mammography: m-channel quadrature mirror filters (QMFs) for microcalcification extraction," *Computerized Medical Imaging and Graphics*, vol. 18, no. 5, pp. 301-314, 1994.

[33] D.E. Rio, R.R. Rawlings, U.E. Ruttimann and R. Momenan, "Study of statistical methods applied in the

spatial, wavelet and Fourier domain to enhance and analyze group characteristics of images: application to positron emission tomography brain images," in Proc. *SPIE Conf Mathematical Methods in Medical Imaging III*, 1994, vol. 2299, pp. 194-206.

[34] U.E. Ruttimann, N.F. Ramsey, D.W. Hommer, P. Thévenaz, C. Lee and M. Unser, "Analysis of functional magnetic resonance images by wavelet decomposition," in Proc. *Int. Conf. on Image Processing*, Washington, DC, 1995, vol. I, pp. 633-636.

[35] U.E. Ruttimann, M. Unser, D. Rio and R.R. Rawlings, "Use of the wavelet transform to investigate differences in brain PET images between patients," in Proc. *SPIE Conf Mathematical Methods in Medical Imaging II*, San Diego, CA, July 11-16, 1993, vol. 2035, pp. 192-203.

[36] R. Sartene, J. Wallet, P. Allione, S. Palka, L. Poupard and J.L. Bernard, "Using wavelet transform to analyse cardiorespiratory and electroencephalographic signals during sleep," in Proc. *IEEE EMBS Workshop on wavelet in Medicine and Biology*, Baltimore, 1994, 18a-19a.

[37] S.J. Schiff, A. Aldroubi, M. Unser and S. Sato, "Fast wavelet transformation of EEG," *Electroencephalogr Clin Neurophysiol*, vol. 91, no. 6, pp. 442-55, 1994.

[38] S.J. Schiff, J. Milton, J. Heller and S.L. Weinstein, "Wavelet transforms and surrogate data for electroencephalographic and seizure localization," *Optical Engineering*, vol. 33, no. 7, pp. 2162-9, 1994.

[39] L. Senhadji, G. Carrault, J.J. Bellanger and G. Passariello, "Comparing wavelet transforms for recognizing cardiac patterns," *IEEE Engineering in Medicine and Biology Magazine*, vol. 14, no. 2, pp. 167-173, 1995.

[40] Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 41, no. 12, pp. 3445-3462, 1993.

[41] R.N. Strickland and H.I. Hahn, "Detection of microcalcifications in mammograms using wavelets," in Proc. *SPIE Conf Wavelet Applications in Signal and Image Processing II*, San Diego, CA, July 11-16, 1994, vol. 2303, pp. 430-441.

[42] P. Thévenaz and M. Unser, "Iterative multi-scale image registration without landmarks," in Proc. *Int. Conf. on Image Processing*, Washington, DC, 1995, vol. III, pp. 228-231.

[43] M.K. Tsatsanis and G.B. Giannakis, "Principal component filter banks for optimal wavelet analysis," *IEEE Trans. Signal Processing*, vol. 43, no. 8, pp. 1766-1777, 1995.

[44] M. Unser, "An extension of the Karhunen-Loève transform for wavelets and perfect reconstruction filterbanks," in Proc. *SPIE Mathematical Imaging : Wavelet Applications in Signal and Image Processing*, San Diego, CA, 1993, vol. SPIE Proc. 2034, pp. 45-56.

[45] M. Unser, "On the optimality of ideal filters for pyramid and wavelet signal approximation," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3591-3596, 1993.

[46] M. Unser, "A practical guide to the implementation of the wavelet transform," in *Wavelets in Medicine and Biology*, A. Aldroubi and M. Unser, Eds., Boca Raton, FL: CRC Press, pp. 40-78, 1996.

[47] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proceedings of the IEEE*, vol. 84, no. 4, 1996.

[48] M. Unser, A. Aldroubi and M. Eden, "Polynomial spline signal approximations : filter design and asymptotic equivalence with Shannon's sampling theorem," *IEEE Trans. Information Theory*, vol. 38, no. 1, pp. 95-103, 1992.

[49] M. Unser, A. Aldroubi and M. Eden, "A family of polynomial spline wavelet transforms," *Signal Processing*, vol. 30, no. 2, pp. 141-162, 1993.

[50] M. Unser, A. Aldroubi and S.J. Schiff, "Fast implementation of the continuous wavelet transform with integer scales," *IEEE Trans. Signal Processing*, vol. 42, no. 12, pp. 3519-3523, 1994.

[51] M. Unser, P. Thévenaz, C. Lee and U.E. Ruttimann, "Registration and statistical analysis of PET images using the wavelet transform," *IEEE Engineering in Medicine and Biology*, vol. 14, no. 5, pp. 603-611, 1995.

[52] P.P. Vaidyanathan, "Quadrature mirror filter banks, M-band extensions and perfect-reconstruction technique," *IEEE ASSP Mag.*, vol. 4, pp. 4-20, 1987.

[53] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice Hall, 1995.

[54] J.B. Weaver, X. Yansun, D.M. Healy, Jr. and L.D. Cromwell, "Filtering noise from images with wavelet transforms," *Magnetic Resonance in Medicine*, vol. 21, no. 2, pp. 288-95, 1991.

TA-1: Applications

Bounds for Estimation of Particle's Velocity from Laser Measurements

Olivier Besson and Frédéric Galtier

ENSICA, Department of Avionics and Systems
Place Emile Blouin, 31056 Toulouse, France
besson,galtier@ensica.fr

Abstract

The problem of estimating the speed of a particle of air passing through the region of interference fringes generated by two coherent laser beams is addressed. The signal detected being of the form $A.\exp\left\{-2\alpha^2.f_d^2t^2\right\}\cos(2\pi f_d t)$ *-where the Doppler frequency* f_d *is related to particle's velocity- this paper is concerned with the best accurate estimation of the parameters* A *and* f_d *in the model considered. Cramér-Rao bounds on the accuracy of estimates of* A *and* f_d *are derived and closed-formed expressions are given. Approximated formulas provide quantitative insights into the influence of* α *and* f_d*. Additionally, a Maximum Likelihood Estimator is presented. Numerical examples illustrate the performance of the MLE and compare it with the CRB. The influence of the SNR, the sample size, the optical parameter* α *and the frequency* f_d *on the estimation performance are emphasized.*

1. Problem statement

Laser velocimeters have gained popularity in the fluid mechanics application, where they have been used to estimate particles velocity in a flow[1][2], mainly for measurements in wind tunnels. Since this system provides a non-intrusive and reliable way of measuring local velocities in fluid flows, it has become an interesting alternative to mechanical systems, for instance in situations where one does not want to disturb the flow. Furthermore, these systems have been reported to yield precise estimation. Examples of laser velocimetry applications include analysis of flow surrounding the blade tips of a hovering rotor, measurements of mean velocity and turbulence intensity in unsteady ultrasonic flow. In aeronautics applications, there is a vital need in having a reliable aircraft's speed measurement system. Moreover, this system must fulfil severe constraints regarding size, weight, accuracy and robustness. With the emergence of a new generation of cheap and small laser diodes, laser anemometers become a conceivable and promising technique for onboard measurement of aircraft's speed. The principle of such a system is now briefly described. Two coherent laser beams are crossed and focused in the vicinity of the aircraft. They generate a symmetric ellipsoidal probe volume composed of equidistant bright and dark fringes. As a particle of air passes through this region, the fringes will cause it alternatively to scatter and not to scatter light, according to the particle's velocity and interfringe width. More exactly, the signal received by the photodetector can be shown[1] to have the form

$$\begin{aligned} x(t) &= A.e^{-2\left(\frac{V}{W}\right)^2 t^2}.\cos(2\pi\frac{V}{I}t) + w(t) \\ &= A.s(t) + w(t) \qquad t = 0, \pm 1, ..., \pm T \quad (1) \end{aligned}$$

where V represents the particle's velocity, $2W$ is the total length of the interference fringes, I denotes the interfringe width. The amplitude A depends on the particle's size, the power emitted and optical transmission coefficients. In (1) the additive noise $\{w(t)\}$ is assumed to be a sequence of i.i.d. Gaussian variables with variance σ_w^2. It should be pointed out that the Gaussian shape of the time-varying amplitude $e^{-2\left(\frac{V}{W}\right)^2 t^2}$ is directly induced from the Gaussian shape of the intensity distribution within the laser beam. In what follows, we note $\alpha \overset{def}{=} \frac{I}{W}$ and let $f_d \overset{def}{=} \frac{V}{I}$ denote the "Doppler" frequency, so that $s(t)$ in (1) can be rewritten as $s(t) = e^{-2\alpha^2.f_d^2.t^2}.\cos(2\pi f_d t)$. Here, we are concerned with the best accuracy that can be achieved when estimating the parameters A and f_d in (1). It should be noted that the model studied here belongs to the class of amplitude modulated sinusoidal signals (see [3] for a thorough overview of multiplicative models). However, in contrast with most approaches, the time-varying amplitude cannot be viewed just as a perturbation term. Moreover, the amplitude and phase are not decoupled from one another, as they both carry information about the frequency of interest.

2. Cramér-Rao Bounds

Let $\boldsymbol{\theta} = \left[A, f_d, \sigma_w^2\right]^T$ be the parameter vector to be estimated from the measurements $\{x(t)\}_{t=-T...T}$. For later use, we define

0-8186-7576-4/96 $5.00

$\mathbf{s} = [s(-T), ..., s(0), ..., s(T)]^T$ and
$\mathbf{w} = [w(-T), ..., w(0), ..., w(T)]^T$ so that (1) can be written in the following compact form

$$\mathbf{x} = A\mathbf{s} + \mathbf{w} \tag{2}$$

2.1. Exact CRB

Under the white Gaussian assumption for $w(t)$, the log-likelihood function is given by[4]

$$\Lambda(\mathbf{x}, \boldsymbol{\theta}) = cte - \frac{2T+1}{2} \ln \sigma_w^2 - \frac{1}{2\sigma_w^2} \|\mathbf{x} - A\mathbf{s}\|^2 \tag{3}$$

Twice differentiating (3) wrt $\boldsymbol{\theta}$ and taking expectations, it is straightforward to show (see [5] for details) that the Fisher Information Matrix (FIM) is given by

$$\mathbf{F} = \begin{bmatrix} \frac{1}{\sigma_w^2}.\mathbf{s}^T\mathbf{s} & \frac{A}{\sigma_w^2}.\mathbf{s}^T\mathbf{s}'_f & 0 \\ \frac{A}{\sigma_w^2}.\mathbf{s}^T\mathbf{s}'_f & \frac{A^2}{\sigma_w^2}.\mathbf{s}'^T_f\mathbf{s}'_f & 0 \\ 0 & 0 & \frac{2T+1}{2\sigma_w^4} \end{bmatrix} \tag{4}$$

where $\mathbf{s}'_f = \frac{\partial \mathbf{s}}{\partial f_d}$. Inverting (4), the diagonal terms of the Cramér-Rao Bound (CRB) are obtained as

$$CRB(\sigma_w^2) = \frac{2\sigma_w^4}{2T+1} \tag{5}$$

$$CRB(A) = \sigma_w^2 . \frac{\mathbf{s}'^T_f \mathbf{s}'_f}{(\mathbf{s}^T\mathbf{s})\left(\mathbf{s}'^T_f\mathbf{s}'_f\right) - \left(\mathbf{s}^T\mathbf{s}'_f\right)\left(\mathbf{s}^T\mathbf{s}'_f\right)} \tag{6}$$

$$CRB(f_d) = \frac{\sigma_w^2}{A^2} . \frac{\mathbf{s}^T\mathbf{s}}{(\mathbf{s}^T\mathbf{s})\left(\mathbf{s}'^T_f\mathbf{s}'_f\right) - \left(\mathbf{s}^T\mathbf{s}'_f\right)\left(\mathbf{s}^T\mathbf{s}'_f\right)} \tag{7}$$

which provides closed-form expressions of the CRB. The influences of α and f_d are of interest as they can guide the selection of the sampling frequency and the optical parameters. However, it turns out that an analytical study from (7) of the dependence of $CRB(f_d)$ on α and f_d is intractable, the derivatives $\frac{\partial CRB(f_d)}{\partial \alpha}$, $\frac{\partial CRB(f_d)}{\partial f_d}$ being difficult to interpret. This influence will therefore be evaluated numerically. However, further insights into the analysis of the CRB can be gained by considering the large-sample case and approximated formulas for the FIM, as shown in the next section.

2.2. Approximated CRB

The aim of this section is to get simplified expressions for the CRB which could provide direct relations between $CRB(f_d)$ and the parameters $\alpha, f_d, A, \sigma_w^2$. We consider that T is "large" (as $\exp\left\{-2\alpha^2 f_d^2 t^2\right\}$ decays very quickly, this assumption is not restrictive). First, observe that the asymptotic FIM depends on the quantities $\lim_{T\to\infty} \mathbf{s}^T\mathbf{s}$, $\lim_{T\to\infty} \mathbf{s}^T\mathbf{s}'_f$ and $\lim_{T\to\infty} \mathbf{s}'^T_f\mathbf{s}'_f$. Therefore, it will consist of a combination of terms of the forms $\lim_{T\to\infty} \sum_{t=-T}^{T} \exp\left\{-4\alpha^2 f_d^2 t^2\right\} .t^n . \begin{matrix}\cos \\ \sin\end{matrix} (4\pi f_d t)$. To get further insights into their values, we propose to use the following approximation:

$$\lim_{T\to\infty} \sum_{t=-T}^{T} \exp\left\{-4\alpha^2 f_d^2 t^2\right\} .t^n . \begin{matrix}\cos \\ \sin\end{matrix} (4\pi f_d t)$$
$$\simeq \int_{-\infty}^{\infty} \exp\left\{-4\alpha^2 f_d^2 t^2\right\} .t^n . \begin{matrix}\cos \\ \sin\end{matrix} (4\pi f_d t).dt \tag{8}$$

This corresponds to a rectangular approximation of the integral. More intuitively, consider either a signal $x(t) = e^{-4\alpha^2 f_d^2 \tau^2}.t^n$ or a random process with autocorrelation $r(\tau) = e^{-4\alpha^2 f_d^2 \tau^2}.\tau^n$. The left-hand (resp. right-hand) sides of (8) are the real and imaginary parts of the Discrete Time (resp. Continuous Time) Fourier Transform of these sequences, evaluated at $2f_d$. Hence, (8) amounts to say that the DTFT fairly approximates the CTFT, which is a common hypothesis. However, it is only an approximation and the ensuing expressions are not exact. Nevertheless, as will be illustrated by numerical examples, it is a very accurate approximation. Based on (8), it can be shown[5] that

$$\lim_{T\to\infty} \mathbf{s}^T\mathbf{s} \simeq \frac{\sqrt{\pi}}{4\alpha f_d} . \left[1 + \exp\left\{-\frac{\pi^2}{\alpha^2}\right\}\right] \tag{9}$$

$$\lim_{T\to\infty} \mathbf{s}^T\mathbf{s}'_f \simeq -\frac{\sqrt{\pi}}{8\alpha f_d^2} . \left[1 + \exp\left\{-\frac{\pi^2}{\alpha^2}\right\}\right] \tag{10}$$

$$\lim_{T\to\infty} \mathbf{s}'^T_f\mathbf{s}'_f \simeq \frac{\sqrt{\pi}\left(3\alpha^2 + 2\pi^2\right)}{16\alpha^3 f_d^3} + \frac{\sqrt{\pi}\left(3\alpha^2 - 2\pi^2\right)}{16\alpha^3 f_d^3} \exp\left\{-\frac{\pi^2}{\alpha^2}\right\} \tag{11}$$

Furthermore, we note that, in general, α (the inverse of the number of interference fringes) is small which implies that terms of the form $\beta . \exp\left\{-\frac{\pi^2}{\alpha^2}\right\}$, $\gamma . \exp\left\{-2\frac{\pi^2}{\alpha^2}\right\}$ are negligible compared to 1. By neglecting these terms in (9)-(11) the (approximated) asymptotic FIM corresponding to $[A, f_d]^T$ is given by

$$\lim_{T\to\infty} \mathbf{F}_{[A,f_d]} \simeq \frac{1}{\sigma_w^2} \begin{bmatrix} \frac{\sqrt{\pi}}{4\alpha f_d} & \frac{-A\sqrt{\pi}}{8\alpha f_d^2} \\ \frac{-A\sqrt{\pi}}{8\alpha f_d^2} & \frac{A^2\sqrt{\pi}\left(3\alpha^2+2\pi^2\right)}{16\alpha^3 f_d^3} \end{bmatrix} \tag{12}$$

With this simplification, inverting (12) and rearranging terms, it comes:

$$CRB(A) \simeq \sigma_w^2 . \frac{2.\left(3\alpha^2 + 2\pi^2\right).\alpha.f_d}{\sqrt{\pi}.\left(\alpha^2 + \pi^2\right)} \tag{13}$$

$$CRB(f_d) \simeq \frac{\sigma_w^2}{A^2} . \frac{8.\alpha^3.f_d^3}{\sqrt{\pi}.\left(\alpha^2 + \pi^2\right)} \tag{14}$$

The approximated formulas (13) and (14) are believed to be of interest from both a theoretical and practical point of view. *It will be shown in the section 4 that they provide very accurate approximations of (6) and (7).* These formulas give a direct expression of the CRB as a function of α (which is a design parameter) and f_d. It should be pointed out first that the CRB tends to a constant as the sample size goes to infinity. This is in contrast with most estimation problems where the Cramér-Rao bound is usually of order $O(1/T)$. Therefore, for T above a threshold, it can be expected that the CRB will not decrease (hence no improvement is achieved): this will be illustrated in section 4. Note also that $CRB(f_d)$ is roughly proportional to $\alpha^3.f_d^3$ whereas $CRB(A)$ is proportional to $\alpha.f_d$. Therefore, increasing the probe volume by a factor of 10 could possibly result in a gain of 1000 on the variance of the frequency estimate. Additionally, observe that

$$\frac{\partial CRB(f_d)}{\partial \alpha} = \frac{8.\sigma_w^2.f_d^3}{\sqrt{\pi}.A^2}\frac{\alpha^2.\left(\alpha^2+3\pi^2\right)}{\left(\alpha^2+\pi^2\right)^2} > 0 \qquad (15)$$

$$\frac{\partial CRB(f_d)}{\partial f_d} = \frac{\sigma_w^2}{A^2}.\frac{24.\alpha^3.f_d^2}{\sqrt{\pi}.\left(\alpha^2+\pi^2\right)} > 0 \qquad (16)$$

which implies that $CRB(f_d)$ monotonically increases with α and f_d.

3. Maximum Likelihood Estimation

In this section, we derive maximum likelihood estimators of A and f_d in the model (1). For any given value of f_d, $\Lambda(\mathbf{x},\boldsymbol{\theta})$ in (3) being a quadratic function of the parameter A, the minimization w.r.t. A reduces to a simple least-squares problem and leads to

$$\hat{A} = \frac{\mathbf{s}^T\mathbf{x}}{\mathbf{s}^T\mathbf{s}} \qquad (17)$$

$\hat{A}$ will be the MLE of A if f_d is replaced by its ML estimate in (17). Reporting (17) into (3), the ML estimate of f_d is then found to be the solution of the following minimization problem:

$$\hat{f}_d = \arg\min_f J_1(f) \qquad (18)$$

$$J_1(f) = \left\|\mathbf{x} - \frac{\mathbf{s}^T\mathbf{x}}{\mathbf{s}^T\mathbf{s}}.\mathbf{s}\right\|^2 \stackrel{def}{=} \|\boldsymbol{\varepsilon}\|^2 \qquad (19)$$

Since $J_1(f)$ is a non-linear function of f, no analytical solution for the problem exists and one has to resort to numerical methods[6]. In an attempt to provide a computationally efficient algorithm, we propose to use a Gauss-Newton procedure which uses exact first-order derivatives and approximated second-order derivatives. Therefore, only $J_1'(f) = \frac{dJ_1(f)}{df}$ needs to be computed. The derivative $\boldsymbol{\epsilon}' = \frac{\partial \boldsymbol{\epsilon}}{\partial f}$ can be written as

$$\begin{aligned}\boldsymbol{\epsilon}' &= \frac{-1}{\left(\mathbf{s}^T\mathbf{s}\right)^2}.\left\{\left[\left(\mathbf{x}^T\mathbf{s}'\right)\left(\mathbf{s}^T\mathbf{s}\right) - 2\left(\mathbf{x}^T\mathbf{s}\right)\left(\mathbf{s}^T\mathbf{s}'\right)\right]\mathbf{s}\right. \\ &\quad \left. + \left(\mathbf{x}^T\mathbf{s}\right)\left(\mathbf{s}^T\mathbf{s}\right)\mathbf{s}'\right\} \qquad (20)\end{aligned}$$

Hence

$$\begin{aligned}J_1'(f) &= 2\boldsymbol{\epsilon}'^T\boldsymbol{\epsilon} \\ &= 2\frac{\mathbf{x}^T\mathbf{s}}{\left(\mathbf{s}^T\mathbf{s}\right)^2}.\left\{\left(\mathbf{x}^T\mathbf{s}\right)\left(\mathbf{s}^T\mathbf{s}'\right) - \left(\mathbf{x}^T\mathbf{s}'\right)\left(\mathbf{s}^T\mathbf{s}\right)\right\} \qquad (21)\end{aligned}$$

The Gauss-Newton makes use of the following approximation for the Hessian

$$J_1''(f) \approx 2\boldsymbol{\epsilon}'^T\boldsymbol{\epsilon}' \qquad (22)$$

The frequency is thus estimated in an iterative way

$$\hat{f}^{(n+1)} = \hat{f}^{(n)} - \left[\left[\boldsymbol{\epsilon}'^T\boldsymbol{\epsilon}'\right]^{-1}.\boldsymbol{\epsilon}'^T\boldsymbol{\epsilon}\right]_{\hat{f}^{(n)}} \qquad (23)$$

The iterations are stopped whenever $\left|\hat{f}^{(n+1)} - \hat{f}^{(n)}\right| \leq \delta\left|\hat{f}^{(n)}\right|$ where δ is a user defined parameter. In order to avoid possible convergence towards a local minima, care is to be taken in order to properly initialize the algorithm. In the simulations presented in the next section, a Fast Fourier Transform of the data followed by a coarse search for the maximum is used.

4. Numerical examples and conclusions

In this section, we present some numerical examples in which we compare the CRB derived in Section 2 with the performance of the MLE. Since the FFT-based estimate is also available as the initial step of the MLE and because it is the most intuitive way for spectral estimation, we will also compare its performance with the CRB. Additionally, we provide a comparison between exact and approximated CRB and we illustrate the influence of various parameters on the estimation performance. We concentrate on the estimation of the frequency f_d which directly provides particle's velocity. The value of α is selected as $\alpha = 0.122857$ and $A = 1$ thorough the simulations. The Signal to Noise Ratio (SNR) is defined as $SNR = \frac{A^2}{2\sigma_w^2}$. First, we study the influence of the number of samples T. Figures 1 and 2 show the MSE of the estimates versus T for different values of f_d and with $SNR = 15dB$. Exact CRB (given by (7)) are shown in solid lines whereas approximated CRB (see (14)) appear in dashed-dotted lines. From these figures, it can be seen that the MLE has a performance very close to the CRB and superior to the

FFT estimator. The approximate formula (14) gives a very accurate approximation of the exact CRB, as long as T is large enough (which is an expected result since the approximated formula is "asymptotic" in T). However, the number of samples needed for the two expressions to be equal is reasonable (this number decreases while f_d increases). Note also that, when T increases above a threshold (typically $T > 1/\alpha f_d$), no improvement is achieved. This is due to the fact that, for large t, $e^{-2\alpha^2 f_d^2 t^2} \approx 0$: hence, the signal essentially contains noise.

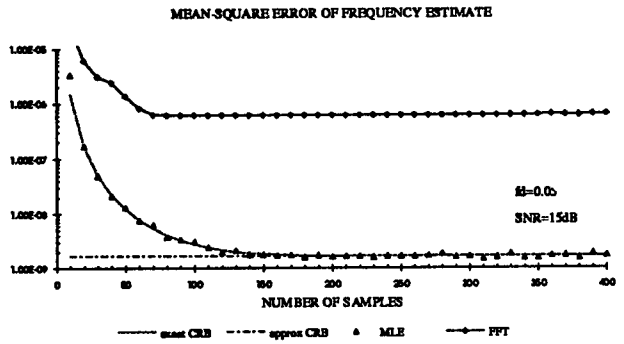


Fig.1. CRB and performance of MLE and FFT estimators versus T. $SNR = 15dB$. $f_d = 0.05$. $A = 1$.

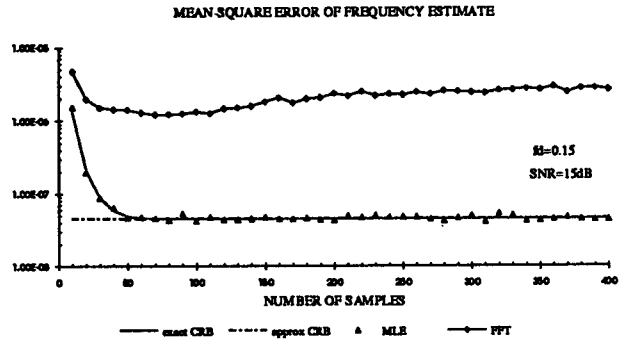


Fig. 2. CRB and performance of MLE and FFT estimators versus T. $SNR = 15dB$. $f_d = 0.15$. $A = 1$.

We now investigate the dependence of $CRB(f_d)$ on f_d and α in Figures 3,4. As can be seen, the CRB increases with f_d or α, which was expected from (15),(16). As f_d (or α) increases, the bandwidth of the time-varying amplitude $\exp\{-2\alpha^2 f_d^2 t^2\}$ increases, which in turns complicates the frequency estimation. Finally, note that $CRB(f_d, \alpha > 0) > CRB(f_d, \alpha = 0)$, this latter case corresponding to the constant amplitude sinusoidal signal. Therefore, although information about f_d is contained in both the amplitude and the phase of $s(t)$, this does not improve the estimation compared with the constant amplitude case where the amplitude does not bring information about the Doppler frequency.

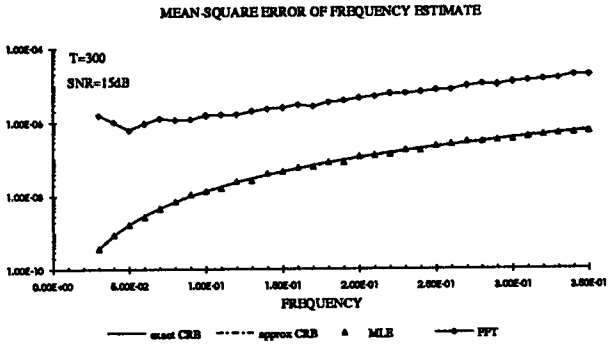


Fig. 3. CRB and performance of MLE and FFT estimators versus f_d. $T = 300$. $SNR = 15dB$. $A = 1$.

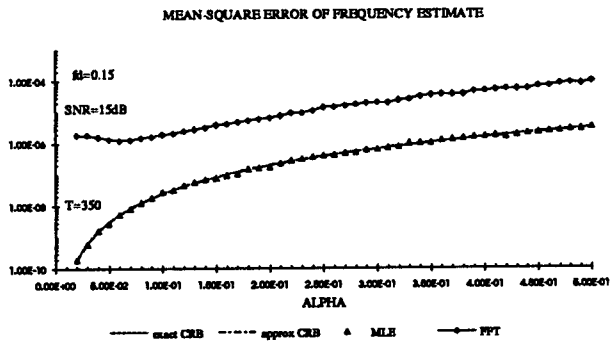


Fig. 4. CRB and performance of MLE and FFT estimators versus α. $f_d = 0.15$. $T = 350$. $SNR = 15dB$. $A = 1$.

References

[1] M. Riethmuller, "Laser doppler velocimetry", in *Measurement of unsteady fluid dynamic phenomena* (B.E.Richards, Editor), Mc Graw Hill, 1977

[2] *Laser Velocimetry,* in Lecture Series of the von Karman Institute for fluid dynamics, September 1981

[3] G. Zhou, *Random amplitude and polynomial phase modeling of nonstationary processes using higher-order and cyclic statistics,* PhD Dissertation, University of Virginia, January 1995

[4] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory,* Prentice Hall, 1993

[5] O. Besson and F. Galtier, "Estimating particle's velocity from laser measurements: Maximum Likelihood and Cramér-Rao Bounds", submitted

[6] G.A.F. Seber and C.J. Wild, *Nonlinear Regression,* John Wiley & Sons, 1989

SEQUENTIAL TEST AND PARAMETER ESTIMATION FOR ARRAY PROCESSING OF SEISMIC DATA

Dmitri V. Sidorovich, Christoph F. Mecklenbräuker, and Johann F. Böhme

Department of Electrical Engineering, Ruhr University,
D-44780 Bochum, Germany

ABSTRACT

We present a sequential test and parameter estimation technique for measured seismic data from the GERESS array situated in Germany. A new approximation of the test statistic distribution and the test threshold is proposed. The sequentially rejecting Bonferroni-Holm test guarantees a global test level. This allows to avoid the computationally expensive bootstrap method and leads to a more simple algorithm. Approximate conditional maximum likelihood estimates in the frequency domain are used to overcome the resolution limits of conventional methods for wideband signal processing and to construct the sequential test. The combination of global optimization via genetic algorithm and a local one using the scoring seems to be a good compromise to handle the problem of cumbersome maximization of the log-likelihood function over the parameters of interest. The algorithm for testing the number of signal phases is applied simultaneously with the estimation of the model parameters.

1. INTRODUCTION

Earthquakes and regional evens give rise to a number of different types of waves, e.g. pressure waves, shear waves, or surface waves. The first waves to be observed on a seismogram are compressional P-waves ("primary" waves). S-waves ("secondary") are transverse (shear) waves with the particle motion in the plane perpendicular to the direction of propagation. The characteristics of the various possible types of seismic waves in a "perfect" medium are well known [1]. The difference in polarization of seismic waves can be used for the analysis with an array containing 3-component sensors [7]. In this contribution we analyse the outputs of an array of vertically sensitive seismometers in order to detect and separate different phases of a regional seismic event. General heterogeneities existing along the travel path and underneath the array reduce the signal coherence and produce travel-time residues between the stations of the array. In this scenario there are more than one phase impinging on the array within a short observation interval. Consequently, approximate conditional maximum likelihood estimates (ACMLE) in the frequency domain [3] can be used to resolve different phases of a seismic event. We combine the global optimization by means of a genetic algorithm and a local one using scoring in order to handle the problem of maximization of the log-likelihood function over parameters of interest and not to be too computationally expensive. Simultaneous usage of the model parameter estimation and testing algorithm with the bootstrap approximation is investigated in [6] and [4]. General results in the seismic application of detection algorithms for narrowband signals can be found in [8]. The sequentially rejecting Bonferroni-Holm test has been formulated in [5]. We present an appropriate test statistic and a method of wideband signal testing.

2. DATA MODEL AND WAVE PARAMETER ESTIMATION

We assume that $m = 1, \ldots, M$ different types of waves arrive at the array. The positions of the sensors of the nth station ($n = 1, \ldots, N$) can be described by a vector $\underline{r}_n$. The outputs of the sensors are Fourier–transformed with a rectangular window of length T:

$$\underline{X}(\omega) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} \underline{x}(t) e^{-j\omega t}. \qquad (1)$$

The reception–propagation situation is described by a $(N \times M)$ matrix $\mathbf{H}(\omega) = [\underline{d}_1, \ldots, \underline{d}_M]'$.
The vectors $\underline{d}_i = [e^{-j\underline{k}_i' \underline{r}_1}, \ldots, e^{-j\underline{k}_i' \underline{r}_N}]'$ are the phase vectors where

$$\underline{k}_i = \frac{\omega}{V_i} \cdot [\cos\phi_i \cos\alpha_i, \cos\phi_i \sin\alpha_i, \sin\phi_i]'$$

is the wavenumber vector of a wave at frequency ω with velocity V_i, and seen at the origin of the array at azimuth α_i and elevation ϕ_i. Since only vertically

0-8186-7576-4/96 $5.00

sensitive seismometers of a plain array are involved in the analysis, we have $\phi_i = 0$ in this application. The wavenumber vectors may be written as $\underline{k}_i = \omega\underline{\xi}_i$ $(i = 1, \ldots, M)$ where $\underline{\xi}_i$ is the slowness vector. The parameters of interest are M slowness vectors $\underline{\xi}_i$, i.e. $\underline{\eta} = (\underline{\xi}_1', \ldots, \underline{\xi}_M')'$, the spectral parameters of the sources, and spectral parameters of noise. Approximately one can express the sensor output vector for each discrete frequency ω_k by

$$\underline{X}(\omega_k) = \mathbf{H}(\omega_k)\underline{S}(\omega_k) + \underline{U}(\omega_k), \qquad (2)$$

where $\underline{S}(\omega_k)$ and $\underline{U}(\omega_k)$ are the Fourier transform of the signals and noise. We assume that $\mathbf{H}(\omega)$ and spectral power of noise $\nu(\omega)$ change slowly with ω, and frequencies $\omega^i (i = 1, \ldots, P)$ suffice to describe the behavior of $\mathbf{H}(\omega)$ and $\nu(\omega)$. If noise is stationary, the $\underline{X}(\omega_k)$ given $\underline{S}(\omega_k)$ is approximately complex normal with mean $\mathbf{H}(\omega^i)\underline{S}(\omega_k)$ and covariance matrix $\nu(\omega^i)\mathbf{I}$. Wideband ACMLE's maximize

$$L(\underline{\eta}, \underline{S}, \underline{\nu}) = -\sum_{i=1}^{P}[pN\log\nu(\omega^i) + \frac{1}{\nu(\omega^i)}\sum_k(\underline{X}(\omega_k) - \mathbf{H}(\omega^i)\underline{S}(\omega_k))^*(\underline{X}(\omega_k) - \mathbf{H}(\omega^i)\underline{S}(\omega_k))],$$

where the inner sum is over the p discrete frequencies ω_k around ω^i. Maximization of the function $L(\underline{\eta}, \underline{S}, \underline{\nu})$ over the $\underline{S}(\omega_k)$ and $\nu(\omega^i)$ leads to the explicit solutions $\tilde{\underline{S}}(\omega_k)$ and $\tilde{\nu}(\omega^i)$ [4]. Then, we can find ACMLE's $\breve{\underline{\eta}}$ by minimizing of

$$q(\underline{\eta}) = \frac{1}{P}\sum_{i=1}^{P}\log\ \mathrm{tr}[(\mathbf{I} - \mathbf{P}^i(\underline{\eta}))\hat{\mathbf{C}}^i_{\underline{X}}]. \qquad (3)$$

$\mathbf{P}^i$ is the projector onto the signal space of M signals $\mathbf{P}^i = \mathbf{H}(\omega^i)[\mathbf{H}^*(\omega^i)\mathbf{H}(\omega^i)]^{-1}\mathbf{H}^*(\omega^i)$.

$$\hat{\mathbf{C}}^i_{\underline{X}} = \hat{\mathbf{C}}_{\underline{X}}(\omega^i) = \frac{1}{p}\sum_k \underline{X}(\omega_k)\underline{X}(\omega_k)^* \qquad (4)$$

denotes a non-parametric estimate of the spectral density matrix (SDM). We smooth here over p discrete frequencies ω_k around ω^i.

3. SEQUENTIAL DETECTION

The main idea of the sequential test is to detect the strongest signal, extract it and continue such procedure until the hypothesis that there exists no further signal will be accepted. Approximate conditional maximum likelihood ratio tests (ACMLRT) for the hypotheses that there is no $m+1$st signal if m signals are already detected, in analogy to [4], results in a test statistic

$$t_{m+1}(\underline{X}) = 2pPN\,[-q_{m+1}(\breve{\underline{\eta}}_{m+1}) + q_m(\breve{\underline{\eta}}_m)]. \qquad (5)$$

$\underline{\eta}_{m+1} = (\underline{\eta}_m', \underline{\xi}_{m+1}')'$ is the vector of wave parameter for $m+1$ signals, q_{m+1} and q_m are defined as in (3) for $m+1$ and m signals, respectively. The assumption that

$$q_{m+1}(\breve{\underline{\eta}}_{m+1}) = \inf_{\underline{\eta}_{m+1}} q_{m+1}(\underline{\eta}_{m+1}) \approx \inf_{\underline{\xi}_{m+1}} q_{m+1}(\breve{\underline{\eta}}_m, \underline{\xi}_{m+1})$$

reduces the numerical effort for sequential ACMLRT. Thus, the hypothesis that there is no $m+1$st signal is rejecteded if

$$Q_{m+1}(\breve{\underline{\eta}}_m, \bar{\underline{\xi}}_{m+1}) = \max_{\underline{\xi}_{m+1}} Q_{m+1}(\breve{\underline{\eta}}_m, \underline{\xi}_{m+1}) > \kappa_{m+1} \qquad (6)$$

with $Q_{m+1}(\underline{\eta}_m, \underline{\xi}_{m+1}) = q_m(\underline{\eta}_m) - q_{m+1}(\underline{\eta}_m, \underline{\xi}_{m+1})$. Q_{m+1} is a kind of geometric mean of i.i.d. F-variables,

$$Q_{m+1}(\underline{\eta}_m, \underline{\xi}_{m+1}) = \frac{1}{P}\sum_{j=1}^{P}\log(1 + \frac{n_1}{n_2}F^i_{m+1}(\underline{\eta}_m, \underline{\xi}_{m+1})),$$

$$F^i_{m+1}(\underline{\eta}_m, \underline{\xi}_{m+1}) = \frac{n_2}{n_1}\frac{\mathrm{tr}[(\mathbf{P}^i_{m+1}(\underline{\eta}_m, \underline{\xi}_{m+1}) - \mathbf{P}^i_m(\underline{\eta}_m))\hat{\mathbf{C}}^i_{\underline{X}}]}{\mathrm{tr}[(\mathbf{I} - \mathbf{P}^i_{m+1}(\underline{\eta}_m, \underline{\xi}_{m+1}))\hat{\mathbf{C}}^i_{\underline{X}}]},$$

where F^i_{m+1} is, under the hypothesis, approximately F-distributed with n_1 and n_2 degrees of freedom and stochastically independent for different frequencies ω^i. The degrees of freedom are as follows,

$$n_1 = 2p(1+r), \quad n_2 = 2p(N - (m+1) - r).$$

r is the number of the model parameters (in this application $r = 2$). It should be pointed out that the degrees of freedom n_1 and n_2 were corrected compared to those in [4] and [6]. This gives more accurate approximation of the test statistic distribution and improve the data analysis. F^i_{m+1} can be interpreted as an increace of spectral signal-to-noice ratio for a possible $m+1$st signal at $\underline{\xi}_{m+1}$. The mean and variance of the random variable $V = \log(1 + \frac{n_1}{n_2}F^i_{m+1}(\underline{\eta}_m, \underline{\xi}_{m+1}))$ can be calculated under hypothesis approximately as

$$\mu_V = \mathrm{E}\,V = \Psi\left(\frac{n_1}{2} + \frac{n_2}{2}\right) - \Psi\left(\frac{n_2}{2}\right) \approx \log(1 + \frac{n_1}{n_2}),$$

$$\sigma_V^2 = \mathrm{Var}\,V = \Psi'\left(\frac{n_2}{2}\right) - \Psi'\left(\frac{n_1}{2} + \frac{n_2}{2}\right) \approx \frac{2n_1}{n_2(n_1 + n_2)}.$$

where the function $\Psi(z)$ is defined as a log-derivative of the Γ-function: $\Psi(z) = \frac{d}{dz}\ln\Gamma(z)$. In this case we can approximate the distribution of Q_{m+1} using the central limit theorem. The detector for the $m+1$st signal then is

$$T = \sqrt{P}\,\frac{Q_{m+1}(\breve{\underline{\eta}}_m, \bar{\underline{\xi}}_{m+1}) - \mu_V}{\sigma_V} \gtrless \kappa^{\alpha}_{m+1},$$

where κ^{α}_{m+1} is the test threshold by the given probability α of the false alarm. We now approximate the test statistic T by a standard normally distributed variable. This formulation allows to avoid the cumbersome bootstrap technique used in [4], [6] and leads to a simple algorithm. To guarantee a global level α we alternatively use the sequentially rejecting Bonferroni-Holm test [5] with maximal possible number of signals $M_0 = 5$. Having detected the $m+1$st signal, we have to determine $\breve{\underline{\eta}}_{m+1}$ by further minimizing of $q_{m+1}(\underline{\eta}_{m+1})$ over $\underline{\eta}_{m+1}$, e.g., using initial value $(\breve{\underline{\eta}}'_m, \bar{\underline{\xi}}'_{m+1})'$.

4. EXPERIMENTS WITH MEASURED SEISMIC DATA

We apply the proposed algorithm to measured seismic data recorded by 24 seismometers of the GERESS array situated in the Bavarian Forest. Detection and localization of regional and local events is one of the main tasks of the GERESS array. The localization depends mainly on the velocity analysis of dominant onsets in the seismograms and the ability to discriminate between P- and S-phases. If we know the structure of the crust, a detailed classification is possible.
The regional seismic event caused by a blasting in an iron mine on the distance of 171 km from the array is analysed. We use a sliding window with length $T = 3.2\,s$ corresponding to 128 sampling values by a sampling frequency of 40 Hz and a shift of 20 values ($0.5\,s$). This short window length does not allow to smooth over frequencies as in (4). Instead we stabilized the estimate $\hat{\mathbf{C}}^i_{\underline{X}}$ by use of Thomson's orthogonal windows [9] with $L = 3$. The frequency band used in the analyse contains $P = 33$ frequency bins in the range from $0,5\,Hz$ to $10,5\,Hz$. The maximization of the likelihood function over the components of the parameter vector is a computationally difficult task. Nevertheless, the problem becomes manageable when a global optimization technique like genetic algorithm [2] is applied to (6) followed by a local optimization technique like scoring around parameters $\breve{\underline{\eta}}_m, \bar{\underline{\xi}}_{m+1}$. The population size for genetic algorithm is 20, the probability of crossing two strings is 0.80, and the probability of mutation is 0.05. We represent the elements of the vector $\underline{\xi}$ by a bit string of length 12. Signals were detected and, afterwards, estimated. Observing Figure 1 we note that detected signals are mostly in the arrival domains of the P-phase ("fast" longitudinal waves) and S-phase ("slow" transverse waves). Analysis of a regional seismic event was carried out using sequential testing with a global level $\alpha = 0,01$ and ACMLEs for azimuth and phase-velocities. The estimates of azimuth and velocities are in agreement with the physics of seismic wave propagation and azimuth of arrival detected by the Institute of Geophysics at the Ruhr University.

5. ACKNOWLEDGMENT

The authors wish to thank the Institute of Geophysics at the Ruhr University Bochum for providing measured data.

6. REFERENCES

[1] K. Aki & P.G. Richards, "Quantitative Seismology: Theory and Methods", Vol. I, Freeman, San Francisco 1980

[2] A.D. Bethke, "Genetic Algorithms as Function Optimizers", PhD Thesis, The University of Michigan, 1980

[3] J.F. Böhme, "Array Processing", in: *Advances in Spectrum Analysis and Array Processing*, Vol. II, S. Haykin ed., Prentice Hall, Englewood Cliffs N.J., pp. 1-63, 1991

[4] J.F. Böhme, "Statistical Array Signal Processing of Measured Sonar and Seismic Data", *Proc. SPIE 2563, Advanced Signal Processing Algorithms*, San Diego, 1995

[5] S. Holm, "A Simple Sequentially Rejective Multiple Test Procedure", *Scand. J. Statist.*, 6, pp. 65-70, 1979

[6] D. Maiwald, "Breidbandverfahren zur Signalentdeckung und -ortung mit Sensorgruppen in Seismik und Sonaranwendungen", (in German), Dr.-Ing. Dissertation, Fakultät für Elektrotechnik der Ruhr Universität Bochum, 1995

[7] D. Maiwald, D.V. Sidorovich, and J.F. Böhme, "Broadband Maximum Likelihood Wave Parameter Estimation using Polarization Sensitive Arrays", *Proc. ICASSP,* Minneapolis, pp. IV356-IV359, 1993

[8] R.H. Shumway, "Replicated time-series regression: an approach to signal estimation and detection", In D.R. Brillinger and P.R. Krishnaiah, editors, *Handbook of Statistics, Vol.3*, pp. 383-408. Elsevier Science Publishers B.V., 1983.

[9] D.J. Thomson, "Spectrum Estimation and Harmonic Analysis", *Proc. IEEE*, 70, pp. 1055-1096, 1982

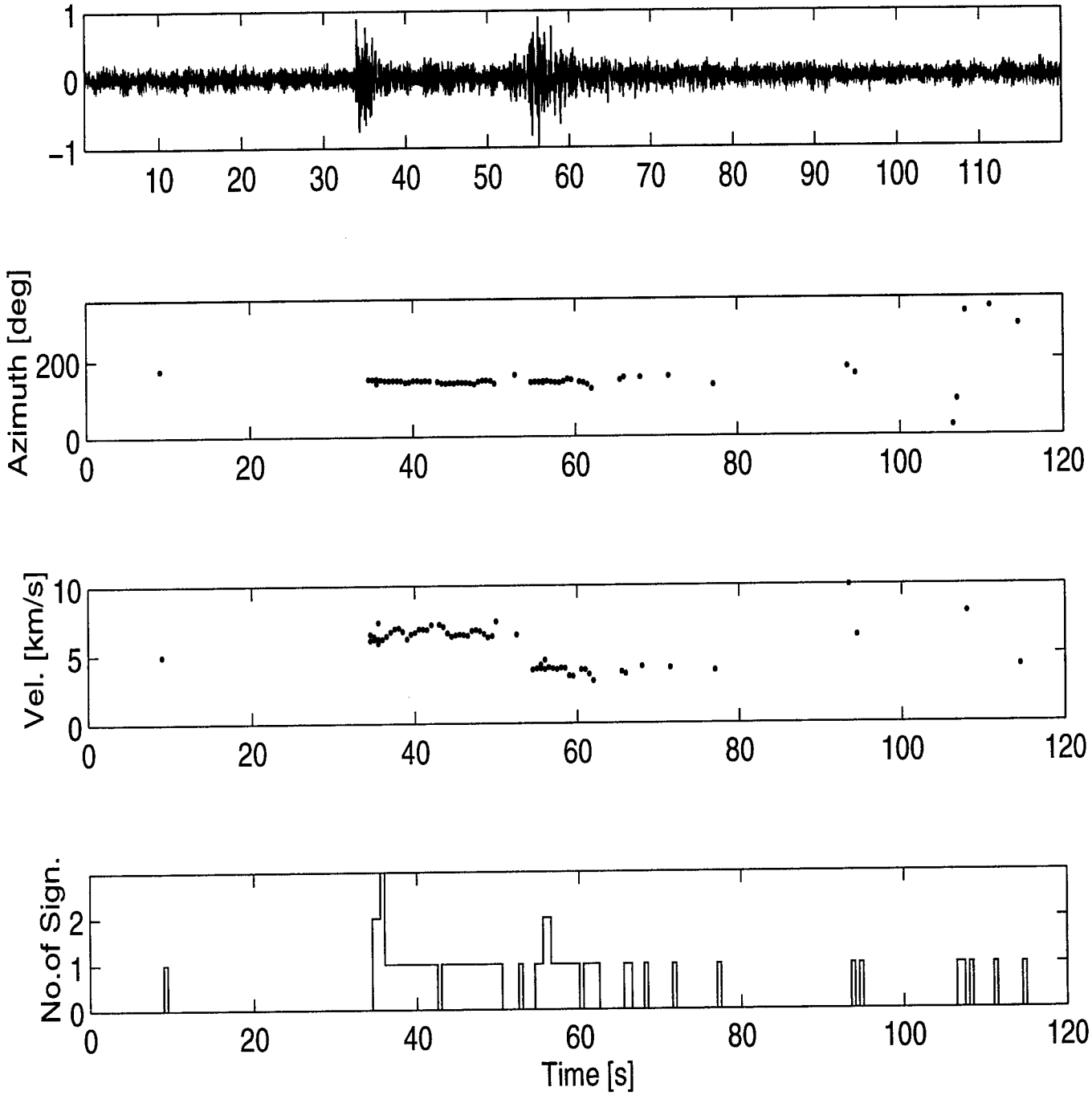


Figure 1: Analysis of a regional seismic event using sequential testing with a global level ($\alpha = 0,01$) control and ACMLEs for azimuth and phase-velocities. The received signal of a reference seismometer is shown on the top.

Time-dependent ARMA Modeling of Continuous Wave Ultrasonic Doppler Signals

M. E. Yüksel, S. Kara, N. Taşpınar
Department of Electronics Engineering, Erciyes University
Kayseri, 38039, TURKEY
yuksel@vm.cc.erciyes.edu.tr

Abstract

This paper presents a new approach to the problem of modeling nonstationary CW ultrasonic Doppler signals. A time-dependent ARMA model whose parameters vary periodically in time is proposed. The time variation of the model parameters is approximated by a weighted sum of a small number of Fourier base functions. It is seen that the spectral characteristics of the model output closely approximate to the spectral characteristics of the nonstationary CW ultrasonic Doppler signals.

1. Introduction

Doppler ultrasound is an important and powerful technique for noninvasive measurement of the velocities of moving particles within the body. The technique is particularly used for the measurement of the blood velocity and a number of other parameters related with blood flow. When employing the technique for blood velocity measurement, an ultrasonic signal is transmitted by an ultrasonic transducer through the blood vessel under examination. This signal is reflected by the red blood cells, causing an echo which is demodulated to yield an audible signal called *Doppler signal.* Frequency content of the Doppler signal is closely related with a number of flow parameters that provide valuable clinical information regarding the diagnosis of various vascular diseases and flow disorders. Time-variation of these parameters can be extracted from the Doppler signal by processing it with appropriate spectral analysis techniques.

In order to obtain clinical diagnostic information from the Doppler signal with minimum error, it is essential to use accurate and reliable techniques for the analysis of the signal. Since the Doppler signal is highly nonstationary and its true time-varying spectrum is unknown, it is impossible to test the accuracy and reliability of a technique with real Doppler signals. For this reason, new Doppler signal analysis techniques are first tested with simulated Doppler signals and then employed for the analysis of real signals if they prove to be reliable. Therefore it is very important to generate artificial signals whose power spectral density functions are known and spectral characteristics are as close to those of real Doppler signals as possible.

Modeling ultrasonic Doppler signals has been of high interest in the last decade since the models developed have not only led to a better understanding of the mechanism governing the generation of the Doppler signal but also provided valuable tools for the assessment of various Doppler signal analysis methods.

The classical method for the simulation of Doppler signals is based on the principle that the output of a linear filter excited by a white Gaussian noise is a Gaussian random process whose power spectral density is equal to the magnitude squared of the filter response. This approach is used by Kristoffersen and Angelsen [1] in a time shared B-mode imaging and Doppler measurement system to generate Doppler signal segments for filling in the gaps due to B-mode interruptions in the output of the Doppler unit.

Another approach to Doppler signal simulation is to first generate some stationary Doppler signals and then modulate their spectral characteristics with appropriate time-varying filters.This approach was used by Leeuwen*et al.* [2] to test some Doppler blood velocity measurement systems.

In all of the methods above, a theoretical time-varying Doppler power spectral density function is assumed. Next, a filter whose magnitude squared time-varying frequency response is similar to the theoretical Doppler spectrum is designed. This filter is excited with a white noise signal to yield nonstationary Doppler signal at its output. Although this approach is very

0-8186-7576-4/96 $5.00

practical, the design of the filter can be quite complicated if the spectral density of the signal changes rapidly in time as in the case of CW Doppler signals.

In order to overcome the diffulties arising from the design of filters having time varying frequency responses, Mo and Cobbold [3] proposed a nonstationary Doppler signal simulation model based on a weighted sum of sinusoidal components. The magnitudes of the components are obtained by evaluating a theoretical Doppler power spectrum. Though this approach does not require complex filter designs, it still relies on a theoretical Doppler power spectrum and is computationally more expensive than the other methods mentioned above.

In this paper, we present a new method for the simulation of the CW ultrasonic Doppler signals. The approach presented here does not depend on a theoretical power spectral density. When the time-frequency spectrogram for a CW Doppler signal is calculated, it is seen that the variation of power spectral density along the time-axis is similar to periodic with a period approximately equal to the average cardiac cycle duration. Then, if this signal is modeled using a parametric model, one expects the model parameters to show a similar periodic time variation provided that the model order, that is the number of the model parameters, is fixed. Based on these observations, a time-dependent autoregressive moving average (ARMA) model is employed in this paper to model the data. The variation of the model parameters is assumed to be periodic with a period equal to the average cardiac cycle. Hence, these parameters are approximated with a weighted combination of a small number of base functions. The Fourier base is chosen here among a number of bases available in the literature [4]. Comparisons in both time and frequency domain show that the method proposed here can successfully be used to model the CW ultrasonic Doppler signals.

2. Method

Let $x(0), \ldots, x(N-1)$ denote the N signal samples to be modeled which are obtained by equally sampling a CW Doppler signal along the cardiac cycle. For the purpose of modeling this signal, we propose a time-dependent ARMA model [5]

$$x(n)+\sum_{i=1}^{p} a_i(n-i)\,x(n-i) = e(n)+\sum_{i=1}^{q} b_i(n-i)\,e(n-i) \tag{1}$$

where $a_i(n-i)$ and $b_i(n-i)$ are the time dependent model parameters, p and q are model orders and $e(n)$ is the driving process which is a zero mean and unity variance white noise process.

Based on *a priori* knowledge, we assume that the time variation of the model parameters is periodic with a period equal to the number of the signal samples N. Then these parameters can be approximated by a weighted combination of a small number of base functions. By employing this approach and choosing the Fourier base, an approximate representation for the time variation of the parameters can be obtained as

$$a_i(n) = \sum_{k=-m}^{m} c_{i,k}\, e^{jknw_0}$$

$$b_i(n) = \sum_{k=-m}^{m} d_{i,k}\, e^{jknw_0}$$

where $2m+1$ is the number of base functions, $c_{i,k}$ and $d_{i,k}$ are the weights and $w_0 = 2\pi/N$. Then the quantities $a_i(n-i)\,x(n-i)$ and $b_i(n-i)\,e(n-i)$ become

$$a_i(n-i)\,x(n-i) = \sum_{k=-m}^{m} c_{i,k}\,x(n-i)\,e^{jk(n-i)w_0}$$

$$b_i(n-i)\,e(n-i) = \sum_{k=-m}^{m} d_{i,k}\,e(n-i)\,e^{jk(n-i)w_0}$$

which can also be expressed in vector form as follows

$$a_i(n-i)\,x(n-i) = \mathbf{u}^T(n-i)\,\mathbf{c}_i \tag{2}$$

$$b_i(n-i)\,e(n-i) = \mathbf{v}^T(n-i)\,\mathbf{d}_i \tag{3}$$

where

$$\begin{aligned}
\mathbf{u}(n-i) &= x(n-i)\,[e^{j(-m)(n-i)w_0} \ \ldots\ e^{jm(n-i)w_0}]^T \\
\mathbf{v}(n-i) &= e(n-i)\,[e^{j(-m)(n-i)w_0} \ \ldots\ e^{jm(n-i)w_0}]^T \\
\mathbf{c}_i &= [c_{i,-m}\ c_{i,-(m-1)}\ \cdots\ c_{i,m}]^T \\
\mathbf{d}_i &= [d_{i,-m}\ d_{i,-(m-1)}\ \cdots\ d_{i,m}]^T
\end{aligned}$$

Here the superscript T denotes matrix or vector transposition.

By substituting (2) and (3) into (1) we obtain

$$\begin{aligned}
x(n) + \mathbf{u}^T(n-1)\,\mathbf{c}_1 + \cdots + \mathbf{u}^T(n-p)\,\mathbf{c}_p = \\
e(n) + \mathbf{v}^T(n-1)\,\mathbf{d}_1 + \cdots + \mathbf{v}^T(n-q)\,\mathbf{d}_q
\end{aligned}$$

which may shortly be rewritten as

$$x(n) + \phi^T(n)\,\theta = e(n) \tag{4}$$

where

$$\begin{aligned}
\phi(n) &= [\mathbf{u}^T(n-1)\ \mathbf{u}^T(n-2)\ \ldots\ \mathbf{u}^T(n-p) \\
&\quad -\mathbf{v}^T(n-1)\ -\mathbf{v}^T(n-2)\ \ldots\ -\mathbf{v}^T(n-q)]^T \\
\theta &= [\mathbf{c}_1^T\ \mathbf{c}_2^T\ \ldots\ \mathbf{c}_p^T]^T
\end{aligned}$$

(4) is identical to (1) but it offers an important advantage: *model parameters of (4) are just constants and not dependent on time.* Therefore, we converted a time-varying modeling problem into a time-invariant modeling problem by making the definitions so far.

We now employ a least squares method to estimate the parameter vector θ of the new time-invariant model of (4) from the original signal samples $x(0), \ldots, x(N-1)$. If we define

$$\hat{x}(n) = -\phi^T(n)\,\theta \tag{5}$$

(4) takes the following form

$$x(n) - \hat{x}(n) = e(n) \quad \Rightarrow \quad x(n) = \hat{x}(n) + e(n)$$

Here $\hat{x}(n)$ may be thought of an approximation to the original signal sample $x(n)$ at time n. In this case $e(n)$ represents the approximation error. If we apply this approach for all the signal samples available, we obtain the following set of equations:

$$\begin{aligned} x(0) &= \hat{x}(0) + e(0) \\ x(1) &= \hat{x}(1) + e(1) \\ &\vdots \quad\quad \vdots \\ x(N-1) &= \hat{x}(N-1) + e(N-1) \end{aligned}$$

By substituting the definition (5) in the above set of equations, we get

$$\begin{aligned} x(0) &= -\phi^T(0)\,\theta + e(0) \\ x(1) &= -\phi^T(1)\,\theta + e(1) \\ &\vdots \quad\quad\quad\quad \vdots \\ x(N-1) &= -\phi^T(N-1)\,\theta + e(N-1) \end{aligned}$$

which can also be written in matrix form as

$$\mathbf{x} = \Phi\,\theta + \mathbf{e} \tag{6}$$

where

$$\begin{aligned} \mathbf{x} &= [x(0)\; x(1)\; \ldots\; x(N-1)]^T \\ \Phi &= [\phi^T(0)\; \phi^T(1)\; \ldots\; \phi^T(N-1)] \\ \mathbf{e} &= [e(0)\; e(1)\; \ldots\; e(N-1)]^T \end{aligned}$$

If we choose a cost function $V(\theta)$ as the sum of squared approximation errors

$$V(\theta) = \frac{1}{2}\,\mathbf{e}^T\,\mathbf{e} = \frac{1}{2}\sum_{n=0}^{N-1} e^2(n)$$

a least squares solution to the parameter vector θ in (6), which minimizes $V(\theta)$, can be found as

$$\theta = (\Phi^T\,\Phi)^{-1}\,\Phi^T\,\mathbf{x} \tag{7}$$

3. Results and discussion

We modeled a CW Doppler signal recording obtained from the carotid artery of a healty subject. The average cardiac cycle duration is found to be approximately 0.86s. The signal at the output of the Doppler unit is sampled along one cardiac cycle at a sampling rate of 10kHz.

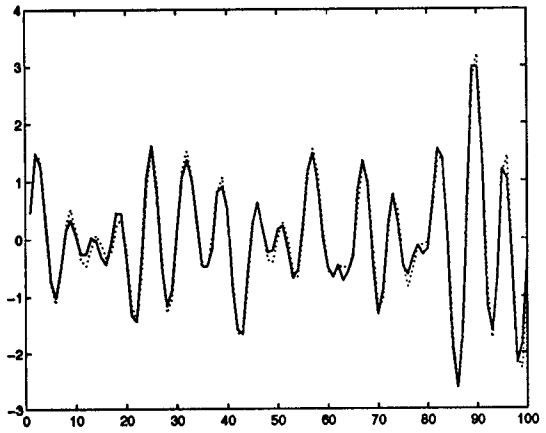
The samples are then modeled using the time-varying ARMA modeling approach proposed in the previous section. Several values for the parameters m and p are tried. It is seen that the performance of the simulation is not strictly dependent on m. The results presented here are for $m = 8$ and $p = q = 5$.

The signal at the output of the ARMA model is compared with the original CW Doppler signal both in frequency and time domains. For time-domain comparison, two sets of signal samples from original CW Doppler recording, one obtained during peak systole and the other at end-diastole, are compared with the corresponding sets of signal samples obtained from the simulated model. Each set contained 100 samples. For frequency domain comparison, time-frequency spectrograms for both signals are calculated by using the fast Fourier transform method via periodogram approach. The frame length is chosen to be 256 with a 50% overlap. Frames are windowed by using a length-256 Hanning window before taking the FFT.

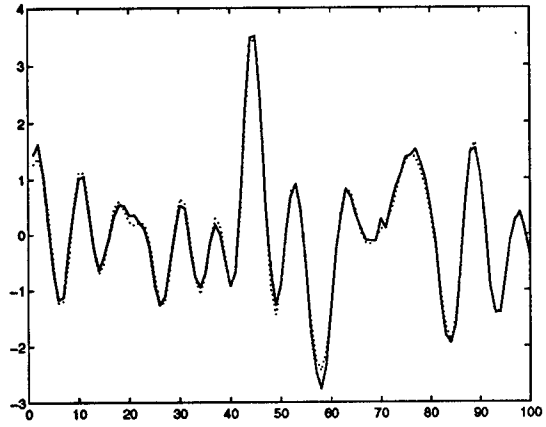
Figure 1 shows the comparison of the original signal and one realization of the simulated signal in time domain. In this figure, solid line represents the model output while the dotted line represents the original signal. Here it is seen that the model output very closely approximates to the original signal both at peak systole (a) and end-diastole (b).

Figure 2 shows the comparison of signal spectrograms for one cardiac cycle. In this figure, the horizontal axis shows time (t), the vertical axis frequency (f) and gray level at the coordinates (t, f) the power of the signal component with frequency f at time instant t. As it can easily be seen, the gray-scale speckle patterns are very similar. This shows that the time variation of the spectral characteristics of the CW Doppler signal is well represented by the simulation model. Furthermore, when the signal obtained at the output of the ARMA model is converted to analog form and played back for an audio comparison, it is found to be almost indistinguishable from the original recording.

The main advantage of the simulation method proposed in this paper over other methods in the literature is that it does not require the assumption of a theoretical power spectral density function. The parameters of the time varying ARMA signal model are estimated

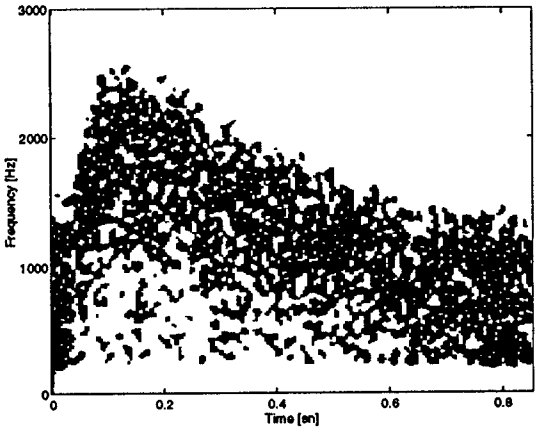


(a)

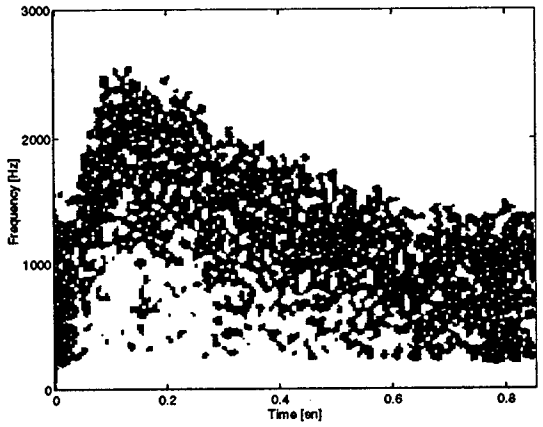


(b)

Figure 1. Comparison of original (dotted line) and simulated (solid line) signals in time domain: (a) peak systole, (b) end-diastole.



(a)



(b)

Figure 2. Comparison of (a) original and (b) simulated CW ultrasonic Doppler signal spectrograms.

by directly using the signal samples available.

4. Acknowledgments

M. E. Yüksel wishes to thank TÜBİTAK (The Turkish Scientific and Technical Research Council, Ankara, TURKEY) and Prof.Dr. A. G. Constantinides (Signal Processing Section, Dept. of Electrical Eng., Imperial College, London, U.K.) for their support to this work.

References

[1] K. Kristoffersen and B. A. J. Angelsen. A time-shared ultrasound Doppler measurement and 2-D imaging system. *IEEE Trans. BME*, 35(5):285-295, 1988.

[2] G. H. van Leeuwen, A. P. G. Hoeks and R. S. Reneman. Simulation of real time frequency estimators for pulsed Doppler systems. *Ultrason. Imag.*, 8:252-271, 1986.

[3] L. Y. L. Mo and R. S. C. Cobbold. A nonstationary signal simulation model for continuous wave and pulsed Doppler ultrasound. *IEEE Trans. UFFC*, 36(5):522-630, 1989.

[4] R. Charbonnier, M. Barlaud, G. Alengrin and J. Menez. Results on AR-modeling of nonstationary signals. *Signal Proc.*, 12:143-151, 1987.

[5] Y. Grenier. Time-dependent ARMA modeling of nonstationary signals. *IEEE Trans. ASSP*, 31(4):899-911, 1983.

A Change Point Detection Method for Elimination of Industrial Interference in Radio Astronomy Receivers

Peter Friedman
Special astrophysical observatory, St.Petersburg dpt.,Russian Academy of Science
196140 Pulkovo, St.Petesburg, Russia
fridman@fsao.spb.su

Abstract

The continuous growth of electromagnetic pollution produces many problems during radioastronomy observations. A method of rejection of industrial interference was proposed and tested.The main idea consits of using change- point detection with implementation in digital signal processor.The modified cumulative sum (CUSUM) method which analyzes several moments of empirical probability distribution of input noise and uses adaptive thresholds was elaborated, simulated on computer and then tested in the real observations on radiotelescope. The obtained sensitivity was much better than without such a processing. The proposed real-time signal processing procedure may be used at many of those radiotelescopes which suffer from the industrial interference, especially at long wavelengths.

1. Introduction

Electromagnetic pollution limits the real sensitivity of modern radiotelescopes,especially at long wavelengths. Industrial interference, radars, cars,radiostations etc. produce a lot of noise which is averaged with the natural noiselike signals from extraterrestrial radiosources. There were several attempts to build new radiotelescopes at radioecologically isolated places. Also radiointerferometry is less sensitive to industrial interference due to the absence of correlation of such a noise at long distances. But the problem of single-dish widesprectrum radioobservations in bad radioecological conditions forces to look for the special methods of signal processing which could improve the output precision of observational data. This report describes work on an real-time digital signal processing (DSP) system using special DSP processors from the TMS320 family - TMS320C53. This system was tested at RATAN-600 radiotelescope during 13cm and 31cm wavelength radioobservations.

2. Proposed Method and Main Results

The block diagram of radiometer with postdetection DSP stages is shown in Fig.1. The sampling frequency f_s of 10-bit analogue-digital converter(ADC) was limited by the serial link between ADC and DSP - 5 Mbit/s. In our case f_s was equal to 200 Ksamples/s.

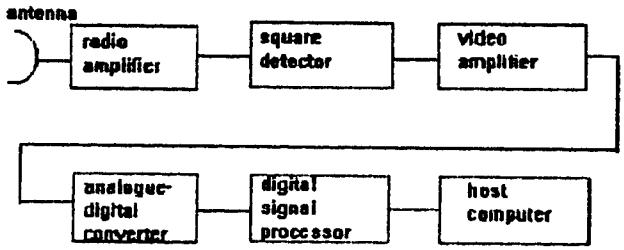


Figure 1. Block diagram of the receiver with digital back end.

First of all the form and structure of industrial noise were studied. The most typical were stochastic bursts with complex structures and radar impulses. The common processing in radiometers includes the averaging of fluctuations after detector during several seconds, minutes and sometimes hours. It is clear that in the presence of such an interferences smoothing of bursts deteriorates the output standard deviation (r.m.s. error). It is necessary to intercept these bursts without smoothing and eliminate them. The algorithm making this procedure must be quick, simple and effective. Several procedures were tested from this point of view and finally the modified cumulative sum

0-8186-7576-4/96 $5.00

method(CUSUM)[1] with adaptive threshold was chosen, simulated on computer and implemented in real radioobservations . The quality criterion of the algorithm was the radiometer fluctuation sensitivity - the r.m.s level of output oscillationsΔT. For the ideal total-power radiometer the minimum detectable change in the radiometer antenna temperature is equal[2]

$$T = \gamma T_{sys}\sqrt{\Delta F/\Delta f} \qquad (1)$$

where Δf - radio-frequency(predetection) bandwidth of the receiver, ΔF - postdetection bandwidth,T_{sys} - effective noise temperature of the radiotelescope. Ideal filters with rectangular passband forms are supposed. The mean value after the square-law detector is proportional to T_{sys} and must be measured with the most possible precision(minimum standard deviation).Factor ($\gamma > 1$ depends on the structure of radiometer[2] and in this work the radiometer with noise adding and antiphase gain modulation was used[3]. In the presence of interference the output postdetection oscillations may be written as

$$y = x_{sys} + \epsilon_{sys} \qquad (2)$$

where x_{sys}- noiselike system signal which mean $\overline{x_{sys}}$ and variance δx_{sys} are proportional to T_{sys}, and ϵ_{int} - oscillations due to interference. In digital case $f_s = 2\Delta F$ and the time of integration $\tau = n_0/f_s$, n_0 - number of averaged samples which characterize the integration interval,$n_0 = N \cdot M, N$ - length in samples of the modulation halfperiod, M - number of modulation periods in total averaging interval. The output standard deviation (1) in these terms is proportional to

$$\delta x1 = \overline{x_{sys}}\sqrt{1/(\Delta f\tau)} \qquad (3)$$

The post detection interference burst ϵ_{int} has mean ϵ_{int} and duration n_{xint} (in samples). The output mean $\overline{y_2}$ (system + interference) is equal to

$$\overline{y_2} = \overline{x_{sys}}(1-\alpha) + (\overline{x_{sys}} + \overline{\epsilon_{int}})\cdot\alpha \qquad (4)$$

where $n_{int}/n_0 = \alpha < 1, n_{int} = \sum_{i=1}^{M} I[i]\{\epsilon_{int} > 0\}$,indicator function $I[i]$ is defined as $I[i] = 1$, $\epsilon_{int}[i] > 0$ and $I[i] = 0$, $\epsilon_{int}[i] = 0$. For the detection of change points of mean at the l-th interval, $l = 1\cdots M$, the following CUSUM procedure was used:

$$S_0 = 0,$$

$$S_n = \sum_{i=1}^{n}(y[l,i] - \overline{y[l-1]} - \nu/2),$$

$$m_n = \min S_k, 0 < k < n \qquad (5)$$

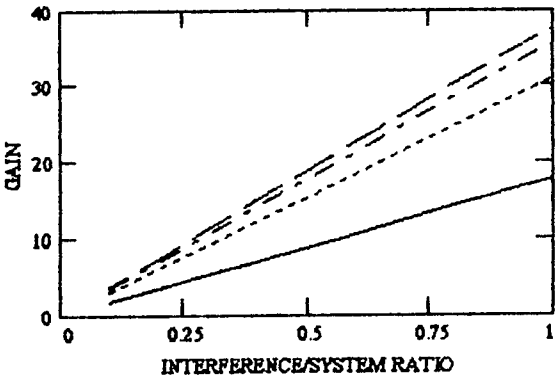


Figure 2. The ratio of the r.m.s.error without interference elimination to the error after this elimination,Q= 10000.
α=0.2------,α=0.4-----,
α=0.6-- -- -- ,α=0.8--·-·-·--

the moment of mean change $i_{ch} = \arg S_n[l,i] - m_n > h, h$ - threshold defined as $h = \beta\cdot\hat{\delta}y_{l-1}, \hat{\delta}y_{l-1}$- estimation of standard deviation at $l-1$ -th interval, β - factor chosen by operator, ν - the minimum expected value of mean chamge, $i = 1\cdots N$ - number of samples at each l-th interval.

After elimination of "contaminated" intervals with total duration n_{int} the standard deviation is

$$\delta y_3 = \overline{x_{sys}}\sqrt{1/2\Delta f(1-\alpha)\tau}, \qquad (6)$$

which is worse comparing with (3). The ratio of standard deviation without elimination of interference to δy_3 may be considered as the gain obtained in consequence of this procedure and is equal to

$$G_0 = \sqrt{(1-\alpha)(1+\alpha^2 Q(\overline{\epsilon_{int}}/\overline{\epsilon_{sys}})^2)}, \qquad (7)$$

where $Q = 2\Delta f\tau$ - radiometer factor. The curves for G_0 versus $\overline{\epsilon_{int}}/\overline{\epsilon_{sys}}$ are shown in Fig. 2, each curve corresponds to definite α , from 0.2 till 0.8.,$Q = 500.$//

The nonideal elimination due to the finite threshold h gives reduced G_1:

$$G_1 = \frac{\sqrt{(\overline{\epsilon_{int}}/\overline{\epsilon_{sys}})^2\alpha^2 + 1/Q)}}{\sqrt{\alpha^2 2\beta/(Q_0 N) + 1/(Q_0 N M(1-\alpha))}}, \qquad (8)$$

where $Q_0 = \Delta f/\Delta F$.

The results of interference elimination during RATAN-600 radioobservations at 31 cm wavelength in real time are given in Fig.3. One can see considerable elimination of interference without losing the total r.m.s. sensitivity.

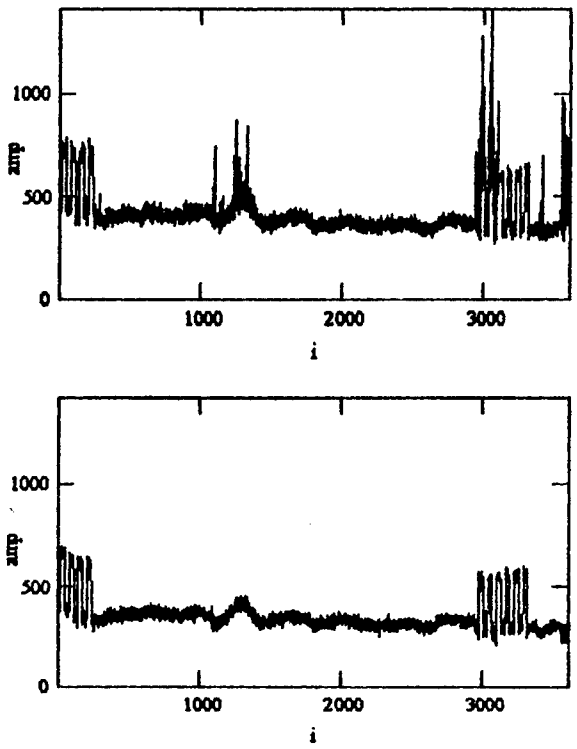


Figure 3. Example of the effectiveness of method: above - the record of radio source without "interference cleaning", below - real-timed cleaned record of the same radiosource,both records were made simultaneously;in the centre- radio source,left and right-calibration intervals.

3. Conclusion

Real time digital signal processing in a wide videoband may improve the observational situation in the presence of industrial interference. Rapid progress in DSP processing gives hope that the whole postdetection processing in radioastronomy technique will be made digitally.

This work was supported by the RFFI grant 95-02-03779.

References

[1] *Detection of Abrupt Changes in Signals and Dynamical Systems*, M.Basseville and A.Benveniste eds.,(1986),Springer-Verlag,Berlin.

[2] M.E.Tuiri,Radio-Telescope Receivers, in *Radio Astronomy*, J.D.Kraus,ed.,(1966), McGraw-Hill, Book Company, New York,Chapter 7.

[3] N.A.Esepkina,D.V.Korolkov,Y.N.Pariiski, *Radio Telescopes and Radiometers*,(1973) Nauka,Moscow, pp.286-289,(in Russian).

A Pseudo-Linear Method for Fractionally Integrated Arma (ARFIMA) Model Estimation

G.N. Fouskitakis and *S.D. Fassois*
Department of Mechanical Engineering
University of Patras
GR-265 00 Patras, Greece

Abstract

A Pseudo-Linear method for the estimation of Fractionally Integrated ARMA (ARFIMA) models is introduced. The method uses a long binomial series expansion of the fractional differencing operator, as well as the relationship of the AR/MA parameters and binomial expansion terms with the model's inverse function. It is based upon a pseudo-linear formulation motivated by the fact that this relationship leads to a special-form regression problem that can be decomposed into a scalar non-linear and a multiple linear regression. The performance characteristics of the method are demonstrated via Monte Carlo experiments and comparisons with the frequency-domain Maximum Likelihood method.

1. Introduction

Most of the work on time series analysis has been concerned with series characterized by the property that distant observations behave independently, or nearly so. Yet, in many empirical studies [1-3] the dependence between distant observations is not negligible and decays very slowly. Series with such long-term persistence are referred to as *long-memory* time series, and their power spectral density increases indefinitely as the frequency approaches zero, while their autocorrelation decays hyperbolically.

Long-memory time series aren't well represented by the usual stationary AutoRegressive Moving Average ($ARMA$) models, which are characterized by limited, at the origin, power spectral density and an exponentially decaying autocorrelation [4].

A class of models that exhibits the foregoing long-memory characteristics is that of *Fractionally Integrated ARMA (ARFIMA)* models. This in essence is an extension of the Integrated ARMA (ARIMA) models of Box *et al.* [4], in which the differencing operator is raised into a fractional, instead of the usual integer, power.

The majority of the available $ARFIMA$ model estimation methods follow a two-step approach, according to which an estimate of the fractional power is obtained (usually in the frequency domain) in the first step, and a standard $ARMA$ estimation technique is applied to the adjusted (filtered by the fractional differencing operator) time series in the second. These methods have been criticized for failing to produce good estimates for relatively short data records [2]. The alternative one-step methods advocate the simultaneous estimation of all model parameters based upon variants of the Maximum Likelihood procedure in either the time or the frequency domains [2,5-7]. A major drawback of this category of methods is their high computational complexity.

In this paper a simple and computationally efficient Pseudo-Linear method for $ARFIMA$ model estimation is introduced. The method uses exclusively time-domain operations and is based upon the decomposition of a special-form regression problem into a scalar non-linear and a multiple linear regression.

2. Problem statement

An $ARFIMA(n,d,m)$ process is of the form:

$$\Phi(B)\cdot(1-B)^d\cdot X_t = \Theta(B)\cdot\alpha_t$$
$$\alpha_t \sim i.i.d.\mathcal{N}(0,\sigma_\alpha^2) \quad d\in(-0.5,0.5) \qquad (1)$$

with t indicating discrete time, X_t the observed time series, α_t an independently identically distributed ($i.i.d.$) Gaussian sequence with the indicated mean and variance, B the backshift operator ($BX_t \triangleq X_{t-1}$), d the fractional power, and $\Phi(B)$, $\Theta(B)$ the autoregressive

0-8186-7576-4/96 $5.00

(AR) and moving average (MA) polynomials:

$$\Phi(B) \triangleq 1 + \sum_{i=1}^{n} \phi_i B^i \qquad \Theta(B) \triangleq 1 + \sum_{i=1}^{m} \theta_i B^i \quad (2)$$

Fractional differencing is defined by the binomial series expansion:

$$(1-B)^d = 1 + \sum_{j=1}^{\infty} \alpha_j \cdot B^j = 1 + \alpha_1 \cdot B + \alpha_2 \cdot B^2 + \ldots \quad (3)$$

$$\alpha_1 = -d, \quad \alpha_j = \frac{\alpha_{j-1} \cdot (j-1-d)}{j} \quad (j = 2, 3, \ldots) \quad (4)$$

The process representation (1) is assumed to satisfy the following standard assumptions:

A1. $d \in \mathcal{D} \triangleq (-0.5, 0.5)$, $\Phi(B) \neq 0$ for $|B| \leq 1$ *(stationarity conditions)*

A2. $\Theta(B) \neq 0$ for $|B| \leq 1$ *(invertibility condition)*

The problem of $ARFIMA$ process estimation may be then stated as follows: "Given time series data X_t $t \in [1, N]$, select a particular model $\mathcal{M}(\bar{\mathbf{p}})$ from the model set[1]:

$$\mathcal{M} = \{\mathcal{M}(\bar{\mathbf{p}}) : \Phi(B) \cdot (1-B)^d \cdot X_t = \Theta(B) \cdot e_t(\bar{\mathbf{p}})$$
$$\bar{\mathbf{p}} \triangleq [d\ \boldsymbol{\phi}^T\ \boldsymbol{\theta}^T\ \sigma_e^2]^T \in \mathcal{D} \times \mathcal{H}(\Phi) \times \mathcal{H}(\Theta) \times \Re^+\} \quad (5)$$

where $e_t(\bar{\mathbf{p}})$ represents the model's one-step-ahead prediction error, σ_e^2 its variance, $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ the AR and MA parameter vectors, respectively, and $\mathcal{H}(\Phi)$, $\mathcal{H}(\Theta)$ the regions of $\Re^n$, $\Re^m$ in which the stationarity and invertibility, respectively, conditions hold."

3. The pseudo-linear estimation method

The substitution of a truncated, p-th order, binomial series expansion of the fractional differencing operator (3) into the $ARFIMA$ representation (1) yields the $ARMA(p+n, m)$ representation:

$$(1 + \rho_1 \cdot B + \ldots + \rho_{p+n} \cdot B^{p+n}) \cdot X_t = \\ = (1 + \theta_1 \cdot B + \ldots \theta_m \cdot B^m) \cdot \alpha_t \quad (6)$$

with ρ_i defined by the convolution expressions:

$$\rho_i = \sum_{k=0}^{i} \alpha_k \phi_{i-k} \qquad (i = 1, 2, \ldots, p+n) \quad (7)$$

[1]Bold face lower-case/capital characters represent vector/matrix quantities.

with $\alpha_0 \equiv 1$, $\alpha_k \equiv 0$ $(k > p)$, $\phi_0 \equiv 1$, $\phi_k \equiv 0$ $(k > n)$. Denoting this representation's inverse function operator as:

$$I(B) \triangleq 1 + \sum_{i=1}^{\infty} I_i B^i \triangleq P(B)/\Theta(B) \quad (8)$$

and combining it with (7) yields:

$$\begin{array}{l} \alpha_1 + \Phi_1 - \Theta_1 = I_1 \\ \alpha_2 + \alpha_1\Phi_1 + \Phi_2 - I_1\Theta_1 - \Theta_2 = I_2 \\ \qquad\qquad \vdots \qquad\qquad \vdots \\ \alpha_p + \alpha_{p-1} \cdot \Phi_1 + \ldots + \alpha_{p-n} \cdot \Phi_n - I_{p-1} \cdot \Theta_1 - \ldots \\ -I_{p-m} \cdot \Theta_m = I_p \\ \alpha_p \cdot \Phi_1 + \ldots + \alpha_{p-n+1} \cdot \Phi_n - I_p \cdot \Theta_1 - \ldots \\ -I_{p-m+1} \cdot \Theta_m = I_{p+1} \\ \alpha_p \cdot \Phi_2 + \ldots + \alpha_{p-n+2} \cdot \Phi_n - I_{p+1} \cdot \Theta_1 - \ldots \\ -I_{p-m+2} \cdot \Theta_m = I_{p+2} \\ \qquad\qquad \vdots \\ \alpha_p \cdot \Phi_n - I_{p+n-1} \cdot \Theta_1 - \ldots - I_{p+n-m} \cdot \Theta_m = I_{p+n} \end{array} \quad (9)$$

3.1 Stage one estimation

In this stage initial parameter estimates are obtained based upon the inverse function operator (8).

Inverse function estimation

Consider the model:

$$I(B, \mathbf{i}) \cdot X_t = e_t^{ar}(\mathbf{i}) \quad (10)$$

that corresponds to the process representation implied by (8). In this model $I(B, \mathbf{i})$ represents a finite (truncated) s-th $(s > p+n)$ order approximation (permitted by way of assumption A2) of the inverse function operator, $\mathbf{i}$ the corresponding parameter vector, and $e_t^{ar}(\mathbf{i})$ the model's one-step-ahead prediction error at time t.

An interval estimate of the inverse function parameter vector is obtained through the expressions:

$$\hat{\mathbf{i}} = \left(\sum_{t=s+1}^{N} \boldsymbol{\psi}_t^{ar} (\boldsymbol{\psi}_t^{ar})^T \right)^{-1} \cdot \left(\sum_{t=s+1}^{N} \boldsymbol{\psi}_t^{ar} X_t \right) \quad (11)$$

$$\hat{C}ov[\hat{\mathbf{i}}] = (\hat{\sigma}_e^{ar})^2 \cdot \left[\frac{1}{N-s} \sum_{t=s+1}^{N} \boldsymbol{\psi}_t^{ar} (\boldsymbol{\psi}_t^{ar})^T \right]^{-1} \quad (12)$$

$$(\hat{\sigma}_e^{ar})^2 = \frac{1}{N-s} \sum_{t=s+1}^{N} [e_t^{ar}(\hat{\mathbf{i}})]^2 \quad (13)$$

with $\boldsymbol{\psi}_t^{ar} \triangleq [X_{t-1}\, X_{t-2} \ldots X_{t-s}]^T$.

Initial parameter estimation

Initial parameter estimation is based upon expressions (9) that relate the fractional power and AR/MA parameters with the model's inverse function. Given inverse function estimates, these lead to:

$$[\ \mathbf{A}\quad \mathbf{J}\ ]\cdot\mathbf{p}=\begin{bmatrix}\mathbf{j}_1-\mathbf{a}_1\\ \mathbf{j}_2\end{bmatrix}+\mathbf{e}\Leftrightarrow$$

$$\Leftrightarrow \mathbf{X}\cdot\mathbf{p}=\mathbf{y}+\mathbf{e} \tag{14}$$

where:

$$\mathbf{A}_{[(p+n)\times n]}(i,j)\triangleq\left\{\begin{array}{cc}0 & i<j\\ 1 & i=j\\ 0 & i\geq p+1+j\\ \alpha_{i-j} & otherwise\end{array}\right\} \tag{15}$$

$$\mathbf{J}_{[(p+n)\times m]}(i,j)\triangleq\left\{\begin{array}{cc}0 & i\prec j\\ -1 & i=j\\ -\hat{I}_{i-j} & otherwise\end{array}\right\} \tag{16}$$

$$\mathbf{p}\triangleq\left[\phi_1\ \ldots\ \phi_n\,\vdots\,\theta_1\ \ldots\ \theta_m\right]^T=\left[\boldsymbol{\phi}^T\,\vdots\,\boldsymbol{\theta}^T\right]^T \tag{17}$$

$$\mathbf{j}_1\triangleq[\ \hat{I}_1\ \ldots\ \hat{I}_p\ ]^T,\quad \mathbf{j}_2\triangleq[\ \hat{I}_{p+1}\ \ldots\ \hat{I}_{p+n}\ ]^T \tag{18}$$

$$\mathbf{a}_1=[\ \alpha_1\ \ldots\ \alpha_p\ ]^T \tag{19}$$

with $\mathbf{e}$ representing an error vector. Expressions (14) define a special-form regression problem that is nonlinear in the fractional power d but linear in the AR/MA parameter vector $\mathbf{p}$. The optimization of the regression cost function:

$$J(d,\mathbf{p})\triangleq\frac{1}{2}\mathbf{e}_k^T\mathbf{Q}_{k\times k}\mathbf{e}_k \tag{20}$$

may be then accomplished through a pseudo-linear two-step procedure, according to which the fractional power is varied through an appropriate search scheme and conditional, upon it, AR/MA parameter estimates are obtained as:

$$\hat{\mathbf{p}}(\hat{d})=\left(\mathbf{X}^T_{k\times(n+m)}\mathbf{Q}_{k\times k}\mathbf{X}_{k\times(n+m)}\right)^{-1}\mathbf{X}^T_{k\times(n+m)}\mathbf{Q}_{k\times k}\mathbf{y}_{k\times 1} \tag{21}$$

In the above k $(n+m<k<n+p)$ refers to the number of scalar equations of (14) actually used in the regression, while $\mathbf{Q}_{k\times k}$ represents a proper weighting matrix. The procedure is terminated once the minimum of $J(d,\mathbf{p})$ is achieved. The innovations variance is then estimated as:

$$\hat{\sigma}_a^2=\hat{\sigma}_e^2(\hat{d},\hat{\mathbf{p}})=\frac{1}{N-q}\sum_{t=q+1}^{N}e_t^2(\hat{d},\hat{\mathbf{p}}) \tag{22}$$

with $q\triangleq p+n+m$.

3.2 Stage two estimation

This stage aims at refining the estimates of stage one. Let:

$$\hat{\bar{\mathbf{p}}}_{i-1}\triangleq\left[\hat{d}_{i-1}\ \hat{\boldsymbol{\phi}}^T_{i-1}\ \hat{\boldsymbol{\theta}}^T_{i-1}\ (\hat{\sigma}_a^2)_{i-1}\right]^T$$

denote the vector of $ARFIMA$ parameter estimates obtained at iteration $i-1$, and initially equal to those provided by stage one. At iteration i these estimates are updated as follows:

Fractional power and AR parameter estimation

Assuming small perturbations in the MA parameter estimates during successive iterations, that is $\Theta(B,\hat{\bar{\mathbf{p}}}_{i-1})\approx\Theta(B,\hat{\bar{\mathbf{p}}}_i)$ the $ARFIMA$ model (5) may, at iteration i, be approximately expressed as:

$$e_t(\hat{\bar{\mathbf{p}}}_i)\approx\Phi(B,\hat{\bar{\mathbf{p}}}_i)\cdot(1-B)^{d_i}\cdot\bar{X}_t^{i-1} \tag{23}$$

with:

$$\bar{X}_t^{i-1}\triangleq X_t/\Theta(B,\hat{\bar{\mathbf{p}}}_{i-1}) \tag{24}$$

The model (23) is of the Fractionally Integrated AutoRegressive $[FIAR(n,d)]$ form, and its parameters may be estimated via a procedure similar to that of Stage 1. In this case equation (14) is such that $\mathbf{X}=\mathbf{A}$, $\mathbf{p}=\boldsymbol{\phi}$, and the weighting matrix $\mathbf{Q}_{k\times k}$ in (20) is selected equal to the corresponding submatrix of the estimated inverse function covariance (12). Due to the form of (14) in this case, this leads to optimal, in the sense of the Gauss-Markov theorem [8], estimates.

MA parameter and innovations variance estimation

The MA parameters are then updated by solving the linear regression problem [obtained from (14)]:

$$\mathbf{J}\cdot\boldsymbol{\theta}=\mathbf{y}^M+\mathbf{e}^M \tag{25}$$

with $\mathbf{e}^M$ denoting the regression error vector, and:

$$y_i^M=\hat{I}_i-\sum_{k=0}^{i}\hat{\alpha}_k\hat{\phi}_{i-k}\qquad(i=1,2,\cdots p+n) \tag{26}$$

where $\alpha_0\equiv 1$, $\alpha_k\equiv 0$ $(k>p)$, $\phi_0\equiv 1$, $\phi_k\equiv 0$ $(k>n)$. The innovations variance is updated through (22).

4. Numerical experiment

Consider the $ARFIMA(1,d,2)$ process with parameters indicated in Table 1. This process is characterized by a sharp spectral valley, owing to the proximity

Parameter	Actual	Estimate ± std. deviation	
		PL	ML
d	0.30	0.313 ± 0.086	0.301 ± 0.067
ϕ_1	0.60	0.582 ± 0.073	0.587 ± 0.062
θ_1	-0.40	-0.396 ± 0.032	-0.380 ± 0.032
θ_2	0.99	0.916 ± 0.031	0.865 ± 0.097
σ_a^2	1.00	0.950 ± 0.086	1.124 ± 0.136

Table 1. Monte Carlo estimation results by the Pseudo-Linear (PL) and Maximum Likelihood (ML) methods ($N = 300$; 20 runs).

of its complex conjugate pair of zeros to the unit circle (magnitude of 0.995).

Monte Carlo estimation results by the Pseudo-Linear (PL) ($p = 20, s = 30$) and frequency-domain Maximum Likelihood (ML) [9] methods are, based upon 300-sample-long data records, summarized in Table 1. Despite the relatively short data record length and the significantly higher computational complexity of the ML method, the performance characteristics of the two methods appear similar. The ML method provides a slight improvement in the fractional power estimate, while, quite interestingly, the PL method achieves a noticeable improvement in the MA parameter estimates.

These observations are additionally confirmed from the estimated spectra, which are contrasted to the theoretical process spectrum in Figure 1. From these it is indeed evident that the PL method achieves a significantly reduced estimation scatter in the neighborhood of the spectral valley, while also providing a few estimates that are practically indistinguishable from the theoretical curve in the proximity of the spectral minimum. Similar behavior is observed with longer ($N = 1,000$) data records and processes characterized by sharp spectral peaks [10].

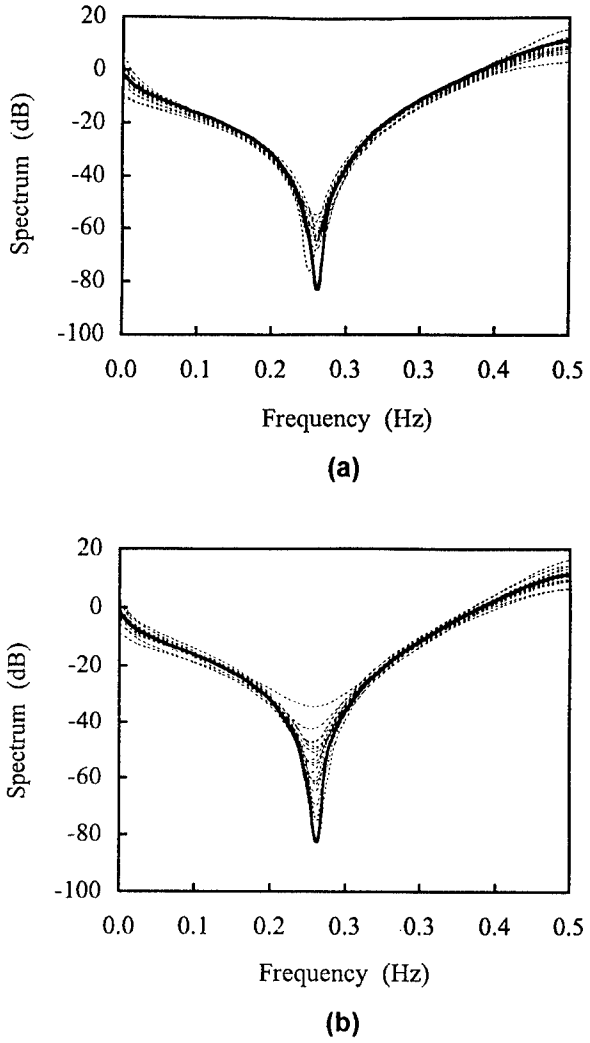


Figure 1. The actual (——) and estimated (- - - -) $ARFIMA(1, d, 2)$ spectra: (a) Pseudo-Linear method; (b) Maximum Likelihood method ($N = 300$; 20 runs).

References

[1] J.R.M. Hosking. Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research*, 20:1898-1908, 1984.

[2] M. Deriche and A.H. Tewfik. Signal modeling with filtered discrete fractional noise processes. *IEEE Transactions on Signal Processing*, 41:2839-2849, 1993.

[3] R. Ben Mrad, S.D Fassois, J.A. Levitt and B.I. Bachrach. On-board prediction of power consumption in automotive active suspension systems - parts I and II. *Journal of Mechanical Systems and Signal Processing*, in press, 1996.

[4] G.E.P. Box, G.M. Jenkins and G.C. Reinsel. *Time Series Analysis: Forecasting and Control.* Third Edition, Prentice-Hall, 1994.

[5] R. Fox and M.S. Taqqu. Large sample properties for strongly dependent stationary gaussian time series. *Annals of Statistics*, 14:517-532, 1986.

[6] R. Dahlhaus. Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Annals of Statistics*, 16:808-841, 1988.

[7] F. Sowell. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53:165-188, 1992.

[8] J.M. Mendel. *Lessons in Digital Estimation Theory.* Prentice-Hall, 1987.

[9] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods.* Second Edition, Springer-Verlag, 1991.

[10] G.N. Fouskitakis. *Pseudo-Linear Estimation of Fractionally Integrated ARMA Processes with Applications.* Diploma Thesis, Department of Mechanical Engineering, University of Patras, 1995.

Moment Analysis of Ambient Noise Dominated by Local Shipping

Lisa A. Pflug and Pam M. Jackson
Naval Research Laboratory, Code 7176, Stennis Space Center, MS 39529-5004
pflug@nrlssc.navy.mil jackson@nrlssc.navy.mil

Juliette W. Ioup and George E. Ioup
Dept. of Physics, University of New Orleans , New Orleans, LA 70148
jwiph@uno.edu

Abstract

The stationarity and local Gaussianity of ambient shipping noise recorded during an experiment conducted in the San Diego port area is investigated. First through fourth order moments are used to identify time periods of nonstationarity in the noise. Comparison of the shipping data with colored Gaussian noise indicates that the third order moments deviate from Gaussianity more than the fourth order moments. The local Gaussianity is quantified using the Kolmogorov-Smirnov test. While the shipping noise at the deeper depth appears somewhat nonGaussian during certain time periods, the shallower depth data appears Gaussian.

1. Introduction

In an effort to detect quieter signals in noisier environments, alternative detection algorithms using higher order statistics have been proposed by various researchers and have shown promise in simulations for a variety of scenarios. Many of these detectors exploit the difference between the higher order statistics of the signal and the noise, which is ideally stationary and Gaussian. However, in shallow-water environments complicated by factors such as heavy shipping, surf noise, and multipath signal distortion, detection algorithms are generally not optimal and the assumption that ambient noise is Gaussian, or even stationary, may not hold.

Computer simulations and theoretical developments have shown that higher order moments can passively detect transient signals in Gaussian noise better than the ordinary cross correlation detector. In studies by Ioup et al. [1] and Pflug et al. [2], [3], correlations are calculated using information from multiple sensors, with the number of sensors required being equal to the order of correlation. In this situation, the second and higher order moments of the noise have no effect on detector performance if the noise is uncorrelated. However, if correlations are formed by repeating information from only one sensor, which is sometimes all that is available for processing, then the higher order moments of the noise can affect detection. To extend this work on higher order correlation transient detectors to complicated shallow water environments, it is important to have a realistic estimate of the higher order moments of ambient noise and their stationarity.

In this work, measured shallow water ambient noise due primarily to shipping is investigated with the goal of determining whether stationarity and Gaussianity assumptions for transient detectors are appropriate, and if so, for what time periods. While it is generally accepted that ambient noise due to nearby shipping is nonGaussian, only a few attempts have been made to explore the nature of the nonGaussianity [4]-[7]. In [4], Brockett et al. examine the third order statistics of noise dominated by distant shipping or by one nearby ship. In [5], Hinich et al. show that the towing platform in an experiment has strong bispectral components. Richardson and Hodgkiss [6] use the bicoherence to determine that a recorded deep-water time series is nonGaussian. Only Dalle Molle and Hinich [7] consider the fourth order statistic, showing that the noise generated by two ships approximately 460m from a sonobuoy is not significantly different from Gaussian noise. However, none of these studies investigates the statistics of ambient noise generated by a multitude of nearby ships, such as would occur in a port area.

2. SWellEX-3 Experiment

Ambient noise recorded during the SWellEX-3 experiment is used to investigate the first, second, and higher order moments of ambient noise due to shipping in a moderately busy port area. The data were taken near the port of San Diego, California, in July-August 1994 [8]. Ambient noise measurements were recorded on a vertical 64-element array with 2 m spacing, located in water approximately 200 m deep. Two channels of data were chosen for analysis, 2 and 43, with respective depths of 192 m and 116 m.

The data are sampled at 1500 samples/second. Only 3-minute data segments have been used for analysis so far. The data are calibrated and mooring platform self-

0-8186-7576-4/96 $5.00

noise is reduced/removed. Additionally, a high-pass Butterworth filter of order nine with a cutoff frequency of 15 Hz was applied in an attempt to reduce the effects of sensor motion, or flow noise, that appeared in the uppermost phones.

Ships in the port were tracked with radar during the experiment and are used to identify times of low, moderate, and high shipping activity for analysis. It should be noted that shipping traffic in the area is always significant, and the terms low, moderate, and high are only relative and describe the number of ships in the general area and to some degree the proximity to the array.

3. Higher Order Moment Analysis

Three different length processing windows are used to calculate the changing first through fourth order moments of the data, defined by

$$m_p = \frac{\Delta t}{T} \sum_{k=0}^{N-1} n^p(k\Delta t)$$

where n(t) is the recorded noise, Δt is the sampling interval, T is the window duration, and p is the order of correlation. A 99% overlap of the sliding window corresponding to a moment sampling rate of 50 samples/second was found sufficient to prevent aliasing in the time-variation of the second, third, and fourth order moments. However, a small degree of aliasing still exists in the first order moment. Calculations of processing window length versus mean moment values indicate that the moments are reasonably stable for one-second intervals.

The moments for channel 2 using a one-second processing window during the moderate noise period are given in Fig. 1. For comparison, simulated stationary Gaussian noise with the same standard deviation and approximate color as the data segments is also analyzed with the results shown in Fig. 2. The moment means and standard deviations are given above each plot. The colored Gaussian noise is high-pass filtered in the same manner as the data. As expected, the filtering has little effect on the moments of the Gaussian noise. A similar analysis is performed for the remaining data, and is summarized by the moment means and standard deviations in Tables 1, 2, and 3 in the Appendix. The three sets of Gaussian noise in each table have the same 3-minute standard deviation as the corresponding shipping noise.

From visual comparison of Figs. 1 and 2, the shipping data appear nonstationary over the 3-minute time segment. Theoretically, the moments of a stationary process are constant over time. The variability seen in Fig. 2 is a result of using finite, rather than infinite, sums in the moment calculations. If the shipping data were also stationary, they should have a similar variation in their moments. Comparison of the standard deviations of the four data moments and the stationary Gaussian noise moments reveals that the moments of the shipping noise vary much more than those of the stationary noise for this case. To some extent, these differences can be used to quantify the degree of nonstationarity present in the data.

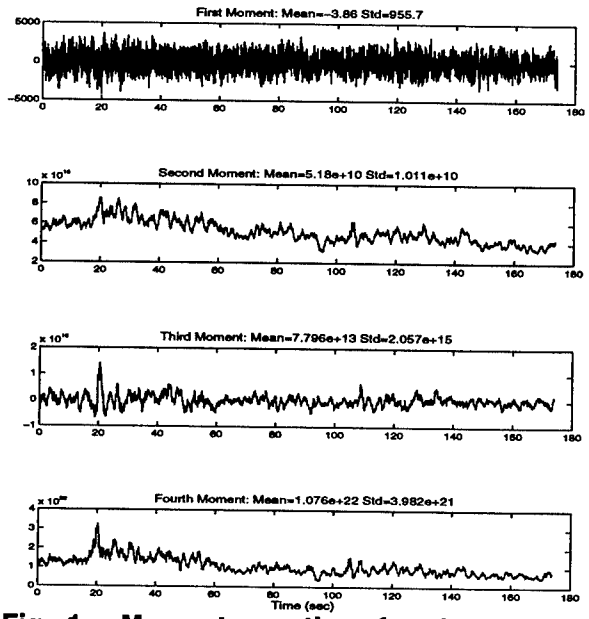


Fig. 1. Moments vs time for the channel 2 moderate-noise shipping data.

Even in the presence of nonstationarities, inferences about the local Gaussianity of the data can be made. For a zero-mean Gaussian process, infinite sums in the moment calculations should result in $m_1 = m_3 = 0$ and $m_4 = 3m_2^2$. The use of finite sums in the calculations results in deviations from these relationships. For example, the mean value of m_4 for the Gaussian noise shown in Fig. 2 shows a 1.35% difference from $3m_2^2$. However, the corresponding comparison for the shipping data shown in Fig. 1 reveals a -34.17% difference, indicating a departure from Gaussianity, averaged over time, due to other than finite sums. For the data analyzed, this is by far the largest difference between the data and simulated Gaussian averages. For the low and high level of noise at channel 2, the differences are -1.18% and -3.02%. At channel 43, the differences between $3m_2^2$ and m_4 for the low and moderate levels of noise are -6.98% and -3.51%. The various sets of Gaussian noise show differences between 1.35% and 2.48%. Except for the moderate noise at channel 2, the means of the fourth moments of the shipping data match those for simulated Gaussian noise closely and have relationships to the second moment that are consistent with Gaussianity. In contrast, the means of the third moments of the shipping data do not match those for the simulated

Gaussian noise except for the high shipping noise at channel 43, which differs by only -7.34%. The remaining sets of data have third moments that differ from the simulated Gaussian noise by at least 80.95%. Although the values for m_3 in Tables 1 and 2 appear large, they are small when compared to the magnitude of the cube of the data.

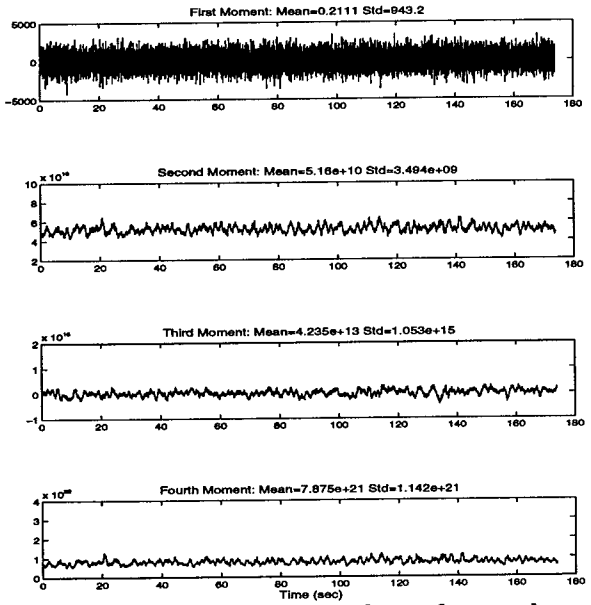


Fig. 2. Moments vs time for colored Gaussian noise.

4. K-S Test for Gaussianity

Although the shipping data has periods of nonstationary, the Kolmogorov-Smirnov (K-S) test can be used to assess the local Gaussianity of the data. Using a 1-second sliding window with 90% overlap, the K-S test was applied to both the shipping data and the simulated stationary Gaussian data. Values of the K-S statistic over the three-minute moderate noise time period for channels 2 and 43 are shown in Figs. 3 and 4. The horizontal lines at 0.035 represent the level above which the K-S statistic is different from a theoretical Gaussian distribution at the 5% significance level. The Gaussian assumption is rejected 18.38% of the time for channel 2. In contrast, Fig. 4 shows that the Gaussian assumption is rejected only 0.0057% of the time for channel 43. The simulated Gaussian noise is never rejected as Gaussian at the 1% level and is rejected less than 0.0458% of the time at the 5% level. The average K-S statistics for all the Gaussian noise sets range from 0.0163 to 0.0175. The average K-S statistics for the shipping data are also within this range, except for the noise at channel 2 during the moderate noise period. However, channel 2 during the moderate and high noise periods exhibits local peaks suggesting local nonGaussianity.

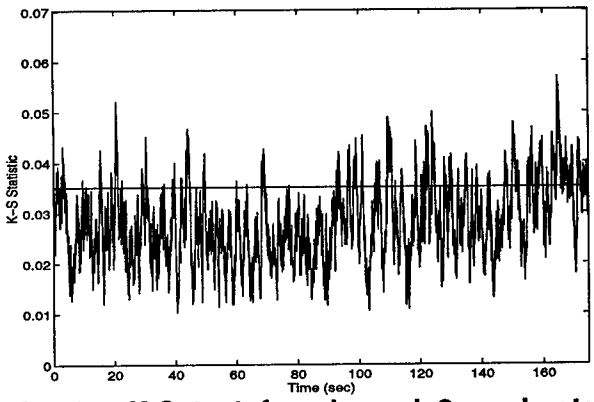


Fig. 3. K-S test for channel 2 moderate level shipping noise.

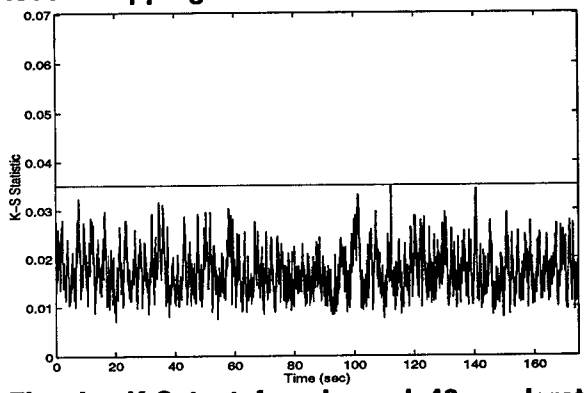


Fig. 4. K-S test for channel 43 moderate level shipping noise.

5. Conclusions

The higher order statistics of ambient noise due primarily to nearby ship traffic in a port area are analyzed. The analysis includes two channels of data from a vertical array for three different periods of noise. Examination of the first through fourth order moments over time reveals apparent nonstationarities in the shipping data. Comparison with moments of stationary colored Gaussian noise supports this conclusion. While the third order moments of the shipping data differ somewhat from that of the Gaussian noise, the fourth order moments differ much less. The Kolmogorov-Smirnov test indicates that the noise at the deeper hydrophone appears to have periods of local nonGaussianity during the moderate and high noise segments, while the noise at the shallower hydrophone, and at both hydrophones during the low noise segment, appears relatively Gaussian.

Acknowledgments

The authors are grateful for the assistance of Dr. Gerald D'Spain and Dr. William Hodgkiss, both of Scripps Institute of Oceanography, in obtaining and

calibrating the SWellEX-3 data, and for helpful suggestions in the analysis.

References

[1] Ioup, G. E., Pflug, L. A., Ioup, J. W., and Field, R. L., "Prefiltering for Higher Order Advantage," Signal Processing Workshop on Higher-Order Statistics, 309-313 (1993).

[2] Pflug, L. A., Ioup, G. E., Ioup, J. W., and Field, R. L., "Prefiltering for Improved Correlation Detection of Bandlimited Transient Signals," J. Acoust. Soc. Am., 95, 1459-1473 (1994).

[3] Pflug, L. A., Ioup, G. E., Ioup, J. W., and Field, R. L., "Prediction of SNR Gains for Passive Higher Order Correlation Detection of Energy Transients," J. Acoust. Soc. Am., 98, 248-260 (1995).

[4] Brockett, P. L., Hinich, M., and Wilson, G. R., "Nonlinear and Non-Gaussian Ocean Noise," J. Acoust. Soc. Am., 82, 1386-1394 (1987).

[5] Hinich, M. J., Marandino, D., and Sullivan, E. J., "Bispectrum of Ship-Radiated Noise," J. Acoust. Soc. Am., 85, 1512-1517 (1989).

[6] Richardson, A. M. and Hodgkiss, W. S., "Bispectral Analysis of Underwater Acoustic Data," J. Acoust. Soc. Am., 96, 828-837 (1994).

[7] Dalle Molle, J. and Hinich, M. J. "Trispectral Analysis of Stationary Random Time Series," J. Acoust. Soc. Am., 97, 2963-2978 (1995).

[8] Bachman, R. T., Schey, P. W., Booth, N. O., and Ryan, F., J., "Geoacoustic Data Bases for Matched-Field Processing: Preliminary Results in Shallow Water off San Diego, California," accepted for publication in the J. Acoust. Soc. Am., March, 1996.

Appendix

Noise Type	m_1	m_2	m_3	m_4
Low Gaussian Noise	0.34 308.60	1.10e10 5.06e8	7.98e12 8.33e13	3.54e20 3.35e19
Moderate Gaussian Noise	0.21 -943.20	5.16e10 3.49e9	4.23e13 1.05e15	7.88e21 1.14e21
High Gaussian Noise	0.26 760.90	4.85e10 2.85e9	3.73e13 9.00e14	6.93e21 8.54e20
Low Shipping Noise	-7.40 670.80	1.11e10 1.69e9	1.52e12 8.83e13	3.74e20 1.21e20
Moderate Shipping Noise	-3.90 955.70	5.18e10 1.01e10	7.80e13 2.06e15	1.08e22 3.98e21
High Shipping Noise	-1.50 1028.00	4.87e10 8.39e9	-1.13e14 9.38e14	7.33e21 2.59e21

Table 1. Channel 2 moment mean and standard deviations using a 1-second processing window.

Noise Type	m_1	m_2	m_3	m_4
Moderate Gaussian Noise	0.45 1007.00	5.91e10 3.62e9	7.27e13 1.13e15	1.03e22 1.30e21
High Gaussian Noise	0.62 964.80	6.31e10 3.70e9	7.22e13 1.37e15	1.17e22 1.44e21
Moderate Shipping Noise	-6.50 1407.00	5.96e10 1.34e10	1.61e14 1.29e15	1.14e22 4.88e21
High Shipping Noise	-4.30 1441.00	6.37e10 1.03e10	7.75e13 1.45e15	1.26e22 4.12e21

Table 2. Channel 43 moment mean and standard deviations using a 1-second processing window.

	Channel 2			Channel 43		
Noise Type	Average K-S Statistic	1% Level	5% Level	Average K-S Statistic	1% Level	5% Level
Low Shipping Noise	0.0163	0.000	0.000	------	------	------
Moderate Shipping Noise	0.0280	3.347	18.28	0.0171	0.000	0.0057
High Shipping Noise	0.0171	0.0011	0.3777	0.0177	0.000	0.0343

Table 3. Average K-S statistic and percentage processing windows for which K-S statistic is above the 1% and 5% significance levels.

Constrained Adaptive Beamforming with an Uncalibrated Array[1]

Dean O. Carhoun[2]
Gemini Industries, Inc.
Bedford, MA 01730
email: carhoun@ll.mit.edu

Abstract

A method of constrained adaptive beamforming employing a self-focussing technique for an uncalibrated array is described and experimental results are presented. Constrained adaptive beamforming was employed to assess the performance of adaptive nulling to suppress non-gaussian atmospheric noise and Doppler-spread ionospheric clutter received during bistatic radar experiments. A self-focussing method based on principal component analysis of the received data was devised and used to estimate the induced steering vectors of sources of interest. The procedure is described and results of subspace projection nulling of the unwanted noise and interference to enhance signal-to-noise ratio are presented.

Introduction

A method of constrained adaptive beamforming employing a self-focussing technique was used with an uncalibrated array. The receiving antenna consisted of a thinned planar array of 96 high frequency vertical elements deployed randomly over a 3-kilometer aperture to produce a narrow pencil beam, but with elevated sidelobes. Constrained adaptive beamforming was employed to assess the performance of adaptive nulling to suppress non-gaussian atmospheric noise and Doppler-spread ionospheric clutter received during bistatic radar experiments. During the experiments the random array had not been calibrated for receiver phase and amplitude differences, or cabling differences, although the surveyed positions of the elements were known. A self-focussing method based on principal component analysis of the received data was devised and used to estimate the induced steering vectors of sources of interest. The procedure will be described and results of subspace projection nulling of the unwanted noise and interference to enhance signal-to-noise ratio are presented below.

The Experiment

An aerial view of the receiver site taken during construction is shown in figure 1, indicating the cable runs to the 96 elements randomly deployed over the essentially planar aperture. Figure 2 shows a schematic diagram of the adaptive processing applied to the elements of the planar antenna array. The principal component inverse version of the generalized sidelobe canceller, described by Kirsteins and Tufts, was used. This was applied after range-Doppler processing the outputs of the antenna array receiver elements, which formed the receiving station of a bistatic over-the horizon radar transmitting a linear fm waveform. When the analysis was performed, the array data had not been calibrated to compensate for receiver phase and amplitude differences, cable length differences, and siting errors. Therefore, instead of computing the beamsteering vectors and beamforming constraint vectors from the known parameters of the array based on a planar wavefront propagation assumption, it was necessary to estimate these vectors from the principal components induced by the incident field, accomplished by applying an eigenvector beamforming technique. To estimate the beamsteering weight vector that points the array in the direction of a received signal at a particular Doppler frequency in a range cell being analyzed, a subset of Doppler samples for the array, local to the chosen Doppler frequency, was used to estimate a signal subspace for that frequency. The principal eigenvector belonging to the largest eigenvalue was used to estimate the beamforming vector $\mathbf{w}_0$. The beamforming constraint matrix $\mathbf{C}$ was formed from beamforming vectors

[1] This work was performed as part of the sponsored research program of The MITRE Corporation while the author was affiliated with Gemini Industries, Inc. [2] The author is currently affiliated with MIT Lincoln Laboratory, Lexington, MA.

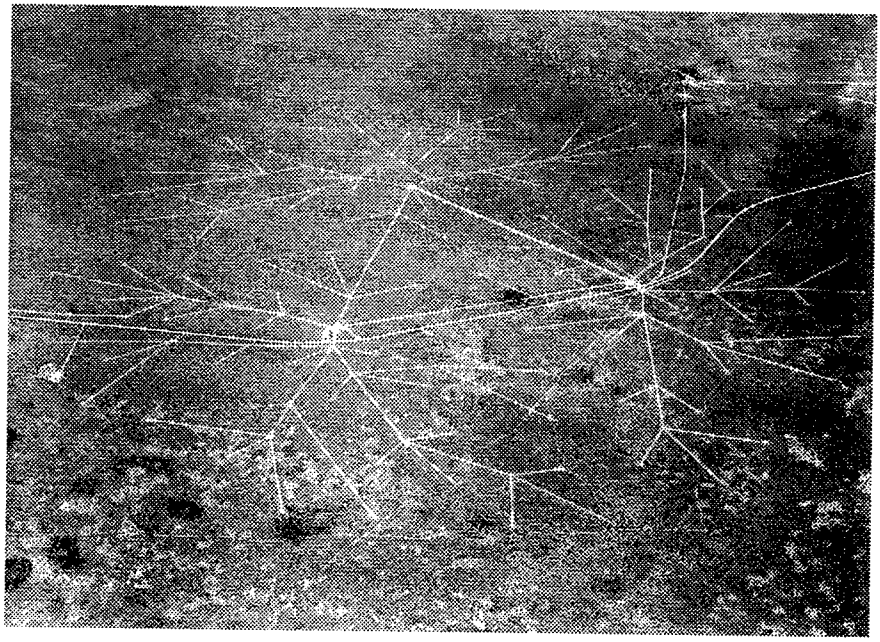


Figure 1. Aerial View of the Receiving Array Site

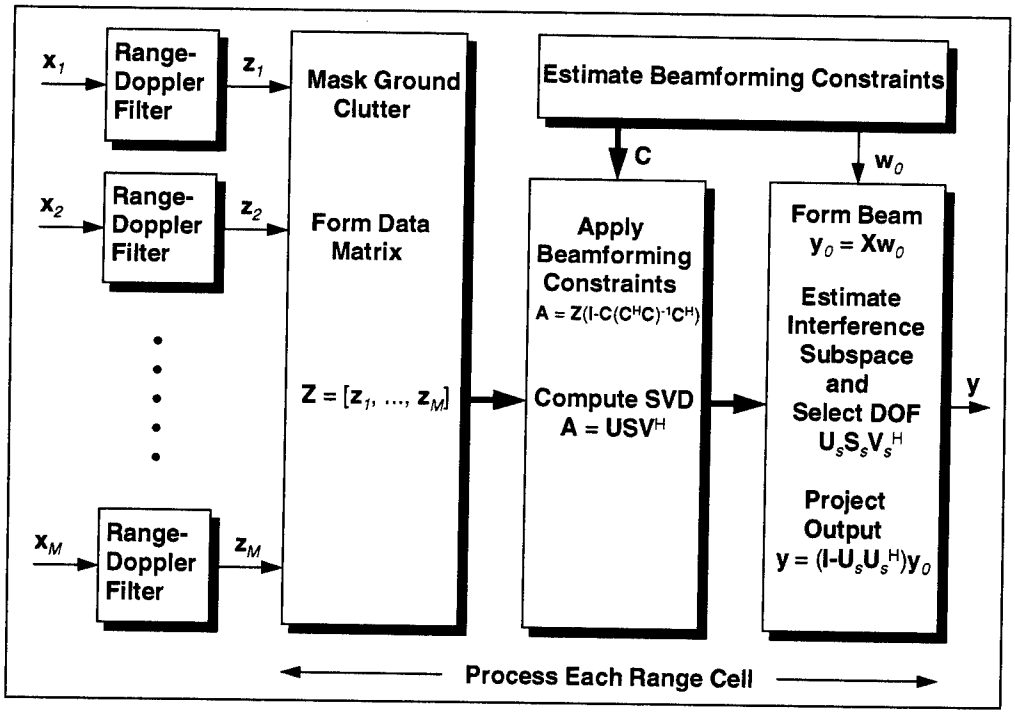


Figure 2. Principal Component Inverse Adaptive Sidelobe Canceller

similarly estimated in adjacent range cells. This procedure was repeated as a sliding window of Doppler cells was scanned across the Doppler spectrum of the range cell to detect Doppler shifted signals and observe the adaptive spatial cancellation of the background noise and Doppler-spread ionospheric clutter. Since the illuminated ground clutter spectra are easily recognised, not important to the detection of Doppler-shifted signals and an unecessary burden for the adaptive spatial processor, the Doppler cells corresponding to the ground clutter were removed before forming the data matrix to be processed and analyzed. The samples to be excised in each range cell were determined by analysis of the principal components of the unexcised data, easily revealing the spectrum of the main ground clutter component.

Adaptive Processing Results

An example of the results of applying this method of adaptive spatial processing to the data is presented next. We applied the self-focussing technique to range cell 166, and a few adjacent range cells, choosing a Doppler band near 2 Hz, to estimate the beamsteering weight vector and linear constraint matrix. The estimated beamforming constraints were then used in an adaptive beamformer applied to all of the range cells in the coherent processing interval. Figure 3 shows the unadapted and spatially adapted Doppler spectra for range cell 166 with the beamsteering weight vector and linear constraint matrix estimated from a band of Doppler cells centered near 2 Hz. Adaptive results are shown for 64 DOF and 87 DOF. Figure 4 displays the unadapted range-Doppler map for 128 range cells for this example (129-256) and figure 5 displays the range-Doppler map after spatial adaptation using 87 DOF. For this example the increase in SINR was measured as 23.4 dB for 64 DOF and 30.3 dB for 87 DOF. Other applications of this technique to the experimental radar data produced similar results. It is evident from the example that substantial suppression of the clutter has been obtained without loss of signal, facilitated by the use of constraints.

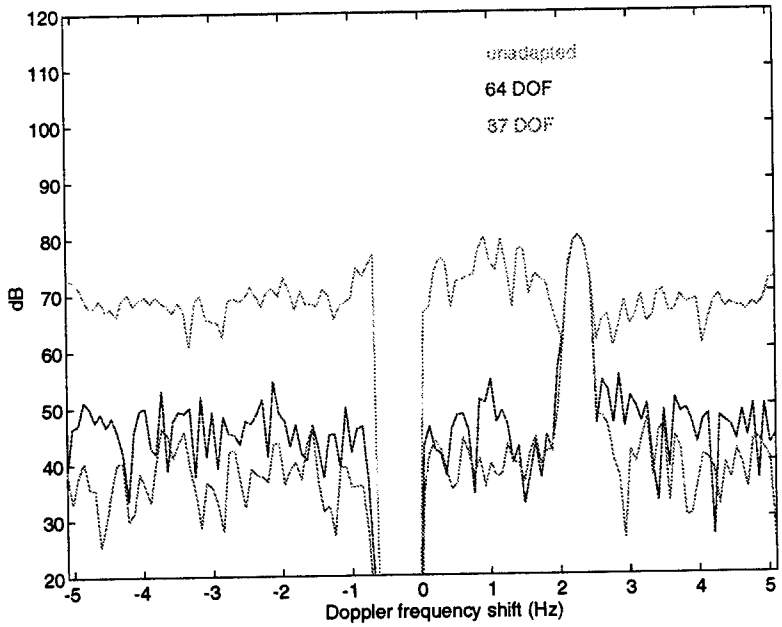


Figure 3. Doppler Spectra for Range Cell 166

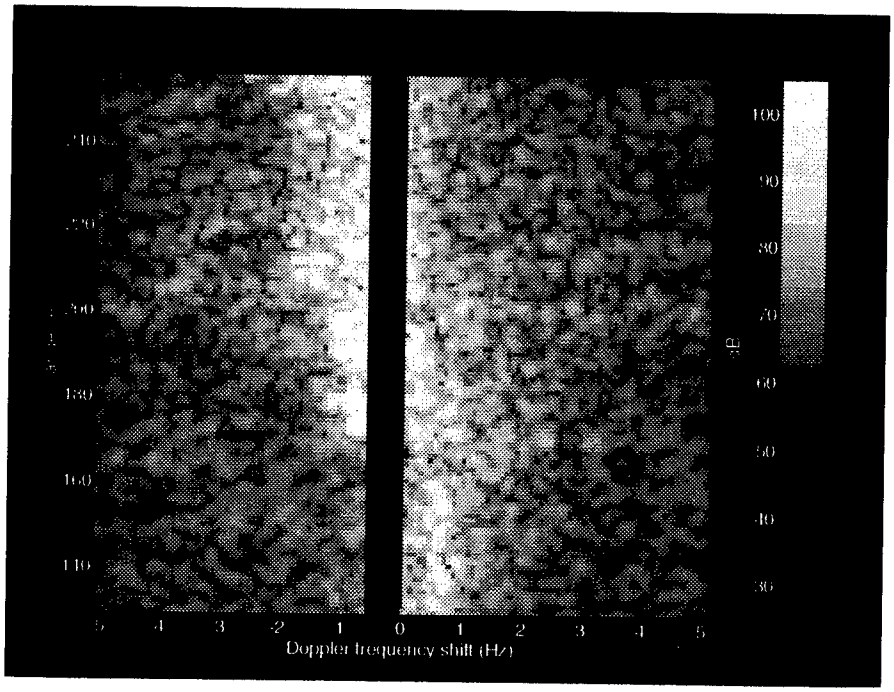


Figure 4. Unadapted Range-Doppler Map

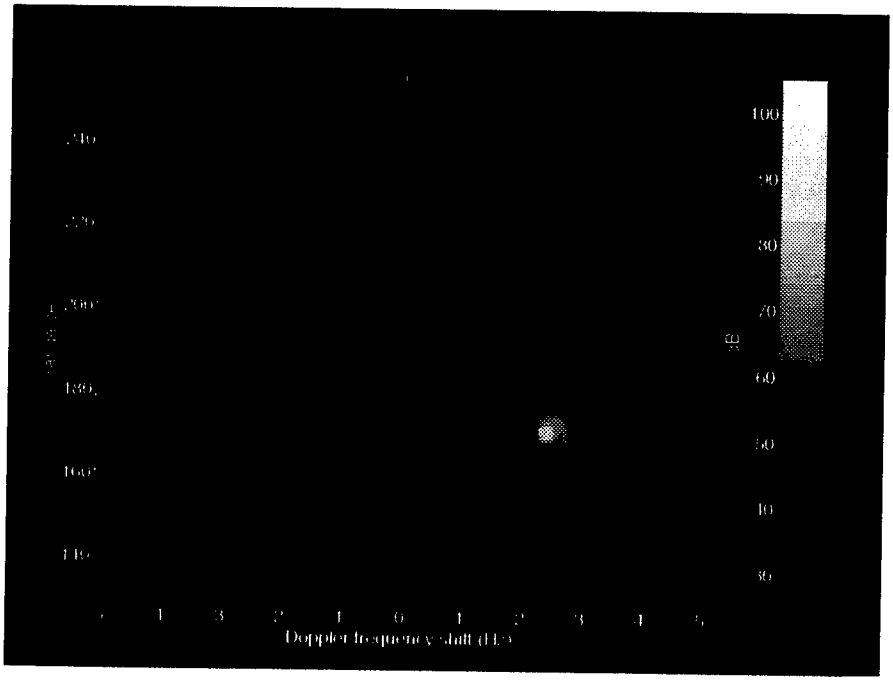


Figure 5. Spatially Adapted Range-Doppler Map

Speech Enhancement through Nonlinear Adaptive Source Separation Methods

Nikos Doukas, Tania Stathaki and Patrick Naylor

Signal Processing Section, Dept of Electrical Engineering,Imperial College,
Exhibition Road, London SW7 2BT, United Kingdom.
e-mail: n.doukas@ic.ac.uk

Abstract

In this paper, a new method that exploits the ideas of independent source separation in the context of Speech Enhancement in single sensor signals, is developed and tested in various situations. The channel distortions of the two sensor case are artificially reproduced by suitable linear and nonlinear filters. Separation is implemented via a Lagrange neural network. Results on speech signals are shown.

1. Introduction

Recently there has been considerable work on the problem of source separation (see e.g [7], [8], [10]). In its simplest form the problem is given a linear mixture of signals (sources), to separate the contribution of each of the sources present assuming they are independent. Other interesting work in the area has been presented in [3], [6] and [9]. Previous research has focused mainly on multisensor approaches to the problem where different mixtures of the source signals arrive at each one of the sensors. Such approaches are difficult to use in practice, because of the increased complexity imposed by the presence of an array. The approach of our work is to produce estimates of the signals present using just one sensor. The different distortions normally suffered by the signals in the channel are modelled locally by suitably filtering the received signal. A Lagrange minimisation problem is formed to be solved by a Lagrange programming neural network ([11]). The results of the application of the method on contaminated speech signal are included.

2. The Source Separation Problem

Consider two independent signals x_1 and x_2 propagating in the same medium and two sensors, each receiving a different mixture of the two signals, i.e. $y_1 = a_{11}x_1 + a_{12}x_2$ and $y_2 = a_{21}x_1 + a_{22}x_2$.

It can then be shown that the initial signals can be recovered as ([1]):

$$s_1 = b_1 x_1 = w_{11}y_1 + w_{12}y_2 \quad (1)$$

$$s_2 = b_2 x_2 = w_{21}y_1 + w_{22}y_2 \quad (2)$$

where b_1 and b_2 are constant gains and the w_{ij} depend only on the a_{ij}s. This recovery may be performed provided that $a_{11}a_{22} - a_{12}a_{21} \neq 0$.

Since the a_{ij}s are of course not known, the w_{ij}s must be estimated through some kind of optimisation procedure. The two signals are by assumption idependent, zero mean implying that their odd powered cross moments are zero. This fact can be exploited for this optimisation. Examples of ways to estimate these moments are given in [1] and [7]. The method of estimation used in our work will be presented later on in this paper.

A typical block diagram of a source separation apparatus, is given in figure 1. The first part of the circuit (marked as 'CHANNEL') reproduces the distortions that would normally be suffered by the signals in the channel. The second part (marked as 'NEURAL NET') is the one that recovers the mixed signals. The weights w_{ij} are controlled by some adaptive mechanism, specific to each method.

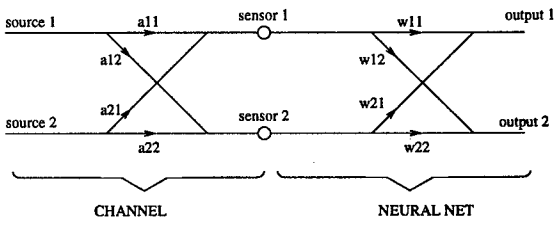


Figure 1: Standard source separation setup

0-8186-7576-4/96 $5.00

3. The single sensor case

The modified arrangement for the new method is depicted in figure 2. In our case there is only one signal available, namely the noise contaminated signal (marked as 'sensor'). A two sensor simulation can be made in such a manner that the distortions that the signal would undergo when travelling through a channel are modelled by passing it through two different filters (shown in figure 2 as $H1$ and $H2$). Some guidelines for chosing these filters are given later in this paper. This produces two pseudo-sensor signals, shown as "sensor 1" and "sensor 2". These two signals are then used as substitutes for the signals from the two sensors.

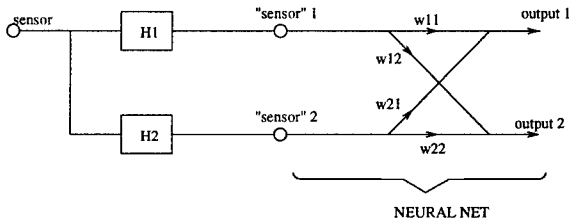


Figure 2: Block diagram of the setup used for the new method

The adaptation mechanism is further assisted by the introduction of constraints. A constrained optimisation problem is set up and its solution implemented through the use of Lagrange Programming Neural Networks. This type of neural networks are based on the Lagrange minimisation theory. They were chosen because they permit the introduction of constraints, but exhibit further advantages in terms of speed of convergence, ability to readapt and good stability. Details about them are given in [11] and [4].

It has already been mentioned that odd power cross moments of the outputs must be zero, and the function to be minimised is therefore taken to be

$$J = \sum_{i,j} \left(E[s_1^{2i+1} s_2^{2j+1}] \right)^2 \qquad (3)$$

subject to the constraint that $s_1 + s_2 = y$ where y is the received signal. This gives the following Lagrange function to be minimised:

$$J = \sum_{i,j} \left(E[s_1^{2i+1} s_2^{2j+1}] \right)^2 + \lambda(s_1 + s_2 - y) \qquad (4)$$

The update equations for w_{ij} and λ can be obtained by using (1) and (refequ2) and differentiating the above expression. A steepest descent adaptation is then performed.

In this study i and j are restricted so that:

$$(i, j) \in \{0, 1\}^2$$

For reasons of simplicity only the two source case is considered.

4. Implementation Issues

The received signal which is assumed to be a linear mixture of the two source signals is passed through two separate filters. The two outputs are used in our setup in the manner of a standard source separation problem ([5], [7]). These filters should not have high stopband attenuation so that both the outputs convey information about all frequency components of the signals. Further investigations as to the choice of these filters are currently under way.

It can be easily seen that the following modification to the objective function, reduces the computational load considerably:

$$J = \left(\sum_{i,j} E[s_1^{2i+1} s_2^{2j+1}] \right)^2 + \lambda(s_1 + s_2 - y) \qquad (5)$$

Possible further implications of this modification are currently under investigation.

Several alternative methods for estimating the cross moments of the signals have been investigated. Clearly, since we are dealing with higher order moments, a large number of samples must be used for reducing the variance of the estimation. The fact however that the signals can not be assumed stationary poses a limit on the number of past samples that can be meaningfully used in the estimation. For these reasons the following recursive formula was used:

$$\left(E[s_1^i s_2^j] \right)_n = \phi \times \left(E[s_1^i s_2^j] \right)_{n-1} + (1-\phi) \times s_{1,n}^i s_{2,n}^j \qquad (6)$$

where

$$\left(E[s_1^i s_2^j] \right)_n$$

is the estimate for the moment at time n, and $s_{k,n}$ is the value of signal s_k at time n and ϕ is a forgetting factor. Equation (6) provides an unbiased estimate for the moments. Clearly it produces good estimates of the value of the moments, since a large number of samples is involved. Additionally, with a suitable choice of ϕ, it can quickly respond to changes in the statistics of the signal.

A variable gain adaptation was used to give better stability and eliminate oscillations of the weights in a dynamic Lagrange neural network realisation. For stationary environments the adaptation gain modification is taken to follow the rule:

$$\mu = \mu_0 \frac{1}{(iteration\ number)^\beta} \tag{7}$$

where β is a positive constant. Typically $0 \leq \beta \leq 2$. This update method is used in current literature ([1]). It gives an initial, near optimal solution quickly and then convergences with small missadjustment.

Solutions for non-stationary cases are currently being explored.

5. Results

Convergence is fast and due to the variable gain there are no weight oscillations after the final values are reached. Sample convergence curves for the weights of the neural network can be seen in figure 3.

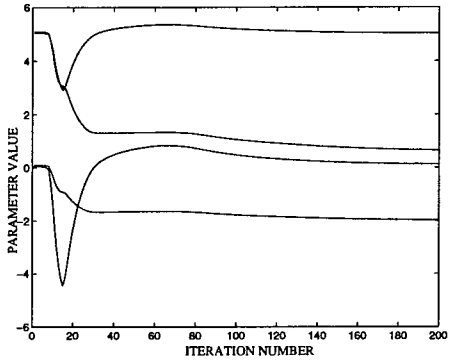


Figure 3: Sample convergence curve for the weights

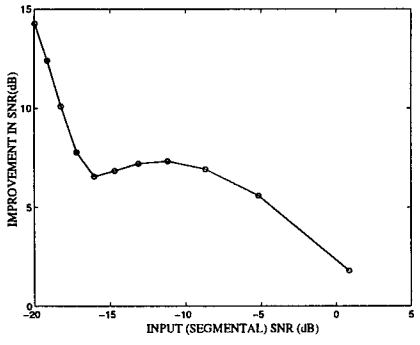
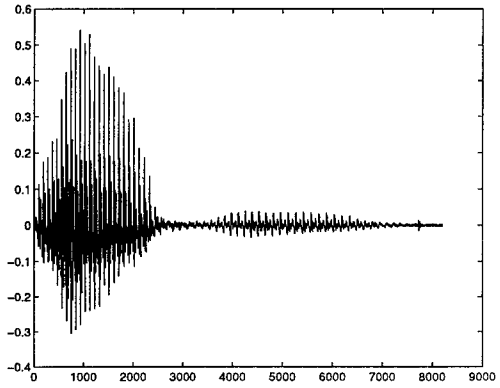
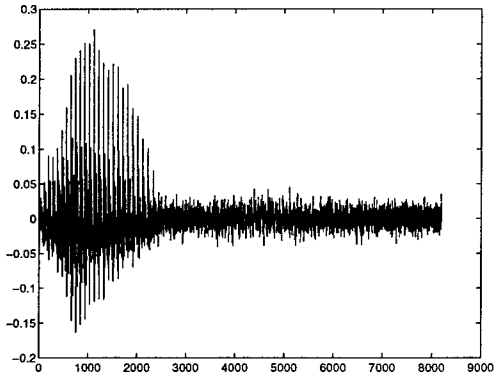


Figure 4: Improvement in SNR after processing versus input SNR (both measured as segmental SNR)

The tests were performed on single sinusoid plus white, zero–mean, gaussian noise, speech plus sinusoid and speech plus white, zero–mean, gaussian noise. Sample results for speech plus white noise, can be seen in figures 5 (the original and the contaminated signals) and 6 (the reconstructed signals).



(a)

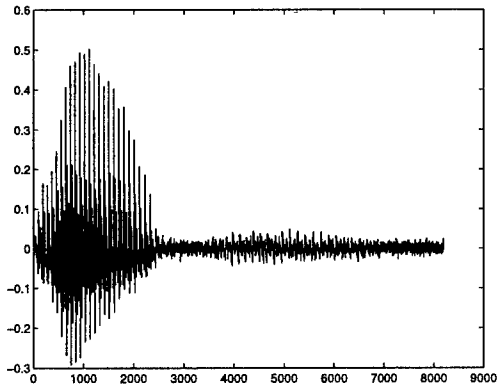


(b)

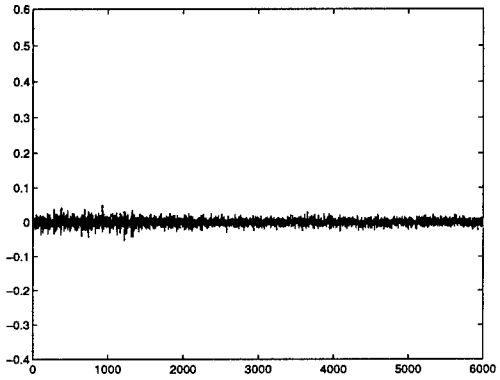
Figure 5: Example of the application of the method: Speech plus White Gaussian Noise. a: original signal, b: contaminated signal

The graphs clearly show a definite improvement of the reproduction of the different signals in each case. The outputs are acoustically close to their original versions. The improvement in SNR versus input SNR is given in figure 4. It can be seen that the proposed method gives good results in very adverse conditions. Note that the SNR displayed is a segmental SNR.

Tests for removing sinusoidal interference from speech were performed. For an input SNR of -3.7 dB, the output SNR was 16.12 dB for a fixed frequency of the sine



(a)



(b)

Figure 6: Example of the application of the method: Speech plus White Gaussian Noise. a: reconstructed signal ,b: reconstructed noise

wave (improvement 19.81 dB) and 12.2 db for a slowly varying one (improvement 15.9 db).

6. Conclusions

A new method to enhance signals, based on source separation techniques is presented. The initial results obtained are quite promising. Several improvements are possible in a variety of directions, for example in using different filters and different objective functions. The method is potentially useful in many applications to other signal processing problems, such as for example Voice Activity Detection. Research is currently under way to explore the fundamental parameters that influence this approach in a decisive manner and to determine the limits of its applicability. Further development of this work is reported in [2].

7. References

[1] A. Cichoki and R. Unbehauen. *Neural Networks for Optimization and Signal Processing.* Wiley, 1993.

[2] N. Doukas, P. Naylor, and T. Stathaki. A single sensor source separation approach to noise reduction. In *CESA 96 IMACS Multiconference*, 1996.

[3] F. Ehrmann, R. Le Bouquin-Jeannes, and G. Faucon. Optimisation of a two-sensor noise reduction technique. *IEEE Signal Processing Letters*, 2(6):108–110, June 1995.

[4] E. Ertin. Lagrange programming neural networks for visual reconstruction. Master's thesis, Imperial College, University of London, September 1993.

[5] C. Jutten and J. Herault. Blind separation of sources, part 1: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1–10, July 1991.

[6] K. Matsuoka and M. Kawamoto. A neural net for blind separation of nonstationary signal sources. In *IEEE International Conference on Neural Networks - IEEE World Congress on Artificial Intelligence*, volume 1, pages 221–226, 1994.

[7] P.Common, C. Jutten, and J. Herault. Blind separation of sources, part 2: Problems statement. *Signal Processing*, 24(1):11–20, July 1991.

[8] E. Sorouchyari. Blind separation of sources, part 3: Stability analysis. *Signal Processing*, 24(1):21–30, July 1991.

[9] E. Weinstein, M. Feder, and A. Oppenheim. Multi-channel signal separation by decorrelation. *IEEE Transactions on Speech and Audio Processing*, 1(4):405–413, October 1993.

[10] D. Yellin and E. Weinstein. Criteria for multichannel signal separation. *IEEE Transactions on Signal Processing*, 42(8):2158–68, August 1994.

[11] S. Zhang and A. G. Constantinides. Lagrange programming neural networks. *IEEE Transactions on Circuits and Systems*, 39(7):441–52, July 1992.

Automatic Segmentation of Piecewise Constant Signal by Hidden Markov Models *

Jong-Kae Fwu and Petar M. Djurić
Department of Electrical Engineering
State University of New York at Stony Brook
Stony Brook, New York 11794-2350
fwu@sbee.sunysb.edu, djuric@sbee.sunysb.edu

Abstract

We propose an automatic signal segmentation algorithm for piecewise constant signals, which is based on Hidden Markov Models (HMM). It segments the observed data without the need for training data and initial conditions. One of the problems for automatic segmentation using HMM models is the determination of their number of states. In this paper, the number of states is estimated according to a maximum a posteriori (MAP) criterion. The proposed algorithm is iterative. Its initial conditions are chosen by a Tree-Structure technique, which is completely data driven. The segmentation is further improved by the multiscale technique. The performance is evaluated by computer simulations.

1. Introduction

Signal segmentation is an important problem that occurs in many applications including speech recognition, biomedical signal processing, and pattern analysis. The commonly used maximum likelihood segmentation tends to have poor performance, since it ignores the temporal correlation among the samples. To include temporal correlation, we use Hidden Markov Models (HMM's) [6]. An optimal segmentation by HMM's can be achieved by the well known Viterbi algorithm [6]. However, this algorithm assumes that the number of states is known and requires sufficient data for training the estimators to achieve good results [6]. Therefore, it is not quite practical when it is applied to data for which such information is unavailable.

In this paper, we propose a novel algorithm that can circumvent these problems. We analyze piecewise constant signals whose levels (states) and number of states are unknown. The number of states and the best segmentation are determined by a maximum a posteriori (MAP) criterion. The proposed algorithm has modest computational requirements even when it is extended to two dimensional data.

*This work was supported by the National Science Foundation under Award No. MIP-9506743.

2. Problem Formulation

Let $\mathbf{y}^T = [y_1 y_2 \cdots y_N]$ be an observed data vector of N samples comprised of a signal embedded in additive noise. Let $\mathbf{x}^T = [x_1 x_2 \cdots x_N]$ be the unobservable vector of signal states which is a realization of an m-state HMM process, i.e. $x_i \in \{1, 2, \cdots, m\}$ for $i = 1, 2, \cdots, N$. Also, suppose that the observed data can be modeled by

$$y_i = g(x_i) + w_i, \quad i = 1, 2, \cdots, N, \tag{1}$$

where $g(x_i)$ is a function that maps the underlying state x_i to a constant μ_{x_i}. The vector $\mathbf{w}^T = [w_1 w_2 \cdots w_N]$ represents noise, and its elements are independently distributed Gaussian random variables with zero mean and unknown variance $\sigma^2_{x_i}$. We model the underlying state vector $\mathbf{x}$ as a first order HMM, so the probability mass function of the vector $\mathbf{x}$ is given by

$$p(\mathbf{x}) = p(x_1) \prod_{i=2}^{N} p(x_i | x_{i-1}) \tag{2}$$

where $p(x_1)$ denotes the state probability of the first sample and $p(x_i|x_{i-1})$ is the state transition probability. The density of $\mathbf{y}$, given the underlying hidden states $\mathbf{x}$, is

$$\begin{aligned} f(\mathbf{y}|\mathbf{x}) &= \prod_{i=1}^{N} f(y_i|x_i) \\ &= \prod_{i=1}^{N} \frac{1}{(2\pi\sigma^2_{x_i})^{\frac{1}{2}}} \exp\left(-\frac{1}{2\sigma^2_{x_i}}(y_i - \mu_{x_i})^2\right). \end{aligned}$$

0-8186-7576-4/96 $5.00 © 1996 IEEE

The number of states m, the parameters associated with each state, $(\mu_{x_i}, \sigma^2_{x_i})$, $p(x_1)$, and the state transition probability $p(x_i|x_{i-1})$, are *unknown*.

The problem is to determine the number of states m, estimate all the unknown parameters, and label the observed data with one of the m states.

3. The Segmentation Approach

For a given number of states $m = k$, we would like to estimate the underlying state vector $\mathbf{x}_k$, where $x_i \in \{1, 2, \cdots, k\}$. From Bayes' theorem we can obtain the posterior probability of the state vector $\mathbf{x}_k$, that is,

$$p(\mathbf{x}_k|\mathbf{y}) = \frac{f(\mathbf{y}|\mathbf{x}_k)p(\mathbf{x}_k)}{f(\mathbf{y})}. \tag{3}$$

Since we would like to adopt as an estimate the most probable value of $\mathbf{x}_k$ given the data $\mathbf{y}$, we write

$$\hat{\mathbf{x}}_k = \arg\max_{\mathbf{x}_k}\{f(\mathbf{y}|\mathbf{x}_k)f(\mathbf{x}_k)\}. \tag{4}$$

Note that $f(\mathbf{y})$ is dropped from (4) because it is not a function of $\mathbf{x}_k$.

To develop an efficient algorithm that searches for the optimal solution, we iteratively optimize individually the x_i, $i = 1, 2, \cdots, N$, according to

$$\hat{x}_i = \arg\max_{x_i \in \{1,2,\cdots,k\}}\{f(\mathbf{y}|\mathbf{x}_k)p(\mathbf{x}_k)\}. \tag{5}$$

When we eliminate the terms which are not functions of x_i, (5) simplifies to

$$\hat{x}_i = \arg\max_{x_i \in \{1,2,\cdots,k\}}\{f(y_i|x_i)p(x_i|x_{i-1})p(x_{i+1}|x_i)\}. \tag{6}$$

This is iteratively solved according to

$$\hat{x}_i^{(j+1)} = \arg\max_{x_i \in \{1,2,\cdots,k\}}\{f(y_i|x_i)p(x_i|x_{i-1}^{(j+1)})p(x_{i+1}^{(j)}|x_i)\} \tag{7}$$

where j denotes iteration, and $p(x_i|x_{i-1}^{(j+1)})$ and $p(x_{i+1}^{(j)}|x_i)$ are the estimates of the transition probabilities from x_{i-1} to x_i and x_i to x_{i+1}, respectively. The optimization of (7) can be implemented, for example by the Iterated Conditional Modes (ICM) algorithm [1]. Since this is an iterative techniques, the initial conditions play an extremely important role, and therefore they need to be handled with great care [4], [5]. In our approach we choose them by a recently developed Tree Structure (TS) scheme [4].

In (7), the labeling of x_i depends only on the x_{i-1} and x_{i+1}. At low signal-to-noise-ratios (SNR's), the initial states may contain many mis-labeled data samples, which could lead to poor results. To overcome this shortcoming, we use a multi-scale technique similar to the one proposed in [2]. We refer to it as Multi-scale HMM (MS-HMM). It is composed of a series of segmentations progressing from coarse to fine scale. This is implemented as follows.

Let the observed data and the underlying labels at scale s, $s = 0, 1, 2, \cdots, s_{max}$, be denoted as $\mathbf{y}_{<s>}$ and $\mathbf{x}_{<s>}$, respectively. The initial label sequence at scale s, $\mathbf{x}_{<s>}$ is obtained from the estimated sequence $\mathbf{x}_{<s+1>}$. The number of data at scale $s+1$ is only half of the number at scale s. For example, $\mathbf{y}_{<0>} = \mathbf{y}$ and $\mathbf{x}_{<0>} = \mathbf{x}$ are the observed data and their labels at the finest (original) scale, respectively. Each sample at the first scale, $\mathbf{y}_{<1>}$ and $\mathbf{x}_{<1>}$, corresponds to two points in the original scale $s = 0$. For the observed data we use $y_{i<1>} = \frac{1}{2}(y_{2i-1<0>} + y_{2i<0>})$, and for the labels $x_{2i-1<0>} = x_{2i<0>} = x_{i<1>}$, for $i = 1, 2, \cdots, \frac{N}{2}$. Similarly, each sample at scale $s = 2$ corresponds to two samples at scale $s = 1$, and so on. Note that, at coarser scales, the noise in the data $\mathbf{y}_{<s>}$ is decreased due to averaging.

4. The MAP Solution for Number of States

In general, the number of states is also unknown. Our objective now is to obtain a criterion for choosing this number. From Bayes' theorem we have

$$p(\mathbf{x}_k|\mathbf{y}) = \frac{f(\mathbf{y}|\mathbf{x}_k)p(\mathbf{x}_k)p(k)}{f(\mathbf{y})} \tag{8}$$

where $f(\mathbf{y}|\mathbf{x}_k)$ is the likelihood function given the hidden states and the number of states, $p(\mathbf{x}_k)$ is the probability mass function of $\mathbf{x}$ given k, and $p(k)$ is the a priori probability mass function of the model with k states. If we let $p(k)$ be uniform, the MAP solution of (8) becomes

$$(\hat{\mathbf{x}}_k) = \arg\max_{\mathbf{x}_k}\{f(\mathbf{y}|\mathbf{x}_k)p(\mathbf{x}_k)\}. \tag{9}$$

Given the number of states k, we can find the underlying labels $\hat{\mathbf{x}}_k$ for $k = 1, 2, \cdots, k_{max}$ by the proposed algorithm. Once we obtain the $\hat{\mathbf{x}}_k$'s for various k's, the number of states is selected according to

$$\hat{k} = \arg\max_k\{f(\mathbf{y}|\hat{\mathbf{x}}_k)p(\hat{\mathbf{x}}_k)\} \tag{10}$$

$$= \arg\min_k\{-\ln f(\mathbf{y}|\hat{\mathbf{x}}_k) - \ln p(\hat{\mathbf{x}}_k)\}. \tag{11}$$

To determine $\ln f(\mathbf{y}|\hat{\mathbf{x}}_k)$ we use

$$f(\mathbf{y}|\hat{\mathbf{x}}_k) = \int_{\Theta_k} f(\mathbf{y}|\hat{\mathbf{x}}_k, \boldsymbol{\theta}_k)f(\boldsymbol{\theta}_k)d\boldsymbol{\theta}_k \tag{12}$$

where $\boldsymbol{\theta}_k = [\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \cdots, \mu_k, \sigma_k^2]^T$ is the parameter vector associated with all the states. If we Taylor

expand $\ln f(\mathbf{y}|\hat{\mathbf{x}}_k)$ around the maximum likelihood estimate of $\boldsymbol{\theta}_k$, $\hat{\boldsymbol{\theta}}_k$, we obtain

$$\ln f(\mathbf{y}|\hat{\mathbf{x}}_k) \simeq \ln f(\mathbf{y}|\hat{\mathbf{x}}_k, \hat{\boldsymbol{\theta}}_k) - \frac{1}{2}(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k)^T \mathcal{H}_k (\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k) \tag{13}$$

where $\mathcal{H}_k$ is the Hessian of $-\ln f(\mathbf{y}|\hat{\mathbf{x}}_k, \boldsymbol{\theta}_k)$ evaluated at $\hat{\boldsymbol{\theta}}_k$. By plugging (13) into (12), we obtain

$$\ln f(\mathbf{y}|\hat{\mathbf{x}}_k) \simeq \ln f(\mathbf{y}|\hat{\mathbf{x}}_k, \hat{\boldsymbol{\theta}}_k) + \frac{1}{2}\ln|\mathcal{H}_k|. \tag{14}$$

Using a similar approach, we can show that

$$\ln p(\hat{\mathbf{x}}_k) \simeq \ln p(\hat{\mathbf{x}}_k|\hat{\boldsymbol{\phi}}_k) + \frac{1}{2}\ln|\mathcal{H}'_k| \tag{15}$$

where $\boldsymbol{\phi}_k$ is the parameter vector of the state transition matrix for k states, and $\mathcal{H}'_k$ is the Hessian of $-\ln p(\mathbf{x}_k|\boldsymbol{\phi}_k)$ evaluated at $\hat{\boldsymbol{\phi}}_k$.

Based on (11), (14), (15), and after some algebra, the MAP criterion becomes

$$\hat{k}_{MAP} = \arg\min_k \{F_y(k) + F_x(k)\} \tag{16}$$

where

$$F_y(k) = -\ln f(\mathbf{y}|\hat{\mathbf{x}}_k, \hat{\boldsymbol{\theta}}_k) + \sum_{i=1}^{k} \frac{1}{2}\ln n_i \tag{17}$$

and

$$F_x(k) = -\ln p(\hat{\mathbf{x}}_k|\hat{\boldsymbol{\phi}}_k) + \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{1}{2}\ln\left(\frac{n_i^2}{n_{ij}}\right) \text{ for } n_{ij} > 0 \tag{18}$$

where n_i is the number of samples that are in the $i-$th state, and n_{ij} is the number of jumps from the $i-$th to $j-$th state.

This criterion is a penalized maximum likelihood with a simple interpretation. $F_y(\cdot)$ has two terms, one corresponding to the fitting error of the applied model and the other to the penalty for overparameterization. $F_x(\cdot)$ on the other hand, reflects the temporal constraints imposed by the HMM and the penalty for imposing the constraints.

5. Implementation of the Proposed Criterion

We implement the procedure by using three different scales. We start with the assumption that the number of segments is equal to one, that is $k = 1$, and initialize all the underlying states to 1 for $s = 0, 1, 2$. We then evaluate $\hat{\mu}_1$, which is simply the sample mean of all the data samples, and $\hat{\sigma}_1^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\mu}_1)^2$.

We continue by setting $k = 2$, and using $\hat{\mu}_1 + \epsilon$ and $\hat{\mu}_1 - \epsilon$, where ϵ is some small positive number, as initial conditions to estimate $\hat{\mathbf{x}}_{2<2>}$ according to

$$x_{i<2>} = l \text{ if } d(y_{i<2>}, \hat{\mu}_l) \leq d(y_{i<2>}, \hat{\mu}_{l'}), \quad l \neq l', \tag{19}$$

where $d(\cdot)$ is the Euclidean distance. Then we update $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ by

$$\hat{\mu}_l = \frac{1}{n_l}\sum_i y_{i<2>}, \quad x_{i<2>} = l \tag{20}$$

$$\hat{\sigma}_l^2 = \frac{1}{n_l}\sum_i (y_{i<2>} - \mu_l)^2, \quad x_{i<2>} = l \tag{21}$$

where $l = 1, 2$, and n_l is the number of samples in the state l. When the labeling process converges, we estimate the transition probability $p(x_{i<2>}|x_{i-1<2>})$ by

$$p(x_{i<2>}|x_{i-1<2>}) = \frac{n_{x_{i<2>}x_{i-1<2>}}}{n_{x_{i-1<2>}}} \tag{22}$$

where $x_{i<2>}, x_{i-1<2>} \in \{1, 2\}$, $n_{x_{i-1<2>}}$ is the number of samples that are in state $x_{i-1<2>}$, and $n_{x_{i<2>}x_{i-1<2>}}$ is the number of jumps from state $x_{i-1<2>}$ to state $x_{i<2>}$. As initial state probabilities $p(x_{1<2>})$, we use the uniform distribution. Now we are ready to start with the iterative process. We iteratively label the state vector $\mathbf{x}_{k<2>}$ using (7) and update all the parameters until the process converges. We then start the next scale $s = 1$, and as its initial parameters we use the final estimates from the previous scale. After we finish the scale $s = 1$, we repeat the procedure for $s = 0$.

Once the segmentation with $k = 2$ is completed, we set $k = 3$. We have two sets of initial conditions $\{\hat{\mu}_1^{(0)} = \hat{\mu}_1 + \epsilon, \hat{\mu}_2^{(0)} = \hat{\mu}_1 - \epsilon, \hat{\mu}_3^{(0)} = \hat{\mu}_2\}$ and $\{\hat{\mu}_1^{(0)} = \hat{\mu}_1, \hat{\mu}_2^{(0)} = \hat{\mu}_2 + \epsilon, \hat{\mu}_3^{(0)} = \hat{\mu}_2 - \epsilon\}$. We then apply the same procedure to each of the initial conditions, evaluate the criterion (16) based on the segmentation results and choose the one with smaller criterion value. The results of the last step are used as initial conditions for $k = 4$. We continue by increasing k and applying the same steps until $k = k_{max}$. Finally, we choose the k that minimizes (16).

6. Simulation Results

To verify the performance of the MS-HMM, we applied it to synthesized and real data. The first experiment was on a synthesized data record with 500 samples, three different states, and eight transitions. The signal intensities of these states were 100, 120, and 140. The SNR is defined by $\min\{\frac{\Delta}{\sigma}\}$, where Δ is the intensity difference of the transitions. The SNR varied between 1 and 5, and there were 100 trials for each different SNR. Figure 2 represents the synthesized noisy

data for SNR=2. The number of different states is equal to three, whereas the maximum number of possible states is equal to five. The segmentation results are shown in Table 1. Figure 3 is the histogram of the detected state transitions for 100 trials. The peaks of the histogram are at the correct locations of state transitions.

In the second experiment we applied the criterion to real patch clamping data, which are used in the study of ion permeation mechanisms in biological membranes [3]. Figure 4 displays the patch clamp data to which we applied our segmentation algorithm. The result is shown in Figure 5 with 8 states detected.

References

[1] J. Besag, "On the Statistical Analysis of Dirty Pictures," *J. R. Stat. Soc.*, ser. B 48, no. 3, pp. 259-302, 1986.

[2] C. Bouman and M. Shapiro, "A Multiscale Random Field Model for Bayesian Image Segmentation," *IEEE Trans. Image Processing*, vol. 3, pp. 162-177, 1994.

[3] P. M. Djurić, J.K. Fwu, S. Jovanovic, and K. Lynn, "On the Processing of Piecewise-Constant Signals by Hierarchical Models With Application to Signal Ion Channel Currents," to be published in the *Proceedings of ICASSP 1996.*

[4] J. K. Fwu and P. M. Djurić, "Unsupervised Vector Image Segmentation by a Tree Structure - ICM Algorithm," submitted to *IEEE Trans. Medical Imaging.*

[5] Z. Kato, J. Zerubia, and M. Berthod, "Unsupervised Adaptive Image Segmentation," *Proceedings of ICASSP*, pp. 2399-2402, 1995.

[6] L. R. Rabiner "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286, 1989.

SNR/K	1	2	3	4	5
1	0	83	17	0	0
2	0	0	100	0	0
3	0	0	100	0	0
4	0	0	100	0	0
5	0	0	100	0	0

Table 1. The entries represent the number of times $\hat{k}$ states were detected in 100 trials.

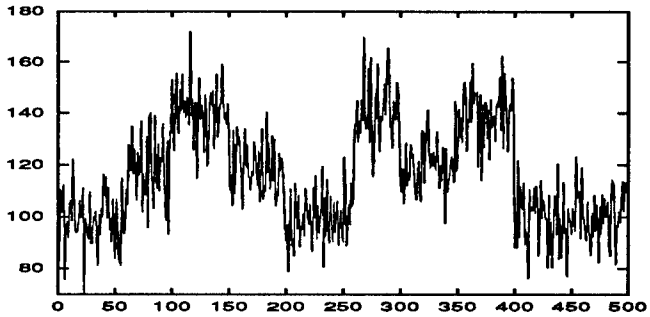


Figure 1. A realization with 3 classes and 8 transitions with SNR=2.

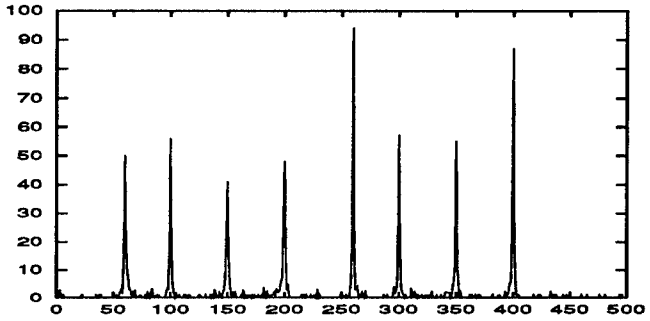


Figure 2. Histogram of the detected edges from 100 trials with SNR=2.

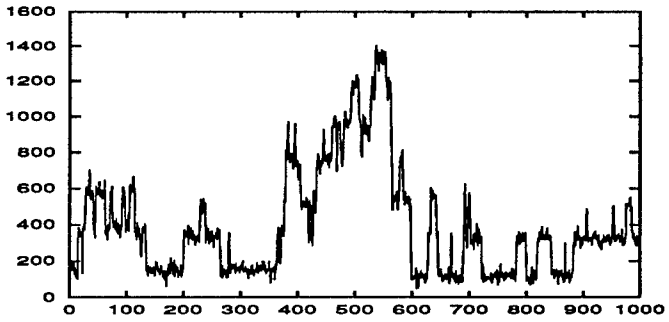


Figure 3. Real data with unknown number of classes.

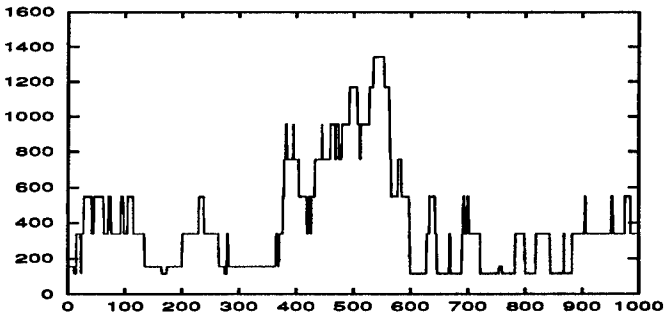


Figure 4. Estimated signal from the data in Figure 3 with 8 states detected.

Position Estimation of Electrocortical Generators

D. Klimovski[1], A.L. Cricenti[2], G.K. Egan[1] and A.A. Sergejew[3]
[1]Monash University - Electrical and Computer Systems Engineering Department
[2]Swinburne University of Technology - School of Biophysical Sciences and Electrical Engineering
[3]Swinburne Centre for Applied Neurosciences (SCAN)

Abstract

Using subspace techniques we present simulated results investigating variations in the geometry of the sensor array for EEG measurements. We have shown that the performance of the subspace techniques degrades as the sources are brought closer to the array. This degradation can be counteracted by changing the curvature of the array. An optimum array curvature exists which exhibits best detection performance for a given angle of arrival. We also present preliminary results applying subspace techniques to a sample of real EEG data.

Introduction

Recent work on the estimation of the direction of arrival (DOA) problem uses the subspace approach to determine the angular location of multiple emitters, for example[1]. These approaches include MUSIC[2], MLM[3] and J&D[4], where these methods have been traditionally applied to the fields of RADAR, SONAR and seismology. The application area of these techniques can be extended to other fields such as biomedical problems[5]. One such application is that of estimating the position of electrocortical generators in the brain from the electroencephalogram (EEG). The estimation problem is complicated by a number of factors:

- The geometry of the array.
- The source distance from the array .
- Noise present in the system.

The work described here investigates the application of subspace techniques to the processing of signals where two of the above factors are varied. This simulates a simplified environment similar to that in which EEG signals are measured. Using subspace techniques on real EEG data the aim is to estimate the position of possible electrocortical generators in the brain.

This paper consists of five sections. Section one reviews some of the work in the area of subspace techniques for solving spectral estimation problems. Section two describes the area of application to both EEG and driven EEG. The results of simulations using subspace algorithms, and a discussion of the limitations of these algorithms under the conditions outlined, are presented in section three. Section four discusses the results of the application of subspace techniques to the EEG context. The final section offers conclusions and comments on possible further work in this area.

1. Signal Subspace Methods

This section briefly reviews signal subspace methods. These methods are primarily derived from the covariance matrix which is constructed from incoming signal data. The covariance matrix can be broken down through the use of appropriate matrix properties and eigen-decomposition techniques into two subspaces, the signal subspace and a noise subspace [1]-[5]

Assuming a system model in which M far-field sources are viewed by N sensors ($N \geq M$). The sensors may exist in any configuration, for example a linear, circular or curved array. Consider the system

$$\mathbf{x} = \mathbf{V}\mathbf{s} + \mathbf{n}, \qquad (1)$$

where

$\mathbf{x} = [x(1), x(2),..., x(n)..., x(N)]^T$ represents the signals at the N sensors at any instant;

$\mathbf{s} = [s(1),s(2),...,s(m),...,s(M)]^T$ represents the plane wavefronts from the M sources;

$\mathbf{n} = [n(1), n(2),....,n(n),...,n(N)]^T$ represents the receiver noise contributions to the signals at the N sensors, and the (N x M) matrix $\mathbf{V}$ represents the response of the N sensors in the M signal directions. The matrix $\mathbf{V}$ cannot be specified until the directions to the sources are known, thus eqn. (1) cannot be solved directly.

The subspace methods use the covariance matrix of the system model which is defined as:

$$\mathbf{C} = E\{\mathbf{x}\mathbf{x}^H\} = E\{(\mathbf{V}\mathbf{s} + \mathbf{n})(\mathbf{V}\mathbf{s} + \mathbf{n})\}^H \qquad (2)$$

where E is the expectation operator and H is the hermitian operator. If the sources are uncorrelated with the receiver noise then:

0-8186-7576-4/96 $5.00

$$E\{\mathbf{ns}^H\} = E\{\mathbf{sn}^H\} = 0 \quad (3)$$

and if the noise is white Gaussian with variance σ^2

$$\mathbf{C} = \mathbf{VC_S V}^H + \mathbf{C_n} = \mathbf{VC_S V}^H + \sigma^2\mathbf{I}, \quad (4)$$

The direction finding (DF) problem in this system is the identification of the direction vectors

$$\mathbf{v_m}^T=[v(1,m),v(2,m),...,v(N,m)], \quad m = 1,...,M \quad (5)$$

Given that all the possible correlations between a pair of individual sensor signals exist in **C** it is possible through the use of eigen-decomposition techniques to decompose the complex space that **C** spans into two mutually orthogonal subspaces. These are the signal subspace and the noise subspace. It can be shown that either the signal or the noise subspace contain all the necessary information required to determine the number of sources and the direction of arrival[2].

Using the hermitian property of **C** we are able to transform it into a real diagonal matrix Λ using a unitary matrix **U** as shown below:

$$\mathbf{U}^H\mathbf{CU} = \Lambda \text{ or } \mathbf{C} = \mathbf{U}\Lambda\mathbf{U}^H \quad (6)$$

where the columns of $\mathbf{U}=[\mathbf{u}_1, \mathbf{u}_2,..., \mathbf{u}_N]$ are the eigenvectors of **C** and Λ holds the eigenvalues.

$$\Lambda = \mathrm{diag}[\lambda_1,\lambda_2,...\lambda_N] \text{ with } \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \quad (7)$$

The transformation can be written as:

$$\mathbf{C} = \sum_{n=1}^{N} \lambda_n \mathbf{u}_n \mathbf{u}_n^H \text{ and } \mathbf{C}^{-1} = \sum_{n=1}^{N} \lambda_n^{-1} \mathbf{u}_n \mathbf{u}_n^H \quad (8)$$

and since $\mathbf{U}^H = \mathbf{U}^{-1}$ (a property of a unitary matrix), $\mathbf{u}_i^H\mathbf{u}_j$ constitute an orthonormal set.

Assuming that there are more sensors than unknown sources, i.e. $M \leq N$, [2] shows that there must be (N-M) eigenvalues λ_n equal to the noise variance σ^2. The corresponding (N-M) eigenvectors form the noise subspace. As a result the M largest eigenvectors of **C** are the M orthonormal vectors which form a subset of the entire complex vector space. This space is known as the signal subspace and it contains the signal vectors.

The subspace approach can be expressed concisely by:

$$\mathbf{C} = \begin{bmatrix}\mathbf{U}_s & \mathbf{U}_n\end{bmatrix} \begin{bmatrix}\Lambda_s & 0 \\ 0 & \Lambda_n\end{bmatrix} \begin{bmatrix}\mathbf{U}_s \\ \mathbf{U}_n\end{bmatrix} \quad (9)$$

It can be shown [2] that $\mathbf{V}^H\mathbf{u}_n=0$ for n=M+1,...,N. By sweeping the direction vector $\mathbf{v}^T(\theta)$ through all possible values of θ and over all noise eigenvectors we derive the $MUSIC_n$[2] and Johnson and DeGraaf ($J\&D_n$)[4] direction finding functions. Whereas MLM_n[3] is derived by sweeping $\mathbf{v}^T(\theta)$ over all eigenvectors.

$$\mathrm{MUSIC}_n(\theta) = 1 \Big/ \left(\sum_{n=M+1}^{N} \left| \mathbf{v}^H(\theta)\mathbf{u}_n \right|^2 \right) \quad (10)$$

$$\mathrm{J\&D}_n(\theta) = 1 \Big/ \left(\sum_{n=M+1}^{N} \lambda_n^{-1} \left| \mathbf{v}^H(\theta)\mathbf{u}_n \right|^2 \right) \quad (11)$$

$$\mathrm{MLM}_n(\theta) = 1 \Big/ \left(\mathbf{v}^H(\theta) \sum_{n=M+1}^{N} \lambda_n^{-1} \mathbf{u}_n \mathbf{u}_n^H \mathbf{v}(\theta) \right) \quad (12)$$

2. Area of Investigation

Since the discovery of the EEG 60 years ago, innumerable studies have investigated the relationships between neural phenomena, the performance of cognitive tasks, and associated changes in the EEG which are called Event Related Potentials[6]. A novel extension of traditional methodology, has been developed by SCAN, based on the technique of Steady-State Visually Evoked Potentials (SSVEP) in which the subject is exposed to a continuously flickering visual driving signal whilst performing cognitive tasks[7]. The signal processing significance of the visual driving signal is that in excess of 38% of all sensory input pathways to the brain's cortex are linked to the visual pathways[8], so that driving the visual pathways presents a substantial known input driving signal to the cortex. The system identification problems which are intrinsic to most EEG signal analysis work are therefore ameliorated to some extent.

The EEG is recorded using a specially designed helmet with 64 sensors. The rigidity of the helmet guarantees the relative position of the electrodes, which are positioned according to the International 10-20 system. Additional electrodes are placed at sites resulting in an average inter-electrode spacing of approximately 2.5 cm[7].

By measuring the spatial distribution of EEG activity under well-defined, stringent test conditions[7][8], estimation of the positions of the electrocortical generators in the brain is equivalent to the classical problem of estimating the location of multiple emitters. The estimation problem is complicated by:

- The geometry of the array. Most of the research in the area of direction finding using spectral estimation techniques is based on either linear or circular planar arrays. This assumption may not be valid in EEG measurements where the sensors are placed on the surface of the scalp.
- The well known model-based direction finding techniques assume that the array is far enough from the sources to ensure planar impinging waves. This may not be true for the EEG measurements.
- Noise present in the EEG usually is not Gaussian, but more likely closer to 1/f noise.

3. Simulation Results

This section describes the results of simulations with the subspace algorithms, and a discussion of the limitations of these algorithms under the conditions outlined below:

- Varying the separation between sources and sensors to investigate the effect of the curvature of the arriving waves.
- Varying the radius of curvature of the array.

The following simulations were based on 1000 data samples, S/N of 20dB and 8 sensor elements. Note the three subspace algorithms produced similar results, thus the figures only show the results for MUSIC. Figures 1-3 present results for one source, whilst figures 4-5 are for two sources.

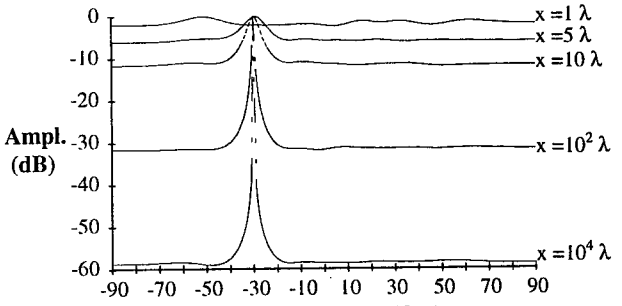


Figure 1 DOA= -30°, linear array, x=distance from array.

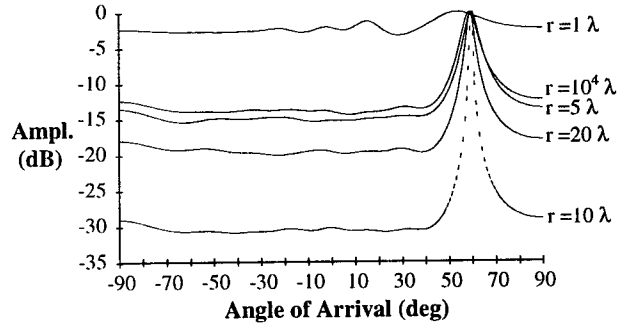


Figure 2 x=5λ, θ = 60°, r= radius of curved array.

Given a linear array, figure 1 shows that as the source is brought closer to the array the performance of the subspace algorithms deteriorates. This is expected since the curvature of the impinging wavefronts increases. To improve the performance in the "close source case" we investigated varying the radius of curvature of the array. Figure 2 shows that there is a particular radius of curvature ($\mathbf{r_{opt}}$) of the array which results in optimum performance of the algorithms.

Figure 3 shows how the performance of the subspace algorithms varies with the direction of the source when the source is close to the array. An interesting observation is that the larger the angle of arrival the lower the noise floor. A larger angle of arrival implies that the distance between the source and the furthest sensor is greater than when the source has a smaller angle of arrival. This extra distance implies the waves travel further thereby better approximating a plane wave.

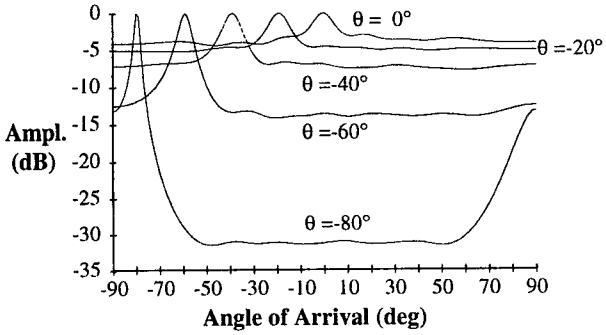


Figure 3 x=5λ, linear array, θ = DOA.

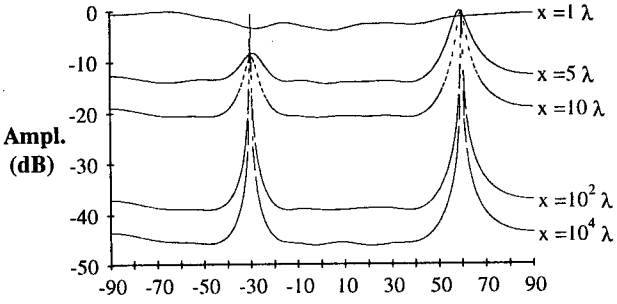


Figure 4 DOA $\theta_1 = 60°$, $\theta_2 = -30°$, $x_1 = x_2 = x$, linear array.

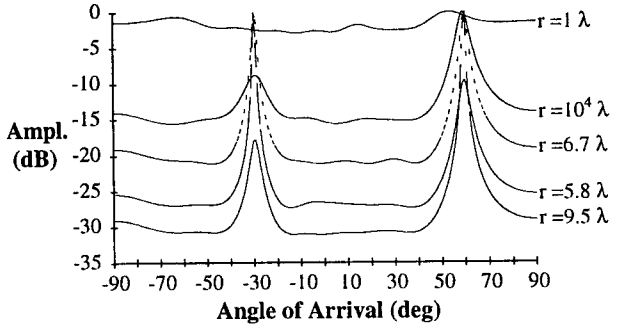


Figure 5 $\theta_1 = 60°$, $\theta_2 = -30°$, $x_1 = x_2 = 5\lambda$, curved array.

As expected, in figure 4 we see similar results to those obtained in figure 1. As the sources are moved closer to the array the performance of the algorithms degrades. Note that the strength of the detected peaks is different. This is expected because the performance of the algorithms is better for larger angles, as can be seen from figure 3. Again, varying the radius of curvature of the array, results in the best performance occurring at $\mathbf{r_{opt}}$ as per the single source case. Figure 5 shows that $\mathbf{r_{opt}}$

depends on the angle of arrival. Given that both sources are 5λ from the array $\mathbf{r}_{opt}$ = 5.8λ for source 1 and $\mathbf{r}_{opt}$ = 9.5λ for source 2. Note peaks of equal strength can be obtained when $\mathbf{r}_{opt}$ = 6.7λ for θ_1= 60°, θ_2= -30°.

4. Analysis of Preliminary EEG Results

In this section we present the preliminary results of applying the subspace techniques to the EEG data. The EEG data was recorded in the presence of visually applied driving signals at a range of frequencies. The particular case under investigation is when the subject is exposed to a sinusoidal visual driving signal of 13 Hz. Seven sensors were chosen located from the front to the back of the head. The sensors chosen were spaced at approximately 35mm. Assuming an average wave velocity in the cortex of 7ms^{-1} the separation becomes approximately 0.065 wavelengths. The EEG signal was filtered to remove unwanted components, the correlation matrix formed and the direction functions plotted.

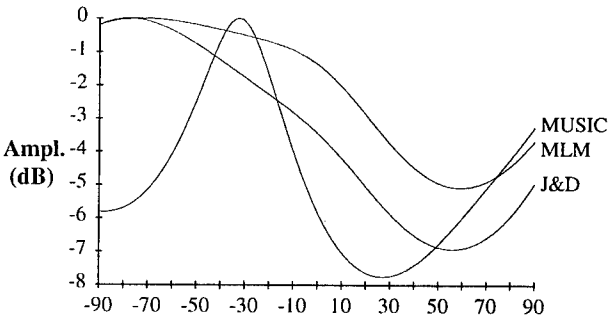


Figure 6

The result shown in figure 6 indicates that there is no localisation of the sources, ie. no distinct angle of arrival is identifiable. This result is promising in that the data used was obtained using a visual stimulus of 13 Hz. This stimulation excites the parts of the brain which are not taking part in any other activity, hence we expect several areas of the brain would pose as possible source locations.

To improve the estimate of the location of the sources the following factors need to be considered;

The curvature of the detection array of sensors needs to be adjusted to obtain $\mathbf{r}_{opt}$. In this particular experiment it wasn't possible to alter the curvature of the detection array. This will be the subject of a subsequent paper.

The result also shows that Music gives a different outcome in comparison to J&D and MLM. This observation may be due to the fact that the MUSIC method doesn't have a weighting function, see equ. **10**, **11** and **12**.

Conclusion

We have shown that the performance of the subspace techniques degrades as the sources are brought closer to the array. This degradation can be counteracted by changing the curvature of the array. An optimum array curvature exists which exhibits best detection performance for a given angle of arrival.

Application of the subspace techniques to the EEG context seems promising as a method for the location of electrocortical generators in the brain. Whilst the results obtained here are encouraging, there is scope for further work. For example the investigation of a compensation filter which would correct the non-optimum array curvature. This may need to be implemented using an artificial neural network, because the exact source distance will be different for different cases.

References

[1] M. Viberg B. Ottersten & T. Kailath, "Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting ", *IEEE Trans., ASSP* vol. 39, no. 11, pp. 2436-2449, Nov 1991.

[2] R.O. Schmidt, "A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation", Ph.D. dissertation, Stanford Univ. CA, 1981.

[3] J. Capon, "High-resolution frequency wavenumber spectrum analysis", Proc. IEEE, vol. 57, no. 8, pp. 1408-1418, 1969.

[4] D.H. Johnson & S.R. DeGraaf, "Improving the Resolution of Bearing in Passive Sonar Arrays by Eigenvalue Analysis", *IEEE Trans.*, ASSP vol. 30, no. 4, pp. 638-647, 1982.

[5] D Klimovski, A. L. Cricenti, A. A. Sergejew, G.K. Egan "Estimation of the Position of Electrocortical Generators Via Subspace Techniques." *Proc. IEEE ICASSP'94*, Adelaide, Australia, pp. IV189-IV192.

[6] A.S. Gevins & B.A. Cutillo, "Signals of Cognition", In: F.H. Lopes da Silva, et. al. (Eds.), Clin. Apps. of Comp. Anal. of EEG and other Neurophysiological Signals Amsterdam: Elsevier, pp. 335-384, 1986.

[7] R.B. Silberstein, et. al., "Steady-State Visually Evoked Potential Topography Associated with a Visual Vigilance Task", *Brain Topography*, vol. 3, no. 2, pp. 337-347, 1990.

[8] A. Brodal, "Neurological Anatomy", Oxford: Oxford University Press, 5th Printing 1976.

ACKNOWLEDGEMENTS

The Authors wish to thank SCAN for their contribution.

A Two Step Incoherent Signal Subspace Averaging to Mitigate Multipath Due to Propeller Scattering

Moeness G. Amin and Mark D. Rosenstock

Department of Electrical and Computer Engineering
Villanova University
Villanova, Pa 19085

Abstract

A two-step incoherent signal subspace averaging technique is applied to locate the sonobuoy in the presence of a highly coherent environment generated by the scattering of the sonobuoy signals from the aircraft propellers. The proposed technique is based on the assumption that accurate modeling of the scattering modulation effects of the propellers is available. This information gives insights into the relative contributions of the direct and multipath components to the signal subspace. The first step of the proposed technique amounts to modifying the MUSIC spectrum by projecting a weighted sum of steering vectors onto the noise subspace. The second step is to perform incoherent subspace averaging across the sonobuoy frequency channels. We show that significant improvement is achieved using the proposed technique over the case of applying noise subspace eigenstructure methods.

1. Introduction

One application of an airborne antenna array is to receive information from sonobuoys as well as to locate their positions. The aircraft drops sonobuoys in the water and starts to monitor their signal transmission. As the aircraft moves along its flight path, the same buoy could be at different positions at different times with respect to the aircraft, due primarily to aircraft motion and to a much lesser extent, movement of the buoy by ocean current. The buoy could be at some distance from the aircraft or directly underneath it. Therefore, the incident angle of the signal on the aircraft array and propellers will not be constant. One of the primary tasks for these flying missions is to locate the sonobuoy with reasonable accuracy.

With the blades of the aircraft propellers in continuous rotation, it can be much expected that the sonobuoy signal scattered from the propellers and arriving at the array to be a modulated version of the direct path signal. Models describing distributed source environment [1], and specifically the propeller return [2] have been recently described. Also, accurate scattering calculations can be performed using the FDTDC [3]. Except for the angle of arrival of the direct path, almost all parameters defining the modulation effects are known [1]. This includes the rotational speed, distances of the blades tips and roots from the center of rotation, number of blades and propellers, and the range of the propeller from the array.

The modulation effects of the propellers depend on several parameters. A number of sidebands often result about the center frequency of the sonobuoy signal. Depending on the number of the propeller blades and the frequency of rotation, one or several sidebands of the propeller reflection fall into the information bandwidth of the direct signal, causing severe multipath degradation effects on the performance of the localization and nulling techniques of the airborne array system. Due to the coherent signal environment, the optimum beamformers (high resolution DOA spectra) not only fail to form nulls (peaks) in the direction of the signals incident on the array, but also it tends to cancel the desired (look direction) signal in the output. This failure occurs even with the decorrelation effects introduced by the motion of the array on the aircraft [4,5].

Preprocessing techniques may prove inefficient for the underlying problem. Spatial filtering methods [6,7] place a low array gain over the interference spatial sector to remove, or at least attenuate the interference outside the sector of interest. Subarray averaging methods [8,9] decorrelate the coherent arrivals under certain conditions which relate the number of sources and array sensors. Both methods reduce the array aperture and are impractical for planar arrays with small number of sensors. Further, the

This work is supported by ONR grant N00014-94-1-1052

0-8186-7576-4/96 $5.00

decorrelation methods are only proper for point sources and do not work using models of distributed sources.

In this paper, we estimate the sonobuoy elevation and azimuth positions by making use of 1) the knowledge of the propeller spatial coordinates, relative to the array, 2) the availability of accurate modeling of the multipath signals. While the former defines the spatial sector of the multipath signals, an accurate model of the propellers provides the means to obtain the relative phase and power of the multipath signal relative to its direct path in each of the sonobuoy's frequency channels.

Due to the narrowness of the interference spatial sector or/and the strong coherence of the mutipath signals, the propellers scattered waveforms can be presented by rank one covariance matrix, i.e., their source representation subspace is spanned by a single vector [10]. This vector can be selected as the directional vector corresponding to the center angle of the spatial sector, or more accurately, can be chosen as the principle eigenvector of the covariance matrix. Each snap shot can, therefore, be modeled as a weighted sum of three directional vectors; one corresponds to the sonobuoy direct path and the others to the two propellers on the aircraft. This paper assumes knowledge of these *fixed* relative weights, and modifies the noise subspace based eigenstruture methods [11] so as their spectra include only peaks whose number and locations, respectively, equal the number and positions of sonobuoys in the field of view.

2 .The Principle Assumption

The covariance matrix of the multipath signals over the sector Θ_i , i=1,2 for the two aircraft propellers can be expressed as

$$R_i = \int_{\Theta_i} \eta(\theta,\varphi) a(\theta,\varphi) a^H(\theta,\varphi) d\theta d\varphi \tag{1}$$

where $a(\theta,\varphi)$ is the directional vector which is a function of both the elevation and the azimuth, and $\eta(\theta,\varphi)$ provides the distribution of the energy over Θ_i. The source representation subspace of the propeller scatters is taken as the principle eigenvector **e** of **R**. Therefore, the vector spanning the subspace of the overall covariance matrix of the coherent direct and mutipath signals takes the form

$$e_o = s(\theta,\varphi) + \alpha\, e_1 + \beta e_2 \tag{2}$$

The weights α, β are complex values and reflect the propeller amplitude and phase changes to the sonobuoy signal. The primary assumption in this paper is the apriori knowledge of these weights from existing distributed sources and propeller scattering models[1-3]. In the simulation section, we assume equal distribution of energy over the spatial sectors of both aircraft propellers. In this case, the covariance matrix in (1) is produced by replacing the integral with a sum and incorporating several directional vectors which uniformly cover the interference sectors.

3. Modified Eigenstructures

For narrowband signals, the multiple signal classification (MUSIC) spectrum is given by

$$S(\theta,\varphi) = \frac{1}{s^H(\theta,\varphi) V V^H s(\theta,\varphi)} \tag{3}$$

where 'H' stands for hermition. The matrix **V** spans the noise subspace, which in the underlying problem, is of dimension Nx(N-L), where L is the number of sonobuoys. In the proposed technique, we use equation (2) to gain insights into the make up of the signal subspace through proper modeling of the aircraft propeller effects. We modify the above equation to

$$S_{mod}(\theta,\varphi) = \frac{1}{s_{mod}^H(\theta,\varphi) V V^H s_{mod}(\theta,\varphi)} \tag{4}$$

where

$$s_{mod}(\theta,\varphi) = s(\theta,\varphi) + \alpha e_1 + \beta e_2 \tag{5}$$

The difference between (3) and (4) is that in (4), the steering vector s along with the two fixed vectors e_1, e_2 is projected onto the noise subspace.

For broadband signals, the above projection is averaged over the entire frequency band of the sonobuoy signal. Assuming M channels, then

$$S_{mod}(\theta,\varphi) = \frac{1}{\sum_{i=1}^{M} \left| s_{mod}{}^H(\theta,\varphi,; f_i) V_i \right|^2} \tag{6}$$

$$s_{mod}(\theta,\varphi; f_i) = s(\theta,\varphi; f_i) + \alpha_i e_1(i) + \beta_i e_2(i) \tag{7}$$

The above type of averaging constitutes the second step of the proposed technique and is similar to incoherent subspace averaging [10], which is proposed for increasing the SNR.

4. Simulations

In the first example, the simulation performed consisted of two groups of completely correlated signals. The first group consists of a desired signal arriving at (θ,φ)= (15,15) degrees with two clusters of multipath signals arriving on the different sides of the desired signal. The two clusters, representing the propeller scattering signals are centered at (56, 170) and (49,10) degrees. Each cluster spans $\Delta\theta = 3, \Delta\varphi = 6$ degrees. The second group has a desired signal arriving at (35,-60) degrees and also with two clusters of propellers multipath signals arriving at the same

angles and spanning the same spatial sector as in the first group. The clusters associated with each desired signal are correlated with the desired signal for that group only. The sonobuoy signals are 20 dB higher than the uncorrelated Gaussian noise. The number of data snapshots taken to generate the estimate of the noise subspace was 1024.

The comparison of the conventional two dimensional MUSIC algorithm and the proposed technique is shown in Figure 1. The proposed technique resolves the direction of arrival (DOA) of the desired signal of each group whereas the MUSIC algorithm completely fails to resolve the DOA of any of the incoming signals. Figure (1-a) shows the contour plot for the MUSIC algorithm applied just for the first group. Figure (1-b) shows the contour plot for the proposed technique where the source location is successfully estimated at (15,15) degrees. Figure (1-c) shows the contour plot for the MUSIC algorithm applied to the above two groups occurring simultaneously. Figure (1-d) shows tha the proposed technique correctly resolve the DOAs for the desired signals from each group, (15,15) and (35,-60).

The second example deals with a broadband signal, where averaging across the frequency band is performed via equation (6). The wavefroms incident on the array consisted of one group of coherent signals, covering half of the total normalized bandwidth. The desired signal arrives at (15,35) degrees with two clusters of multipath signals arriving on its different sides as in the first example.

The comparison of the incoherent signal-subspace (ISS) of the two-dimensional MUSIC algorithm and the proposed technique is shown in Figure 2. The ISS for the MUSIC algorithm completely fails to resolve the DOA of the incoming signal, as shown in the contour plot of Figure 2 (a). The contour plot of Figure 2 (b) shows the clear resolution of the DOA of the direct path signal.

5. Conclusions

The problem discussed in this paper is the estimation of sonobuoy position in the presence of highly coherent environment generated by the propeller scatters of the sonobuoy signals. A two-step incoherent subspace averaging technique was introduced which mitigates the effect of multipath on the noise subspace-based eigenstructure methods. This technique is based on the knowledge of the propeller spatial coordinates, relative to the array, and the availability of accurate modeling of the propeller multipath reflections. The two steps correspond to two types of averaging. The first is to project a weighted sum of the steering vector and the source subspace representations of the multipath spatial sectors on the noise subspace. The second averaging is designated for broadband signals, and is performed over the above projections at different frequency bands. It was shown that the proposed technique performs properly in the presence of one and two sonobuoy signals. The use of coherent subspace averaging in place of the second step is expected to improve resolution and was not explored in this paper.

References

[1] S. Valaee, B. Champagne, and P. Kabal," Parametric localization of distributed sources," IEEE Trans. Signal Processing, September 1995.

[2] J. Martin and B. Mulgrew, "Analysis of the theoretical radar return signal from aircraft propeller blades," Proceedings of the 1990 IEEE International Radar Conference (RADAR-90), Washington, D. C., May 1990.

[3] K. Kunz, R. Luebbers, The Finite Difference Time Domain Method for Electromagnetics, CRC Press, 1993.

[4] J. LI and R. Compton," Angle of arrival estimation of coherent signals using array doublet in motion," IEEE Transactions on Aerospace and Electr. Systems, January 1994.

[5] F. Harber and M Zoltowski," Spatial spectrum estimation in a coherent signal environment using an array in motion," IEEE Transactions on Antenna and Propagation, March 1986.

[6] C. Zhou, F. Haber, and Q Shi," On spatial filtering for AOA estimation in a scattering environment by eigendecomposition-based methods," IEEE Transactions Antenna and propagation, June 1990.

[7] A. Bassias and M. Kaveh,"Coherent signal-subspace processing in a sector," IEEE Transactions on Systems, Man, and Cybernetics, September/October 1991.

[8] T. J. Shan and T. Kailath, "Adaptive beamforming for coherent signals and interference," IEEE Trans. Acoust., Speech, and Signal Processing, vol. 33, pp. 527-536, June 1985.

[9] S. U. Pillai and B. H. Kwon, "Forward/Backward spatial smoothing techniques for coherent signal identification," IEEE Trans. Acoust., Speech, and Signal Processing, vol. 37, pp. 8-15, January 1989.

[10] K. Buckley and X. Xu," Spatial-spectrum estimation in a location sector," IEEE Transactions on Acoustics, Speech, and Signal processing, November 1990.

[11] M. Wax, T.J. Shan, and T. Kailath, "Location and the spectral density estimation of multiple sources," Proceedings of 16th Asilomar Conference on Cir., System, and Comp., Nov. 8-12, 1982.

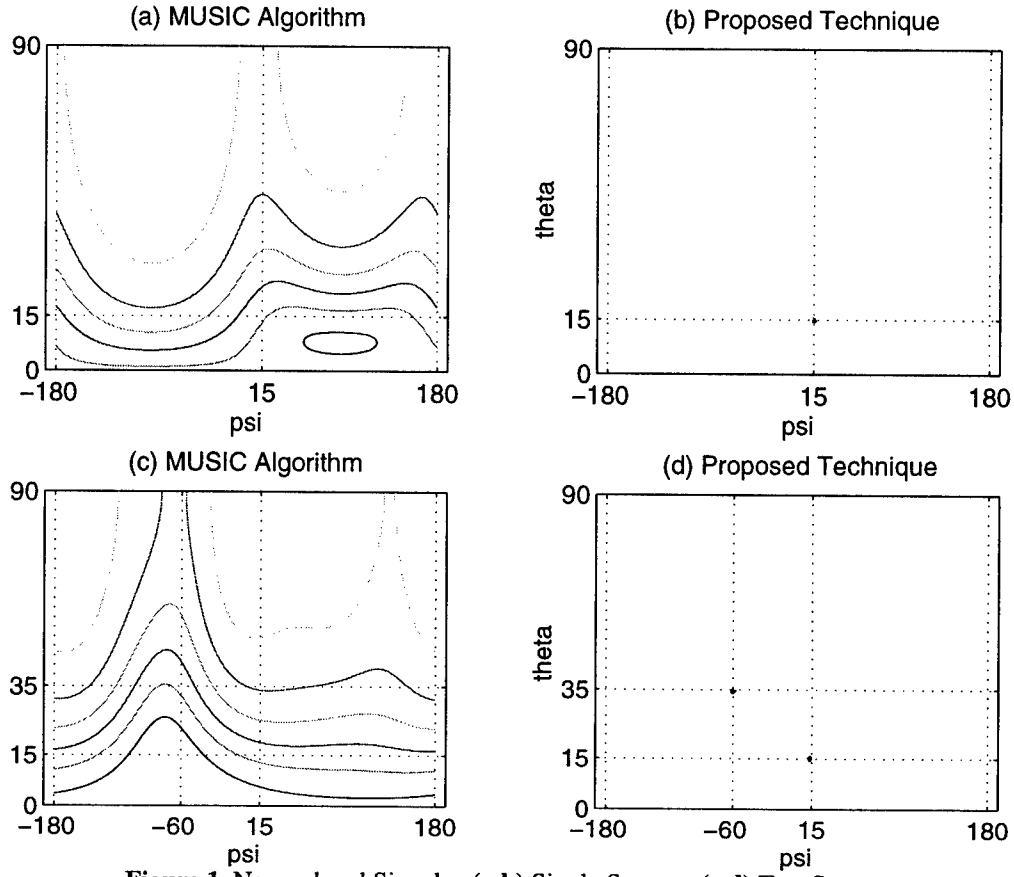


Figure 1 Narrowband Signals, **(a,b)** Single Source **(c,d)** Two Sources

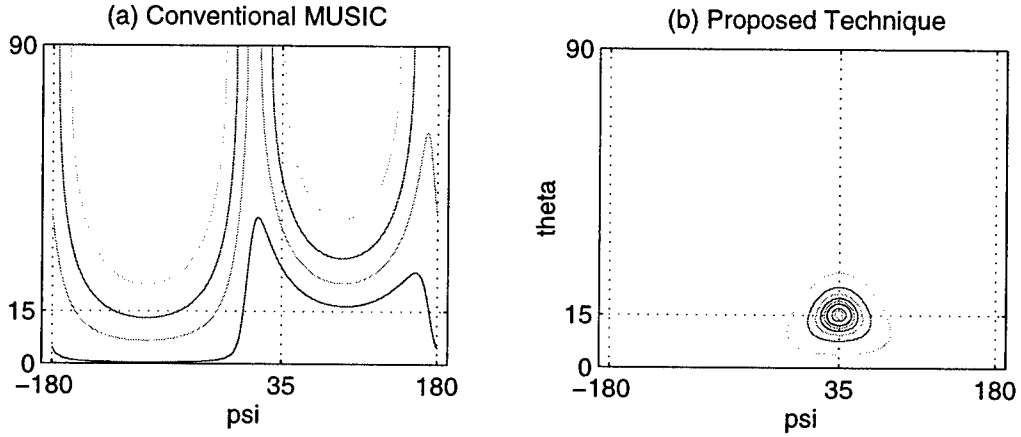


Figure 2 Broadband Signals, **(a)** Conventional MUSIC Algorithm **(b)** Proposed Technique

Adaptive Wideband Aeroacoustic Array Processing

Tien Pham & Brian M. Sadler
Army Research Laboratory
Adelphi, MD 20783-1197
tpham@arl.mil

ABSTRACT

Incoherent and coherent wideband array processing techniques for aeroacoustic detection and tracking of ground vehicles are contrasted. Experimental results for a circular array are presented, illustrating complexity and performance tradeoffs. Incoherent and coherent MUSIC are used for comparison. Complexity is dominated in both cases by singular value decompostion (SVD) calculation performed M times for the incoherent case and S times for the coherent case, where M is the number of frequency bins and S is the number of look angles. Good results are obtained with the incoherent method for small M provided adequate narrowband SNR is available. The coherent approach is more statistically stable, and S can be reduced by employing a priori coarse direction estimates.

1. Introduction

We contrast coherent and incoherent wideband array processing techniques for aeroacoustic detection and tracking of ground vehicles. Experimental results for a circular array of 6 sensors plus 1 at the array center are presented, illustrating complexity and performance tradeoffs in coherent versus incoherent processing.

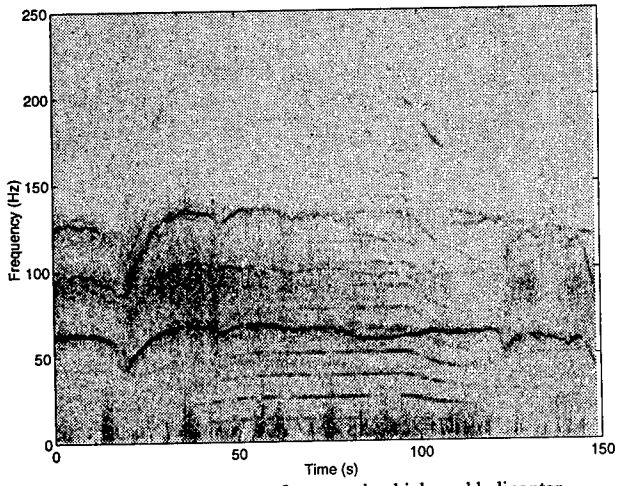


Figure 1. Spectrogram of a ground vehicle and helicopter.

In this application, array baselines are physically constrained by system requirements and variable spatial coherence, motivating use of super-resolution methods [1]. The problem is made difficult by a number of factors. Source acoustic signatures are generally nonstationary and undergo severe fading. The usable channel is largely restricted to [20,200] Hz due to wind noise at low frequencies and poor propagation at higher ones. The channel response is generally nonstationary due to a variety of atmospheric and terrain factors. There may also be significant time-varying multipath.

A typical spectrogram of a moving vehicle at close range, with a helicopter flying nearby, is shown in figure 1. The helicopter's signature consists of sharp and stable harmonics emanating from the main rotor blade. These are evident in the time interval 50-100 seconds. The ground vehicle also exhibits a harmonic structure but it is very nonstationary and exhibits strong fades during vehicle maneuvering. Note the lack of acoustic energy beyond 200 Hz. The combined effects of source and channel nonstationarities produce significant signal variability, even at relatively close ranges of hundred of meters.

2. Incoherent and Coherent Processing

A natural extension of narrowband high resolution subspace methods is to combine narrowband beampatterns over many temporal frequencies [2]. This approach is useful for aeroacoustics if there is sufficient SNR in multiple frequency bins, such that narrowband methods such as MUSIC yield good results independently for each bin. In addition to the relatively high narrowband SNR requirement, disadvantages of this *incoherent* approach include degradation in the presence of correlated multipath, and a general lack of statistical stability when compared to wideband coherent methods. The incoherent averaging can lead to false peaks in the resulting averaged beampattern.

To overcome the nonstationary nature of the source, the data is segmented before processing into fixed blocks, and stationarity is assumed over each block. We have found that this is a reasonable assumption for intervals on the order of 1 sec. Over each processing interval it is assumed that a single frequency bin is occupied by a single source only. This takes advantage of the nonstationarity of the sources, and simplifies the algorithmic complexity as well as estimation of the number of sources. This assumption is justified because different wideband sources are not likely to occupy all of the same bins in any given processing interval, and change bins as a function of time. In practice, the direction of arrival (DOA) estimates are fed into a

0-8186-7576-4/96 $5.00

tracker that is reasonably robust and therefore able to fill in missing or remove outlying data.

Wideband *coherent* processing gain is possible using the steered covariance method (STCM) originated by Wang and Kaveh [3,4]. STCM is based on forming the composite covariance matrix given by

$$\hat{R}(\theta) = \sum_{m=1}^{M} T(\omega_m, \theta)\hat{R}_Y(\omega_m)T(\omega_m, \theta)^H \qquad (1)$$

where M is the number of narrowband frequency bins, and $\hat{R}_Y(\omega_m)$ is the estimated spatial correlation matrix at frequency ω_m. The steering or focusing matrix $T(\omega_m, \theta)$ is a function of both frequency and look angle θ. Here it is defined as

$$T(\omega_m, \theta) = \begin{bmatrix} e^{2\pi f_m \Delta t_1} & 0 & \dots & 0 \\ 0 & e^{2\pi f_{mk} \Delta t_2} & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & e^{2\pi f_m \Delta t_N} \end{bmatrix} \qquad (2)$$

where $\Delta t_i = \frac{d}{c}\sin\phi_i$, $\phi_i = \theta - \alpha_i$ where α_i is the relative angle to the normal for sensor i for i=1, 2, ..., N, d is the radius of the circular array, and c is the speed of sound in air. Other forms for $T(\omega_m, \theta)$ have been suggested to reduce focusing errors. The resulting focused covariance matrix $\hat{R}(\theta)$ is such that signals in the respective narrowband correlation matrices are mapped into the same subspace, yielding coherent processing gain over multiple frequencies. Conventional subspace methods such as MUSIC can then be applied to $\hat{R}(\theta)$, thus requiring eigenanalysis for each θ.

The complexity of the coherent approach is increased due to the need for computing $\hat{R}(\theta)$ for every θ. However, the computational load can be lowered by using preliminary estimates of the source locations, obtained, e.g., by conventional beamforming [3]. Also, it is assumed that there is at most a single source for a single look angle θ. As we will see, the relative computational complexity between the coherent and incoherent techniques depends on the relative size of the number of look directions versus the number of narrowband frequency bins over which wideband processing occurs.

3. Implementation

In this section the processing schemes are described, and estimates of complexity presented for comparing the coherent and incoherent approaches. In both cases MUSIC is used as the means of computing the beampattern. The basic steps are (i) use block-adaptive pre-processing to adaptively select the narrowband frequency bins, (ii) apply incoherent or coherent techniques, and apply MUSIC, and (iii) estimate the directions of the sources from the resulting beampatterns.

Let $y_i(n)$ denote the output of the ith sensor from an array of N sensors, and let $Y_i(k)$ denote $DFT\{y_i(n)\}$. The average sum of the $|Y_i(k)|^2$ is obtained in order to adaptively select frequency bins of interest. This can be performed in a variety of ways, from simple thresholding based on bin SNR, to more complex schemes such as harmonic association. Here, we simply select the M highest power bins within the range ω_{low} to ω_{high}.

The conventional narrowband MUSIC beampattern is computed M times. For each look angle θ, we compute

$$P_{Incoh}(\theta) = \sum_{m=1}^{M} \left[E(\omega_m, \theta)^H \Pi^{\perp}(\omega_m) E(\omega_m, \theta)\right]^{-1} \qquad (3)$$

where $E(\omega_m, \theta) = diag\{T_s(\omega_m, \theta)\}$ is the steering vector and the noise orthonormal projector is defined as

$$\Pi^{\perp}(\omega_m) = \hat{U}_n(\omega_m)\hat{U}_n(\omega_m)^H. \qquad (4)$$

Taking $\hat{R}_Y(\omega_m)$ to be N x N then, by assumption, the noise subspace consists of N-1 eigenvectors corresponding to the N-1 smallest eigenvalues of $\hat{R}_Y(\omega_m)$, and these form $\hat{U}(\omega_m)$.

The computational complexity is approximately $M[O(N^2) + O(N^3) + S \cdot O(N^2)]$, where M is the number of frequency bins and S is the number of look angles. The first squared term in the bracket corresponds to the formation of the correlation matrix $\hat{R}_Y(\omega_m)$, the cubic term is for an SVD calculation to form $\hat{U}_n(\omega_m)$, and the last term corresponds to (3) which is computed for each look angle.

The STCM approach requires focusing as a function of look direction. Experimental results shown in the next section are based on computing over 360 degrees in 1 degree steps. After computation of $\hat{R}(\theta)$ for some angle θ, the SVD of $\hat{R}(\theta)$ yields the unitary noise subspace estimate $\hat{U}_n(\theta)$. We assume only one target for each look angle, so that the signal subspace consists of one eigenvector, with the other N-1 eigenvectors forming the noise subspace. The coherent wideband MUSIC spatial-spectrum is then calculated via

$$\hat{P}_{Coh}(\theta) = \left[L^H \hat{U}_n(\theta)\hat{U}_n(\theta)^H L\right]^{-1} \qquad (5)$$

where L is an N-element vector of ones.

The computational complexity is approximately $S[M \cdot O(N^2) + O(N^3) + O(N^2)]$. The first term in the bracket corresponds to the formation of $\hat{R}_Y(\omega_m)$ and the focusing operation of (1). Note the diagonal form of $T(\omega_m, \theta)$ reduces the computation in (1). These

operations must be performed over the range of frequencies (M), and the range of look angles (S). The cubic term is for the SVD of $\hat{R}(\theta)$, and the last term is for calculation of $\hat{P}_{Coh}(\theta)$ in (5). These are repeated for each look angle, i.e., S times.

For both methods, the most expensive computational cost is the SVD which is $O(N^3)$. This term tends to dominate the complexity comparison. Note that for the incoherent method it is $M \cdot O(N^3)$ while for the coherent it is $S \cdot O(N^3)$, so that the relative complexity is controlled by the relative size of M and S. By assuming a single target for each distinct frequency bin (for incoherent processing) and a single target for each look angle (for coherent processing), we can potentially apply faster eigenanalysis algorithms than the SVD. To reduce the number of frequency bins, harmonic line association techniques can be used to group a set of frequency bins for each source and then only applying MUSIC to the largest narrowband frequency for each set. To reduce the number of look angles S, coarse angle estimates can be used to narrow the field of view.

4. Experimental results

In this section experimental results for DOA estimation of ground vehicles traveling on a 2 km^2 area of open grass field are presented. For each test run, one of the vehicles was equipped with a GPS sensor to provide accurate positioning ground truth. Figure 2 shows raw experimental DOA estimates, for a single source, for incoherent and coherent wideband MUSIC versus the GPS angles on a test run of 250 seconds in length. Mean square error (MSE) and mean absolute error (MAE) results are shown in table 1 for various sets of M frequency bins. The M frequency components are selected based on the highest bin SNR's in the frequency range of [20,200] Hz.

The MSE's and MAE's are calculated with the outliers removed using the criteria

$$|\varepsilon - median(\varepsilon)| > 3\sigma_{MAD} \qquad (8)$$

where ε is the angle error or angle difference between the DOA estimate and the true angle measured by GPS, and σ_{MAD} is the mean absolute deviation [5]. An example of this is shown in figure 3, with $\pm 3\sigma_{MAD}$ shown as straight lines. The outliers can be caused by several factors including fading, wind noise and acoustic source variations. For the error analysis in table 1, the number of outliers ranges from 15 to 24 out of a total of 125 processing intervals of length 1 sec each, sampling rate of 1 kHz, and 1024-pt FFT's. For $M = 1$, incoherent and coherent wideband MUSIC reduce to the narrowband case. Processing gain is evident for both methods, in that the estimates generally improve for increasing M. For this single source experiment, the incoherent approach produced smaller errors both in terms of MSE and MAE, reflecting the generally high SNR in this experiment.

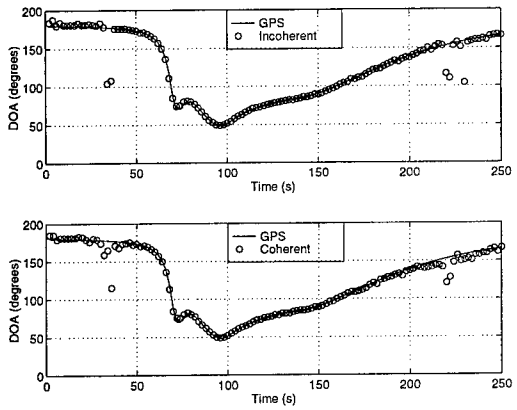


Figure 2: Raw DOA estimates for (a) incoherent and (b) coherent wideband MUSIC for M=50 and GPS ground truth.

	Incoherent MUSIC		Coherent MUSIC	
M	MSE	MAE	MSE	MAE
1	3.558	1.419	3.558	1.419
10	2.144	1.083	3.948	1.422
20	1.235	0.870	3.684	1.366
50	1.221	0.863	2.345	1.130
100	1.172	0.838	2.684	1.178

Table 1. MSE and MAE for wideband processing over M frequency bins between [20,200] Hz.

While the single source, high SNR case can be handled with incoherent MUSIC and small M, the situation changes with multiple sources and low SNR's. A two-source example is illustrated in figure 4, with sources at 50 and 180 degrees. Here beampatterns are shown for a single processing interval, with M varying over 10, 20, 50 and 100. The incoherent method accurately locates the directions of the sources for all four cases. It produces more distinct and sharp beampatterns than the coherent method. However, for M=50 (figure 4c) and especially M=100 (figure 4d), the incoherent method produced additional spurious peaks in the beampattern that can be misconstrued as sources. The explanation for this behavior is that there is high SNR in only a few of the spectral components and no significant SNR elsewhere in the data signature. By incoherently averaging additional beampatterns from low SNR spectral components, the overall beampattern degrades. The beampattern for the coherent method on the other hand, becomes more pronounced as M increases, and exhibits very good statistical stability.

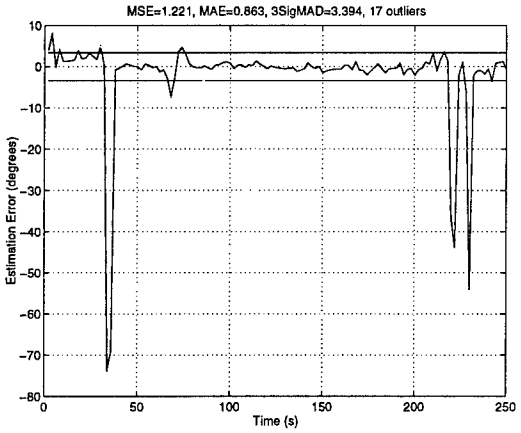


Figure 3: DOA error estimates for incoherent wideband MUSIC (M=50) illustrating outlier removal for MSE and MAE calculation.

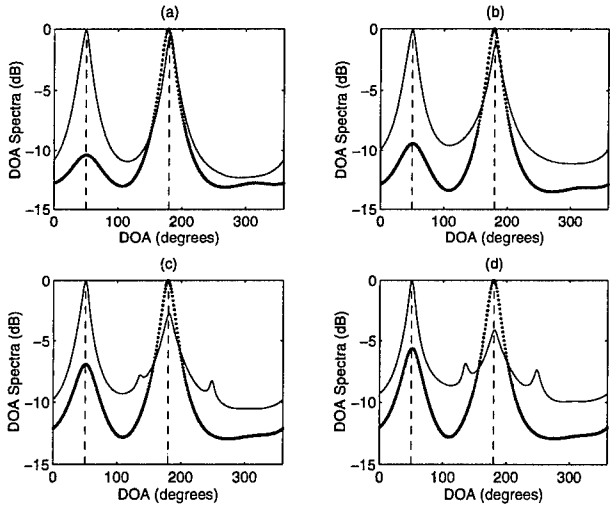


Figure 4: DOA spectra estimates for 2 targets located at θ=50 and θ=180 for incoherent (thin line) and coherent wideband MUSIC for (a) M=10 (b) M=20 (c) M=50 and (d) M=100.

5. Conclusions

Both the incoherent and coherent wideband MUSIC methods provide processing gain over narrowband MUSIC, as exhibited by experiment. Here the sources used are generally characterized as a sum of narrowband frequency components for a majority of the time. Thus, given adequate SNR, incoherent MUSIC performed well and yielded sharp and distinct peaks in the beampattern. However, frequency selection is an issue as the inclusion of low SNR bins tends to degrade the resulting beampattern, reducing source peaks and introducing spurious ones. In contrast, the coherent MUSIC approach is much more statistically stable, with a beampattern that generally improves (rather than degrades) with the addition of lower SNR bins. However, the coherent approach requires more frequency bins be included (i.e., larger M) to achieve the same accuracy, although this is a function of SNR as well. The bias introduced in the coherent processing has been ignored, and results in table 1 may partly reflect this fact. It appears that the coherent approach will degrade more gracefully as the SNR is decreased. Thus, further experiments are warranted for lower SNR (longer range) cases, as well as including sources that do not exhibit strong narrowband signatures.

The computational complexity comparison between the two methods is largely governed by the SVD calculation which is $O(N^3)$, with a multiplier given by the number of spectral components M (incoherent) or number of look angles S (coherent). As we have seen, M can be made small when the source signatures consist of a sum of high SNR narrowband components, enabling use of the incoherent approach. The number of look angles can be reduced by incorporating coarse DOA estimates obtained in a preprocessing step such as a conventional beamformer. Thus, the complexity of the coherent approach can be made manageable (with respect to the incoherent complexity), and is likely to be warranted for more difficult sources at longer ranges.

It is of interest to consider methods for calculating the signal subspace only, as opposed to the full SVD calculation, because in both methods the signal subspace is assumed to consist of one component only. Further work of interest includes reducing computation by exploiting the radial symmetry of the circular array, e.g., see Doron [6], and effects of calibration and sensor placement errors, e.g., see Swindlehurst [7]. We note that only rudimentary effort was made to calibrate the array used in the experiments.

6. References

[1] T. Pham and B. Sadler, "Aeroacoustic wideband array processing for detection and tracking of ground vehicles," *Journal of the Acoust. Soc. of America*, Vol. 98, No. 5, pt. 2, p. 2969, 1995.

[2] G. Su and M. Morf, "The signal subspace approach for multiple wide-band emitter location," *IEEE Trans. ASSP*, Vol. 31, No. 6, pp. 1502-1522, December 1983.

[3] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wideband sources," *IEEE Trans. ASSP*, Vol. 33, pp. 823-831, August 1985.

[4] J. Krolik, "Focused wideband array processing for spatial spectral estimation," chapt. 6 in *Advances in Spectrum Analysis and Array Processing Vol. II*, S. Haykin ed., Prentice-Hall, 1991.

[5] L. Ljung, "*System Identification for the User*," Prentice-Hall, p. 400, 1987.

[6] M. Doron, E. Doron and H. Weiss, "Coherent wide-band processing for arbitrary array geometry," *IEEE Trans. SP*, Vol. 41, No. 1, pp. 414-417, January 1993.

[7] A. Swindlehurst and T. Kailath, "A performance based analysis of subspace-based methods in the presence of model errors, part 1: the MUSIC algorithm," *IEEE Trans. SP* Vol. 40, No. 7, pp. 1758-1774, July 1992.

REDUCED-STATE MAXIMUM-LIKELIHOOD SEQUENCE DETECTION FOR MAGNETIC RECORDING CHANNELS

Jindong Li and John G. Proakis

Center for Communication and Digital Signal Processing
Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115

Abstract

This paper is focused on the design of partial response equalized channels and maximum-likelihood sequence detection for high density magnetic recording systems. Methods for designing linearly equalized partial response channels are presented and applied to a Lorentzian model for the magnetic recording channel. Reduced-state maximum-likelihood sequence detection methods are employed for the partial response equalized channels and the error rate performances of the detectors are evaluated.

1. Introduction

A major factor that limits the density of magnetic recording (MR) systems is intersymbol interference (ISI). To reduce the effects of ISI on high density MR channels, various types of equalizers have been employed [1]. Among these are linear equalizers (LE), decision-feedback equalizers (DFE), and maximum-likelihood sequence detection (MLSD). The latter is efficiently implemented by means of the Viterbi Algorithm (VA).

MLSD is known to be the optimum detection criterion in a channel with ISI, in the sense that the error rate for a sequence of symbols is smallest among the class of equalization methods. However, the computational complexity of the MLSD criterion increases exponentially with the length of the channel memory [2]. Hence, when the span of the ISI is large, the computational complexity of MLSD becomes prohibitive. On the other hand, a LE is significantly simpler to implement. Its major limitation is that it enhances the additive noise in a channel with ISI. The loss in performance of a LE due to noise enhancement is unacceptably high in a high density MR system.

A commonly used method for reducing the computational complexity of MLSD is to combine a LE with MLSD. In particular, the LE is used to equalize the MR channel to a partial response of the type $(1\text{-}D)(1+D)^n$, where D represents a delay of one symbol and n is a non-negative integer that is selected to take the values $n = 0, 1, 2, \cdots$. In general, the optimum choice of n increases as the density of the MR system is increased. By employing a LE to equalize the MR channel to a partial response of short duration, the noise enhancement of the LE is significantly reduced compared to a full response LE. The MLSD that follows the LE is used to detect the data symbols in the partial response signal. Thus, the combination of a LE-MLSD (or LE-VA) is a practical method for achieving high density magnetic recording with a lower computational complexity than MLSD.

In this paper, we investigate another partial response (PR) equalization method for reducing the computational complexity of the MLSD method. For the Lorentzian channel model, we determine the optimum PR targets and the corresponding noise enhancement values for different lengths of the equalized channel. We have found that the optimum method for designing the PR equalized target results in a lower noise enhancement compare to the PR target $(1\text{-}D)(1+D)^n$. However, the noise reduction is achieved at the expense of an increase in the length of the equalized PR target. For PR targets of large length, the computational complexity of MLSD is still prohibitive for high rate MR systems. We have addressed the problem by investigating reduced state MLSD methods. We have found that delayed decision-feedback sequence estimation (DDFSE) as described in [3] is particularly effective for the PR equalized channel.

2. Optimization of the LE

With the conventional LE-VA technique, a linear prefilter(P) or LE is used to adjust the channel (H) to a desired impulse response (Q), which is seen by the VA.

0-8186-7576-4/96 $5.00

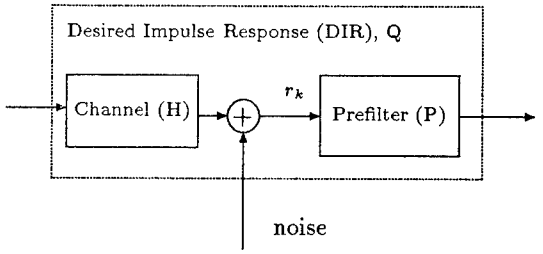


Figure 1: Channel truncation using linear prefilter

We consider methods for selecting P. In all cases, the LE is an FIR filter of length 2M+1. Furthermore, the channel is also modeled as an FIR filter.

A. Optimization of the LE for a Specified DIR

The mean-squared error (MSE) at the output of the LE may be expressed as:

$$MSE = P'AP + Q'Q - 2P'HQ \quad (1)$$

where P is the LE impulse response vector, Q is DIR vector, H is the channel impulse response vector , A is the channel covariance matrix with elements $a_{ij} = \overline{r_i r_j}$, and

$P = [p_{-M} \cdots p_M]$

$Q = [q_0 \cdots q_{L-1}]$

$H = [h_{-N} \cdots h_N]$

For a specified DIR Q, the impulse response P of the LE that minimizes the MSE is

$$P = A^{-1}HQ \quad (2)$$

B. Optimization of the LE for a DIR with Energy Constraint

Falconer and Magee [4] considered the problem of finding the optimum LE response that minimizes the MSE in (1) where the DIR of a specified length is constrained such that Q'Q = 1. The solution to the optimization problem is also given by (2) where the DIR Q is the eigenvector of the matrix (I-H'A^{-1}H) corresponding to the minimum eigenvalue.

C. Optimization of the LE for a DIR with Element Constraint

Suppose we specify the first element q_0 of the DIR to be unity and leave the remaining elements of Q unspecified. That is, Q = [1 q_1 q_2 $\cdots$ q_{L-1}]. The LE impulse response P and the remaining values of Q are selected to minimize the MSE

$$MSE = P'AP + Q'Q - 2P'HQ - 2\lambda(J'Q - 1) \quad (3)$$

where J is an L-element column vector whose first element is 1 and all the other elements are zero. Taking the derivatives of the right hand side of (3) with respect to P, Q, and λ, respectively, and setting the resulting expression to zero, we obtain:

$$\lambda = \frac{1}{J'(I - H'A^{-1}H)J} \quad (4)$$

$$Q_{opt} = \lambda(I - H'A^{-1}H)^{-1}J \quad (5)$$

$$P_{opt} = A^{-1}HQ_{opt} \quad (6)$$

where I is an identity matrix and λ is equal to the minimum mean-squared-error.

3. Linear Equalizer for MR Channels

The magnetic recording (MR) channel is modeled as a linear filter whose step response is the Lorentzian pulse

$$s(t) = \frac{1}{1 + (\frac{2t}{pw50})^2} \quad (7)$$

where pw50 is the "half-amplitude width" of the pulse, which is equivalent to the amount of time that s(t) is greater than or equal to half of its peak value. The input to this channel is a binary data sequence$\{a_k = \pm 1\}$, the channel output is assumed to be corrupted by additive white Gaussian noise(AWGN). The bit rate is $\frac{1}{T_b}$, where T_b is the bit interval. The ratio $S = \frac{pw50}{T_b}$ is called the normalized information density. For the continuous-time system model, Bergmans [5] has derived an equivalent discrete-time system model which is illustrated in Fig. 2, where the parameters are

$$g_k = [\frac{pw50S}{8} tanh(\frac{\pi S}{2})]^{\frac{1}{2}} \frac{k + \frac{S}{2}}{k^2 + (\frac{S}{2})^2} \quad (8)$$

$$h_k = g_k - g_{k-1} \quad (9)$$

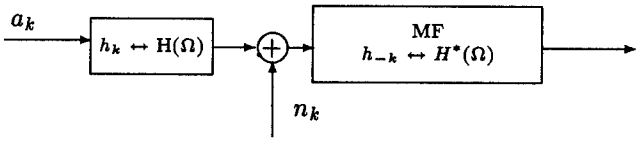


Figure 2: Discrete-time MR system model

4. Performance of the LE-VA system

When the LE is followed by the MLSD that is efficiently implemented by the VA, the error probability of the system may be approximated by

$$P(e) \tilde{>} Kerfc(\sqrt{\frac{d^2_{min}}{8MSE}}) \qquad (10)$$

where K is a constant that depends on the characteristics error events and MSE is the mean-squared-error at the output of the LE. d_{min} is the minimum Euclidean distance for error events.

We note that the probability of error depends on the ratio

$$Ra = \frac{d^2_{min}}{8MSE} \qquad (11)$$

Here, the ratio may be used as a performance index for comparing MR systems with different DIR. It should be emphasized that in using the ratio R_a as a performance index, we have ignored the fact that the noise at the input to the VA is generally correlated. The VA is assumed to ignore this correlation in the computation of the metrics.

5. Reduce-state MLSD

We have observed that by constraining the DIR to have the characteristics Q = [1 q_1 $\cdots$ q_{L-1}], most of the energy in the DIR is contained in the first few coefficients, say q_1,q_2,$\cdots$,q_{L2} where L2< L. This observation suggests that we reduce the computational complexity of the VA by truncating the channel response. Duel-Hallen and Heegard [3] have described an algorithm , called delayed decision-feedback sequence estimation (DDFSE), which performs channel truncation. The complexity of the DDFSE algorithm is controlled by a parameter μ that can be varied from zero to the memory of the channel(in our case, the DIR). The algorithm is based on a trellis search with the number of states equal to 2^{μ}. When $\mu = 1$, the DDFSE reduces to the DFE. When μ equals the length of the channel(DIR), the DDFSE is identical to the full complexity VA. Hence, for 1< μ <L, the DDFSE is a reduced-state VA with feedback incorporated into the structure of path metric computations.

In the Duel-Hallen and Heegard paper [3], it is suggested that the channel be minimum phase. However, this condition is not necessary. As long as most of the channel energy is contained in the first few coefficients of Q, the parameter μ can be selected accordingly to include the larger terms.

For the MR channel with S = 3 and S = 4, the designed DIR's with element constraint are shown in Tables 1 and 2. For these two cases, we have selected μ = L2 = 3 and μ = L2 = 4, respectively. The values of d^2_{min} for the truncated channel and the ratio Ra are given in Tables 3 and 4. For comparison, we also give the corresponding values of d^2_{min} and Ra for the LE-VA when Q is chosen as (1-D)(1+D)^n in Table 5 and Table 6.From Tables 1 and 2, we can see the main energy of the DIR is contained in the first few coefficients of Q. We note that for L> 9 and S=3 and for L>12 and S = 4, there are relatively small performance gains.

6. Simulation Results

We used the two estimation methods, LE-VA with Q chosen as (1-D)(1+D)^n, and LE-DDFSE, to perform simulations with the MR channel S=3 and S=4. The performances are shown in Fig. 3 and Fig. 4. The input SNR is defined as $(SNR)_{in} = \frac{1}{2\sigma^2}$. We can see the LE-DDFSE has a 2dB and 3dB performance improvement over the LE-VA for S=3 and S=4 respectively.

7. Conclusion

In this paper we have presented a method for reducing the computational complexity of the MLSD for high density MR systems. We showed that by proper design of the DIR, a LE followed by a DDFSE algorithm yields superior performance to existing methods.

References

[1] Dennis J. Tyner and John G. Proakis, "Partial Response Equalizer Performance in Digital Magnetic Recording Channels", IEEE Trans. Magn., vol. 29, No. 6, Nov. 1993, pp4194-4208.

[2] Forney, G.D., Jr., "Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol interference", IEEE Trans. Inform. Theory, IT-18, vol. 3, May 1972, pp363-378.

[3] Duel, A. and Heegard, C., "Delayed Decision-Feedback Sequence Estimation", IEEE Trans. Commun., vol 37, No 5, May 1989, pp428-435.

[4] Falconer, D. D., and Magee, F.R., Jr, "Adaptive Channel Memory Truncation for Maximum Likelihood Sequence Estimation", the Bell System Technical Journal, vol. 52, No. 9, Nov. 1973, pp1541-1562.

[5] J. W. M. Bergmans, "Discrete-time Models for Digital Magnetic Recording", Phillips J. Res., vol. 41, No. 6, 1986, pp531-558.

Table 1. DIR at $(SNR)_{in}$ = 13dB (S=3)

L	Q
5	1 0.66 − 0.33 − 0.63 − 0.34
6	1 0.74 − 0.18 − 0.54 − 0.47 − 0.22
7	1 0.77 − 0.11 − 0.46 − 0.43 − 0.31 − 0.14
8	1 0.78 − 0.08 − 0.41 − 0.39 − 0.29 − 0.20 − 0.09
9	1 0.79 − 0.07 − 0.39 − 0.36 − 0.27 − 0.19 − 0.13 − 0.06
10	1 0.79 − 0.06 − 0.38 − 0.35 − 0.25 − 0.17 − 0.13 − 0.09 − 0.04
11	1 0.79 − 0.06 − 0.38 − 0.34 − 0.24 − 0.16 − 0.11 − 0.08 − 0.06 − 0.03
12	1 0.79 − 0.06 − 0.38 − 0.34 − 0.24 − 0.16 − 0.10 − 0.07 − 0.06 − 0.04 − 0.02
13	1 0.80 − 0.06 − 0.38 − 0.34 − 0.23 − 0.15 − 0.10 − 0.07 − 0.05 − 0.04 − 0.03 − 0.01

Table 2. DIR at $(SNR)_{in}$ = 13dB (S=4)

L	Q
5	1 0.86 − 0.13 − 0.67 − 0.41
6	1 1.00 0.10 − 0.61 − 0.67 − 0.32
7	1 1.08 0.28 − 0.44 − 0.67 − 0.53 − 0.23
8	1 1.12 0.37 − 0.32 − 0.58 − 0.56 − 0.39 − 0.16
9	1 1.14 0.42 − 0.25 − 0.50 − 0.51 − 0.42 − 0.28 − 0.11
10	1 1.15 0.44 − 0.21 − 0.45 − 0.46 − 0.39 − 0.30 − 0.20 − 0.08
11	1 1.15 0.46 − 0.19 − 0.43 − 0.43 − 0.36 − 0.28 − 0.22 − 0.14 − 0.06
12	1 1.15 0.46 − 0.17 − 0.41 − 0.40 − 0.33 − 0.26 − 0.20 − 0.16 − 0.11 − 0.05
13	1 1.16 0.47 − 0.17 − 0.40 − 0.40 − 0.32 − 0.24 − 0.19 − 0.15 − 0.12 − 0.08 − 0.02

Table 3. DIR (L2=3) at $(SNR)_{in}$ = 13dB LE-DDFSE (S=3)

L	Q(truncated)	d^2_{min}	MSE	Ra (dB)
5	1 0.66 − 0.33	6.19	0.053	11.65
6	1 0.74 − 0.18	6.32	0.051	11.95
7	1 0.77 − 0.11	6.43	0.050	12.12
8	1 0.78 − 0.08	6.49	0.049	12.20
9	1 0.79 − 0.07	6.52	0.049	12.23
10	1 0.79 − 0.06	6.53	0.049	12.25
11	1 0.79 − 0.06	6.54	0.049	12.26
12	1 0.79 − 0.06	6.54	0.049	12.26
13	1 0.79 − 0.06	6.54	0.049	12.26

Table 4. DIR(L2=4) at $(SNR)_{in}$ = 13dB with LE-DDFSE (S=4)

L	Q (truncated)	d^2_{min}	MSE	Ra (dB)
5	1 0.86 − 0.13 − 0.67	7.51	0.096	9.91
6	1 1.00 0.10 − 0.61	7.89	0.086	11.62
7	1 1.08 0.28 − 0.44	7.56	0.081	11.68
8	1 1.12 0.37 − 0.32	7.36	0.079	11.68
9	1 1.14 0.42 − 0.25	7.31	0.078	11.71
10	1 1.15 0.44 − 0.21	7.30	0.077	11.74
11	1 1.15 0.46 − 0.19	7.30	0.077	11.76
12	1 1.15 0.46 − 0.17	7.30	0.077	11.77
13	1 1.16 0.47 − 0.17	7.31	0.077	11.77

Table 5. DIR at $(SNR)_{in}$ = 13dB with LE-VA (S=3)

n	Q	d^2_{min}	MSE	Ra(dB)
0	1 − 1	8	0.37	4.36
1	1 0 − 1	8	0.11	9.45
2	1 1 − 1 − 1	16	0.12	12.29
3	1 2 0 − 2 − 1	24	0.21	11.69
4	1 3 2 − 2 − 3 − 1	48	0.46	11.15
5	1 4 5 0 − 5 − 4 − 1	120	1.227	10.89

Table 6. DIR at $(SNR)_{in}$ = 13dB with LE-VA (S=4)

n	Q	d^2_{min}	MSE	Ra(dB)
0	1 − 1	8	0.87	0.59
1	1 0 − 1	8	0.29	5.40
2	1 1 − 1 − 1	16	0.26	8.80
3	1 2 0 − 2 − 1	24	0.38	8.98
4	1 3 2 − 2 − 3 − 1	48	0.73	9.20
5	1 4 5 0 − 5 − 4 − 1	120	1.66	9.58

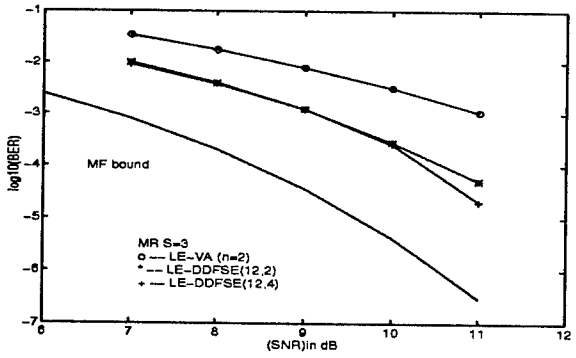


Figure 3 The simulation results for S=3 with LE-DDFSE and LE-VA

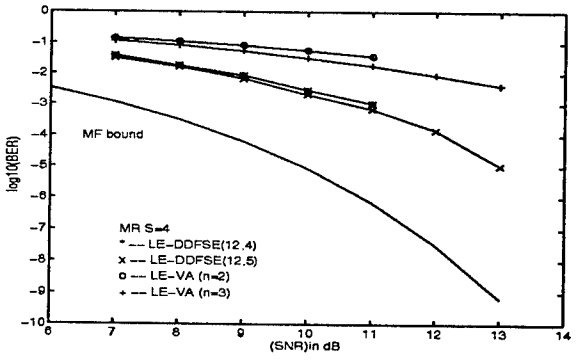


Figure 4 The simulation results for S=4 with LE-DDFSE and LE-VA

Detection and Modeling of Acoustic Emissions for Fault Diagnostics

D. West G. Venkatesan A. Tewfik K. Buckley
M. Kaveh
Department of Electrical Engineering
University of Minnesota
Minneapolis, MN 55455
dwest@ee.umn.edu

Abstract

The formation of microcracks in a material creates propagating ultrasonic waves that are called Acoustic Emissions (AEs). These AEs provide an early warning to the onset of material failure. In practical cases, however, these AEs have to be detected at very low SNRs, amongst strong interference and random noise. This paper presents some preliminary results from an ongoing investigation into the modeling and detection of AEs as a viable technique for predictive diagnostics.

1 Introduction

Automatic monitoring techniques are being considered as a means to safely simplify or dispense of periodic fault inspection procedures. One such automatic monitoring technique is based on the detection of AEs, that are generated due to the formation of microcracks in a material.

AE signals have been extensively studied (e.g. [7][2][1]). However, these studies were using data acquired from isolated material specimens in controlled laboratory conditions. Hence, they do not directly relate to a practical case wherein the AE signal has to be detected in the presence of strong interference, caused due to mechanical motion in the machine. This paper addresses the problem of detecting the AE signal in such a "real world" scenario. The various stages of the proposed procedure is shown in Figure 1. The paper presents some preliminary results obtained on *real* AE[1] and interference data.

*This work was supported in part by ONR under URI.

[1]Thanks to Professor Gerberich and David Bahr of the Material Science Department for their kind assistance in providing us with data.

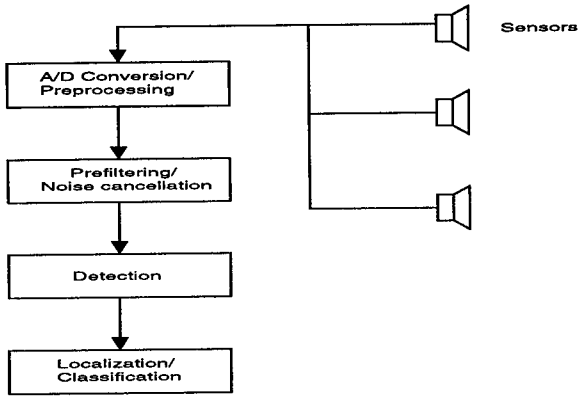


Figure 1. Block diagram of proposed procedure for AE signal processing

2 Acoustic Emissions

AEs are transient in nature, and can be modeled as a sum of decaying complex exponentials [4] as,

$$\sum_{k=1}^{K} A_k e^{\alpha_k[t-T]} \cos(2\pi f_k[t-T]+\phi_k)u(t-T) \quad (1)$$

where $u(t)$ is the step function, and A_k, ϕ_k, α_k, and f_k are amplitude, phase, decay rate, and frequency of the k^{th} AE signal component. AE signals may be broad band, with energy ranging as high as several MHz [7].

Figure 2 shows an AE obtained from a 600nm thick Tantalum Nitride specimen deposited on sapphire, and the corresponding model obtained via Prony's method, ([3]). The estimated parameters are tabulated in Table 1.

This AE was generated by a microcrack that was initiated using a nanoindenter (a device which pushes a diamond tip into a material in a controlled manner).

0-8186-7576-4/96 $5.00

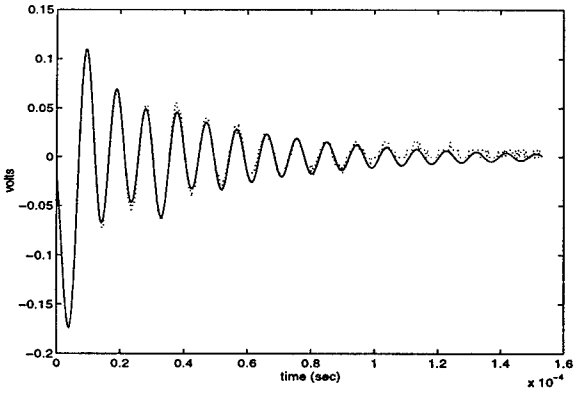


Figure 2. Tantalum Nitride acoustic emission measured, (dotted), and model, (solid).

Table 1. Exponentially decaying sinusoid parameters corresponding to Tantalum Nitride AE.

A(V)	ϕ(radians)	α(/sec)	f(Hz)
0.102	0.075	-2.225e+04	1.057e+05
0.099	2.720	-8.116e+04	6.218e+04
0.080	1.938	-7.214e+04	3.923e+04

The setup to generate such a "microevent" is shown in Figure 3.

3 Prefiltering

We are looking at scenarios where the AE signal is buried in strong interference (which could be periodic) due to mechanical motion, like the movement of a piston in an engine or the rotation of helicopter blades. In consequence, it becomes necessary to first mitigate this interference using a suitable prefiltering technique.

Figure 4 shows typical interference data recorded from a lawnmower at 2MHz. The duration of the observation was one-twentieth of a second. The power spectrum of this lawnmower noise is dominated by the power at frequencies below 30 kHz. Hence, it would

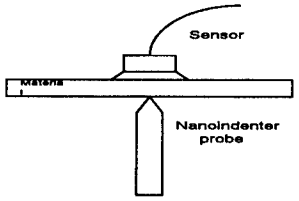


Figure 3. Setup to generate a microevent

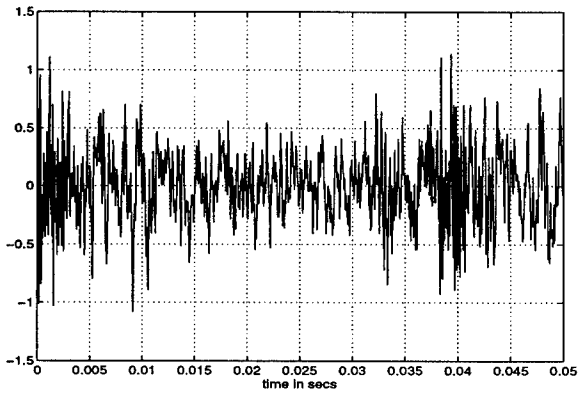


Figure 4. Measured lawnmower noise

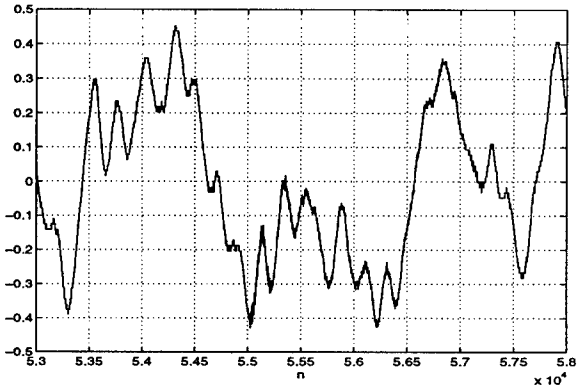


Figure 5. Typical input of prefilter

be reasonable to conclude that a simple high pass filter with a cutoff of about 30KHz should suffice to filter out this periodic noise from a Tantalum Nitride AE, which has a single dominant component at about 105 KHz.

Another possible method for performing the prefiltering is to use a linear prediction (LP) filter designed using the method in [8], using the first 8000 samples of the data. The coefficients can be updated or redesigned after a certain period of time to reflect any changes in the characteristics of the machine noise. We found that LP filtering worked better than an HP filter, for some types of AEs, (cf. Figures 5 and 6). Of course, a combination of HP filtering and LP filtering can also be considered.

4 Detection

At the output of the prefilter the AE signal, if present, is usually buried in additive noise at very low SNRs. In this section, we present results from two methods that we have considered for possible robust detection of the signal.

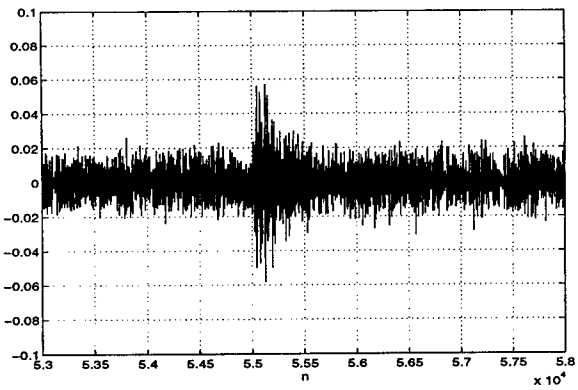


Figure 6. Output of linear prediction filter

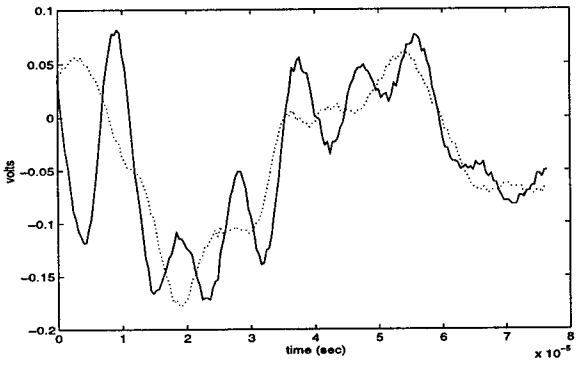


Figure 7. Tantalum Nitride signal embedded in lawnmower noise.

4.1 Optimal Tapers

We applied the techniques described in [5] to determine whether each exponentially decaying sinusoid detected via Prony's method was actually present, or a false reading produced by noise. The method consists of applying an optimal window function to the data, (optimal in the sense that it minimizes spectral leakage for constant SNR), then testing the fit of the model to the data with an F-statistic.

The optimal windows are determined by parameters P, ν, and β: We chose time-bandwidth product $P = 4$, $(NW = 8\pi)$. Experiments showed that noise parameter ν had a small effect on the final result, so we let $\nu = 0$. The exponential decay parameter β was chosen to match the value returned from Prony's method.

The method works well for high SNR. The Tantalum Nitride signal was embedded in lawnmower noise, (cf. Figure 4), as shown in Figure 7. Figure 8 shows the resulting F-statistic. There is a clear peak near 106 kHz, the frequency identified by Prony's method.

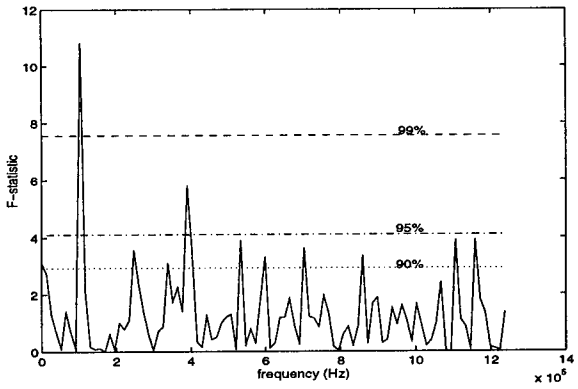


Figure 8. F-statistic for Tantalum Nitride signal embedded in lawnmower noise.

4.2 Dominant Component

At the output of the prefilter, the AE signal (if present), is usually buried in additive noise at very low SNRs. Under the assumption that both the noise and the AE can be modeled as Gaussian random vectors, we can state the detection hypothesis problem as,

$$H_0 : \vec{Y} = \vec{N} \quad , \quad H_1 : \vec{Y} = \vec{N} + \theta\vec{S} \qquad (2)$$

where $\vec{N} \sim \mathcal{N}(\vec{\mu}_n, R_n)$ is the noise vector and $\vec{S} \sim \mathcal{N}(\vec{\mu}_s, R_s)$ is the signal (AE) vector. θ is the unknown amplitude of the signal vector. No uniformly powerful test (UMP) exists for the above hypothesis with respect to θ. However, a locally most powerful (LMP) test can be found assuming low SNRs.

The LMP test statistic T_{lo} for the above hypothesis after pre-whitening of the additive Gaussian noise is given by, ([6]),

$$T_{lo}(\vec{y}) \propto \vec{y}^T R_s \vec{y} \qquad (3)$$

where $\vec{y}$ is the observation vector. This is equivalent to,

$$T_{lo}(\vec{y}) \propto \sum_{k=1}^{r} \lambda_k < \vec{y}, \vec{v}_k >^2 \qquad (4)$$

where λ_k is the k^{th} eigenvalue and $\vec{v}_k$ is the k^{th} eigenvector of R_s. r is the rank of R_s. The decision statistic T_{lo} given in (4) can be implemented with a bank of r causal, linear filters in parallel as shown in Figure 9. The impulse response of the k^{th} channel filter is given by,

$$h_k(n) = v_k(N - n) \qquad (5)$$

where $n = 1, 2...N$. N is the number of samples in a data block. The output of each channel is squared and

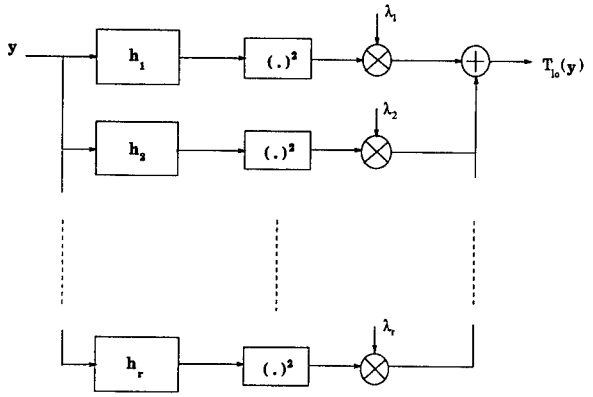


Figure 9. Implementation of test statistic

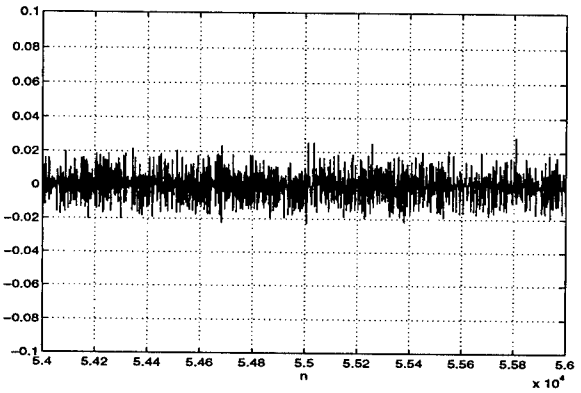


Figure 10. Output of prefilter

the test statistic is obtained by a linear combination (with the corresponding eigenvalues as the weights) of these squared outputs.

Four AEs from the Tantalum Nitride specimen were used to estimate the signal covariance matrix. The rank of the estimated covariance matrix was found to be 2. An AE signal from a Tantalum Nitride specimen was then added to the lawnmower noise to simulate a practical case, wherein the AE signal is measured at very low SNRs amidst strong correlated noise. It is important to note that the AE signal which was added to the noise was was *not* one of the AE signals used to estimate the covariance matrix. Since the rank is 2, only two channels of filtering is required. The impulse response for these two filters are given by (5). A typical output of the prefilter is shown in Figure 10. The AE (starts at n=55000) signal is buried in noise. The corresponding test statistic obtained with the above output of the prefilter is plotted in Figure 11. It is evident that the test statistic performs quite well in identifying the occurence of the acoustic emission.

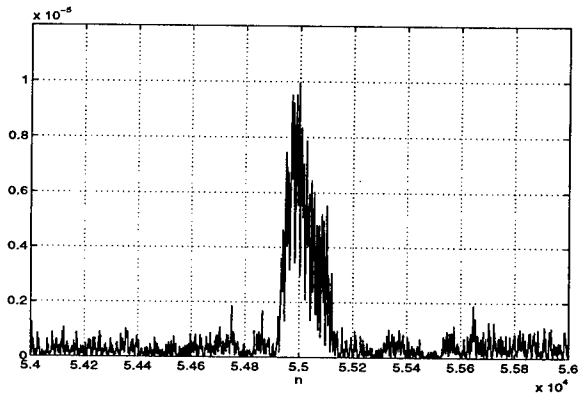


Figure 11. Test statistic (Equation 4)

5 Conclusions

Some preliminary results from our investigation into AEs as a viable technique for predictive diagnostics have been presented. HP filtering and LP filtering approaches have been considered for prefiltering. Two possible detection methods have also been considered. The choice of the technique to be used for prefiltering and detection is heavily dictated by the specific application under consideration.

Classification and crack localization are possible future directions in our research.

References

[1] R. Isermann. Fault diagnosis of machines via parameter estimation and knowledge processing - tutorial paper. *Automatica*, 29:815–835, 1993.

[2] L. Jacobs. Characterization of acoustic emission signals from mode i crack. *Journal of Engineering Mechanics*, 117(8):1878–1889, 1991.

[3] R. Kumaresan and D. W. Tufts. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-30(6):833–840, December 1982.

[4] D. Mitrakovic, I. Grabec, and S. Sedmak. Simulation of ae signals and signal analysis systems. *Ultrasonics*, 23:227–232, 1985.

[5] J. Park, C. R. Lindberg, and D. J. Thomson. Multiple-taper spectral analysis of terrestrial free oscillations: part i. *Geophysical Journal of the Royal Astronomical Society*, 91(3):755–794, December 1987.

[6] V. H. Poor. *An Introduction to Signal Detection and Estimation, Second Edition.* Springer-Verlag, 1994.

[7] C. Scruby. An introduction to acoustic emission. *J. Phys. E: Sci. Instrum.*, 20:946–953, 1987.

[8] D. W. Tufts and R. Kumaresan. Singular value decomposition and improved frequency estimation using linear prediction. *IEEE Trans on ASSP*, 30(4):671–675, August 1982.

TA-2: Multichannel ID – Source Separation

Blind System Identification Using Minimum Noise Subspace

Yingbo Hua, Karim Abed Meraim, and Mati Wax*

Department of Electrical and Electronic Engineering
The University of Melbourne, Parkville, Victoria 3052, Australia
email: yhua@ee.mu.oz.au
* Rafael 83, P.O. Box 2250, Haifa 31021 Israel

Abstract

Developing fast and robust methods for identifying multiple FIR channels driven by an unknown common source is important for wireless communications. In this letter, we present a new method that exploits a minimum noise subspace (MNS). The MNS is computed from a set of channel output pairs which form a "tree". The "tree" exploits with minimum redundancy the diversity among all channels. The MNS method is much more efficient in computation and only slightly less robust to channel noise than the subspace method by Moulines et al.

1. Introduction

Blind identification of multiple-channels FIR system driven by a common source has recently received much attention due to its potential applications in wireless communications. In contrast to the traditional cost-function based adaptive approaches and the more recent higher order statistics (HOS) based methods, the second order statistics (SOS) based methods appear to be a "hot" topic in this community, e.g., see [2]. Apparently, this trend started from the work by Tong et al [3]. Among many SOS based methods known so far, the subspace (SS) method by Moulines et al [1] is an outstanding one. The SS method applies the MUSIC concept to a relation between the channel impulse responses and the noise subspace associated with a covariance matrix of the system output. In this paper, we present a new variation of the SS method. Instead of exploiting the full noise subspace, this new method exploits a minimum noise subspace (MNS). The MNS method represents a solid extension of an observation made by Moulines et al [1] that the full noise subspace of the system output covariance matrix is generally not necessary to asymptotically yield the unique (up to a constant) estimate of channel responses. We will show that the minimum dimension of the noise subspace required for unique system identification is M-1 where M is the number of FIR channels, and each of the required M-1 noise vectors can be computed from one of M-1 covariance matrices corresponding to properly chosen M-1 (distinct) pairs of channel outputs. Any M-1 pairs of channel outputs that span a "tree" pattern (Figure 1) are a proper choice. The MNS method is much more efficient in computation than the SS method. Simulations have shown that the MNS method is only slightly less robust to channel noise than the SS method.

2. Channel model and the SS method

We consider M parallel FIR channels driven by a common source. The output vector of the i^{th} channel can be written as

$$\mathbf{y}_i(n) = \mathbf{H}_i\mathbf{s}(n) + \mathbf{w}_i(n)$$

where

$$\mathbf{y}_i(n) = [y_i(n) \quad y_i(n+1) \quad \dots \quad y_i(n+N-1)]^T$$

$$\mathbf{s}(n) = [s(n-L) \quad s(n-L+1) \quad \cdots \quad s(n+N-1)]^T$$

$$\mathbf{w}_i(n) = [w_i(n) \quad w_i(n+1) \quad \dots \quad w_i(n+N-1)]^T$$

$$\mathbf{H}_i = \begin{bmatrix} h_i(L) & \cdots & h_i(0) & 0 & \cdots & 0 \\ 0 & h_i(L) & \cdots & h_i(0) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & h_i(L) & \cdots & h_i(0) \end{bmatrix},$$

$N\times(N+L)$.

$y_i(n)$ denotes the output sequence of the ith channel; $s(n)$ the input sequence; $w_i(n)$ the noise sequence on the i^{th} channel; and $h_i(k)$ the impulse response of the i^{th} channel. L denotes the maximum order of the M channels; and N the window length on each channel output. Then we write

$$\mathbf{y}(n) = \mathbf{H}\mathbf{s}(n) + \mathbf{w}(n)$$

0-8186-7576-4/96 $5.00

where $\mathbf{y}(n)=\begin{bmatrix}\mathbf{y}_1(n)\\ \vdots \\ \mathbf{y}_M(n)\end{bmatrix}$, $\mathbf{w}(n)=\begin{bmatrix}\mathbf{w}_1(n)\\ \vdots \\ \mathbf{w}_M(n)\end{bmatrix}$, and $\mathbf{H}=\begin{bmatrix}\mathbf{H}_1\\ \vdots \\ \mathbf{H}_M\end{bmatrix}$. The matrix $\mathbf{H}$ is known as $MN\times(N+L)$ generalized Sylvester matrix [6] which has the full column rank $N+L$ under the assumptions: **A1)** the M channels do not share a common zero; and **A2)** $N\geq L+1$. The blind identification problem here is to find $\mathbf{H}$ from the sequence $\{\mathbf{y}(n)$ for $n=1,2,...,T\}$. The SS method [1] exploits the covariance matrix of all channel outputs: $\mathbf{R}_y=\frac{1}{T}\sum_{n=1}^{T}\mathbf{y}(n)\mathbf{y}(n)^H$ where ^{H} denotes the conjugate transpose. This matrix has the inherent structure: $\mathbf{R}_y=\mathbf{H}\mathbf{R}_s\mathbf{H}^H+\mathbf{R}_w$ with $\mathbf{R}_s=\frac{1}{T}\sum_{n=1}^{T}\mathbf{s}(n)\mathbf{s}(n)^H$ and $\mathbf{R}_w=\frac{1}{T}\sum_{n=1}^{T}\mathbf{w}(n)\mathbf{w}(n)^H$. The SS method then computes the eigendecomposition of $\mathbf{R}_y$: $\mathbf{R}_y=\begin{bmatrix}\mathbf{U}_s & \mathbf{U}_n\end{bmatrix}\begin{bmatrix}\Sigma_s & \\ & \Sigma_n\end{bmatrix}\begin{bmatrix}\mathbf{U}_s & \mathbf{U}_n\end{bmatrix}^H$ where the matrix $\mathbf{U}_n$ consists of the $MN-N-L$ non-principal eigenvectors of $\mathbf{R}_y$. In addition to the assumptions **A1-A2**, if **A3)** the source covariance matrix $\mathbf{R}_S$ has the full rank $N+L$, and **A4)** the noise covariance matrix $\mathbf{R}_W$ is proportional to the identity matrix (which is true when the noise is white and T is very large), then it can be shown [1] that range($\mathbf{U}_n$) is the orthogonal complement of range($\mathbf{H}$). Hence, range($\mathbf{U}_n$) is referred to as the noise subspace. The SS method yields an estimate $\mathbf{H}_e$ of $\mathbf{H}$ by solving the equation $\mathbf{U}_n{}^{\mathrm{H}}\mathbf{H}_e=\mathbf{0}$ in a least square sense (where $\mathbf{H}_e$ is subject to the same structure as $\mathbf{H}$). This estimate is uniquely (up to a constant scalar) equal to $\mathbf{H}$ under the assumptions **A1-A4** [1].

3. The MNS method

In the MNS method, we first select M-1 distinct pairs from the M channel outputs $\{\mathbf{y}_i(n), i=1,...,M\}$. The M-1 pairs must span a "tree" which connects all M channel outputs. The channel outputs are the "nodes" of the tree as shown in Figure 1.

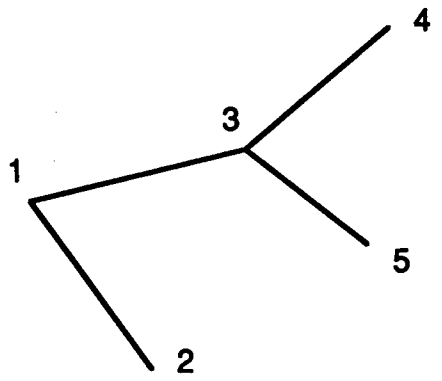


Figure 1: This illustrates a "tree" which connects $M=5$ channel outputs as its "nodes". A tree must have no loop and connect all its nodes. Here, the nodes 2, 4, and 5 are "ending" nodes, and the nodes 1 and 3 are "branching" nodes. (The tree spanned by M-1 pairs of channel outputs is the same as the tree by M-1 pairs of the columns of $\overline{\mathbf{H}}(z)$ in the proof for Lemma 3.)

Then for each pair of channel outputs, we compute the covariance matrix $\mathbf{R}_y^{i,j}=\frac{1}{T}\sum_{n=1}^{T}\begin{bmatrix}\mathbf{y}_i(n)\\ \mathbf{y}_j(n)\end{bmatrix}\begin{bmatrix}\mathbf{y}_i(n)\\ \mathbf{y}_j(n)\end{bmatrix}^H$ and its least dominant eigenvector $\tilde{\mathbf{v}}^{i,j}$. Let $\tilde{\mathbf{v}}^{i,j}=\begin{bmatrix}\tilde{\mathbf{v}}^{i,j}(i)\\ \tilde{\mathbf{v}}^{i,j}(j)\end{bmatrix}$ where each subvector has the dimension $N\times 1$. Then define "zero padded" vector $\mathbf{v}^{i,j}=\begin{bmatrix}\mathbf{v}^{i,j}(1)\\ \vdots \\ \mathbf{v}^{i,j}(M)\end{bmatrix}$ where $\mathbf{v}^{i,j}(k)=\begin{cases}\tilde{\mathbf{v}}^{i,j}(i) & k=i\\ \tilde{\mathbf{v}}^{i,j}(j) & k=j\\ \mathbf{0} & otherwise\end{cases}$. Then we form a $MN\times(M-1)$ matrix $\mathbf{V}_n$ of the M-1 vectors $\{\mathbf{v}^{i,j}\}$. Similar to the SS method, the MNS method yields an estimate $\mathbf{H}_e$ of $\mathbf{H}$ by solving the equation $\mathbf{V}_n{}^{\mathrm{H}}\mathbf{H}_e=\mathbf{0}$ in a least square sense (where $\mathbf{H}_e$ is subject to the same structure as $\mathbf{H}$). The significant compuational advantage of the MNS method over the SS method is obvious. In particular, the SS method requires a full eigendecomposition of an $MN\times MN$ matrix, but the MNS method computes the single least dominant eigenvector of a $2N\times 2N$ matrix in parallel for each of M-1 pairs of channel outputs.

We will now establish that under the assumptions **A1-A4**, a) the MNS method yields the unique estimate of $\mathbf{H}$, and b) M-1 is the smallest number of vectors from the noise subspace in order for an equation like $\mathbf{V}_n{}^{\mathrm{H}}\mathbf{H}_e=\mathbf{0}$ to yield the unique estimate of $\mathbf{H}$. We note that due to the

limited space, the proofs shown below might be too brief for some readers.

Lemma 1 (easy to prove): For any equation $\mathbf{v}^H\mathbf{H}=\mathbf{0}$ where $\mathbf{v}=\left[\mathbf{v}(1)^T \cdots \mathbf{v}(M)^T\right]^T$ with $\mathbf{v}(i)=\left[v_i(0) \cdots v_i(N-1)\right]^T$ and $\mathbf{H}$ is a $MN\times(N+L)$ generalized Sylvester matrix, there uniquely corresponds a polynomial equation $\sum_{i=1}^{M} V_i(z)H_i(z)=0$ of degree $N+L-1$, where $H_i(z)=\sum_{l=0}^{L} h_i(l)z^{-l}$ of degree L and $V_i(z)=\sum_{l=0}^{N-1} v_i(l)z^{-l}$ of degree $N-1$. The converse is also true.

Lemma 2: If there are q $MN\times1$ vectors $\{\mathbf{v}_i$ for $i=1,...,q\}$ such that $\{\mathbf{v}_i{}^H\mathbf{H}=\mathbf{0}$ for $i=1,...,q\}$ where $\mathbf{H}$ is a $MN\times(N+L)$ generalized Sylvester matrix, then $\mathbf{H}$ is (possibly) unique up to a constant scalar *only if* $q\geq M-1$.

Proof: Using Lemma 1, it is straightforward to show that $\{\mathbf{v}_i{}^H\mathbf{H}=\mathbf{0}$ for $i=1,...,q\}$ is equivalent to the polynomial matrix equation $\overline{\mathbf{V}}(z)\overline{\mathbf{h}}(z)=\overline{\mathbf{0}}$ of degree $N+L-1$, where $\overline{\mathbf{V}}(z)$ is a $q\times M$ polynomial matrix of degree $N-1$ uniquely corresponding to $\{\mathbf{v}_i$ for $i=1,..,q\}$ and $\overline{\mathbf{h}}(z)$ is an $M\times1$ polynomial vector of degree L uniquely corresponding to $\mathbf{H}$. But using the polynomial matrix theory [5], $\overline{\mathbf{h}}(z)$ is determined by the equation $\overline{\mathbf{V}}(z)\overline{\mathbf{h}}(z)=\overline{\mathbf{0}}$ uniquely up to a polynomial (or constant) scalar *only if* $q\geq M-1$.

It is easy to show that under the assumptions **A1-A4**, the vector $\mathbf{v}^{i,j}$ satisfies $(\mathbf{v}^{i,j})^H\mathbf{H}=\mathbf{0}$. Since the MNS method only relies on $M-1$ noise vectors, Lemma 2 has now established that the MNS method exploits a "minimum" noise subspace.

Lemma 3: The MNS method yields the unique (up to a constant scalar) estimate of the channel responses under the assumptions **A1-A4**.

Proof: From Lemma 1, the equation $(\mathbf{v}^{i,j})^H\mathbf{H}=\mathbf{0}$ is equivalent to a polynomial equation $V_j(z)H_i(z)+V_i(z)H_j(z)=0$ of degree $N+L-1$, where $V_i(z)$ and $V_j(z)$ are of degree $N-1$ and $H_i(z)$ and $H_j(z)$ are of degree L. Similarly, each sub-equation $(\mathbf{v}^{i,j})^H\mathbf{H}_e=\mathbf{0}$ in the overall MNS estimation equation $\mathbf{V}_n{}^H\mathbf{H}_e=\mathbf{0}$ is equivalent to a polynomial equation $V_j(z)H_{e_i}(z)+V_i(z)H_{e_j}(z)=0$ where the degrees of all polynomials are the same as in the previous polynomial equation. Combining these two polynomial equations yields $H_j(z)H_{e_i}(z)-H_i(z)H_{e_j}(z)=0$. Using this equation for each of the $M-1$ pairs of channels, one can show that the solution to $\mathbf{V}_n{}^H\mathbf{H}_e=\mathbf{0}$ is equivalent to that of the polynomial matrix equation $\overline{\mathbf{H}}(z)\overline{\mathbf{h}}_e(z)=\overline{\mathbf{0}}$ of degree $2L$, where $\overline{\mathbf{H}}(z)$ is an $(M-1)\times M$ polynomial matrix of degree L uniquely corresponding to $\{H_i(z)$ for $i=1,...,M\}$, and $\overline{\mathbf{h}}_e(z)$ is an $M\times1$ polynomial vector of degree L uniquely corresponding to $\{H_{ei}(z)$ for $i=1,...,M\}$ (or equivalently $\mathbf{H}_e$). Note that each row of $\overline{\mathbf{H}}(z)$ only has two nonzero entries and hence defines a pair of columns. The $M-1$ pairs of columns defined by the $M-1$ rows of $\overline{\mathbf{H}}(z)$ also span a "tree" which connects all M columns of $\overline{\mathbf{H}}(z)$ as its "nodes". This tree is identical to the tree spanned by the pairs of channel outputs (Figure 1). Because of this structure in $\overline{\mathbf{H}}(z)$, one can show by induction that $\overline{\mathbf{H}}(z)$ has the full row rank $M-1$. (Note that removing a column and a row of $\overline{\mathbf{H}}(z)$ associated with an "ending node" decreases the rank of $\overline{\mathbf{H}}(z)$ by one, and when $\overline{\mathbf{H}}(z)$ is 1×2 its rank is one.) Therefore, the solution for the $M\times1$ vector $\overline{\mathbf{h}}_e(z)$ to the equation $\overline{\mathbf{H}}(z)\overline{\mathbf{h}}_e(z)=\overline{\mathbf{0}}$ must be unique up to a polynomial scalar [5]. Furthermore, since $\overline{\mathbf{h}}(z)$ is a solution of degree L to $\overline{\mathbf{H}}(z)\overline{\mathbf{h}}_e(z)=\overline{\mathbf{0}}$ and there is no common zero among all channels (**A1**), $\overline{\mathbf{h}}(z)$ must be the unique solution up to a constant scalar.

Lemma 3 has established that the MNS method yields asymptotically the unique estimate of $\mathbf{H}$. This section has provided a much stronger result than a discussion in [4] on the MNS method.

4. Performance of the MNS method

In our simulation, we used a system of four ($M=4$) parallel FIR channels. The first channel is given by the GSM test channel [7] with 6 ($L=5$) delayed paths. The other three channels are generated by assuming a plane propagation model for each path with corresponding electric angles uniformly distributed in $[0, \pi/3]$. A realization of the

channel impulse responses is shown in the table shown below. The output observation noise is an i.i.d. sequence of zero-mean Gaussian variables. The input signal is an i.i.d. sequence of zero-mean, unit-variance QAM-4 variables independent from the noise. The performance is measured by $MSE(dB) = 10\log_{10}\left\{\frac{1}{N_r}\sum_{r=1}^{N_r}\|\mathbf{h}_r - \mathbf{h}\|^2\right\}$ where N_r is the number of independent runs (N_r=100), $\mathbf{h}$ is the true (unit-norm) vector of the impulse responses $\{h_i(k)$ for i=1,..,M and k=0,...,$L\}$, $\mathbf{h}_r$ is the estimated (unit-norm) vector of impulse responses at the r^{th} run. (The equation $\mathbf{V}_n^H\mathbf{H}_e$=0 was solved subject to $\|\mathbf{h}_e\| = 1$. For each run, $\mathbf{h}_r$=$\alpha\mathbf{h}_e$ where α=$\mathbf{h}_e^H\mathbf{h}$ is a phase adjuster.) The signal-to-noise ratio is defined as $SNR(dB) = 20\log_{10}\left(\frac{\|\mathbf{h}\|\sigma_s}{\sqrt{M}\sigma_w}\right)$ where σ_S and σ_W denote the deviations of the input and the noise respectively. Figure 2 compares the performances of the SS and NMS methods. This figure (associated with the case defined by the table) is quite typical among all the cases that we considered in our simulation. In the operational region where the *MSE* is relatively small, the MNS method required *SNR* no more than 3 dB higher than the SS method to yield a given value of *MSE*.

References

[1] E. Moulines, P. Duhamel, J. Cardoso and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. on SP*, pp. 516-525, Feb 1995.

[2] *Proceedings of IEEE ICASSP '94 and '95.*

[3] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," *Proc of the 25th Asilomar Conference*, CA, pp. 856-860, Nov. 1991.

[4] Y. Hua, H. Yang, M. Wax and M. Zhou, "Blind system identification using multiple sensors," *Proc of IEEE ICASSP'95*, pp.3171-3174, May 1995.

[5] G. Forney, "Minimal bases of rational vector spaces with applications to multivariable linear systems", *SIAM J. on Control*, Vol. 13, pp.493-520, May 1975.

[6] R.R.Bitmead, S.-Y. Kung, B. Anderson, and T. Kailath, "Greatest common division via generalized Sylvester and Bezout matrices," *IEEE Trans. on AC*, Vol. 23, No.6, pp. 1043-1047, Dec. 1978.

[7] J. Proakis, *Digital Communications*. New York: McGraw-Hill, 1989.

[8] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," IEEE Transactions on Signal Processing, March 1996.

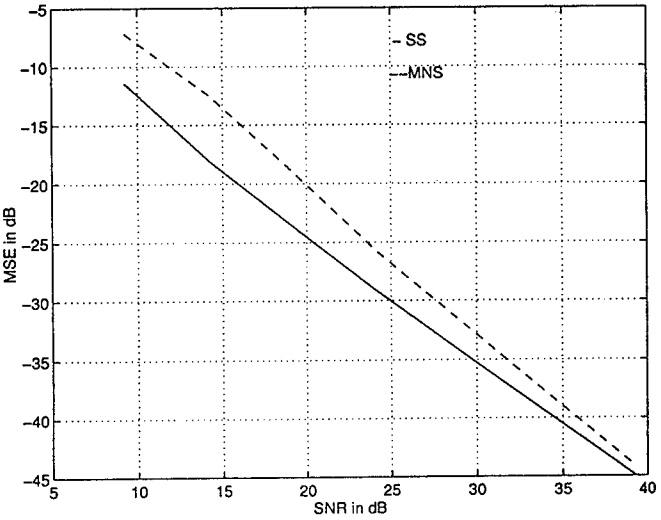


Figure 2: Performance comparison of the SS method and the MNS method. *MSE* versus *SNR*.

Acknowledgement: This work was supported by the Australian Research Council and the Australian Cooperative Research Centre on Sensor Signal and Information Processing.

	$h_1(k)$	$h_2(k)$	$h_3(k)$	$h_4(k)$
k=0	0.4972-1.2784i	1.3516-0.2333i	0.8970+1.0377i	-0.4264+1.3037i
k=1	-0.0370+0.7256i	-0.5251+0.5021i	0.7265+0.0046i	-0.5314-0.4955i
k=2	1.4158+0.2768i	0.8012+1.1996i	-0.2867+1.4138i	-1.2052+0.7927i
k=3	0.6417+0.4440i	0.2181+0.7492i	-0.3031+0.7190i	-0.6886+0.3669i
k=4	-1.2418+0.5984i	-1.2837+0.5023i	-1.3182+0.4032i	-1.3450+0.3018i
k=5	0.0235+1.1777i	-0.7049+0.9438i	-1.1360+0.3118i	-1.0879-0.4518i

Table: Impulse responses of the M-channel system. Other parameters are M=4, L=5, N=6, T=245.

On Blind Separation of Convolutive Mixtures of Independent Linear Signals

Jitendra K. Tugnait
Dept. of Electrical Engineering
Auburn University
Auburn, Alabama 36849, USA
tugnait@eng.auburn.edu

ABSTRACT

This paper is concerned with the problem of blind separation of independent signals (sources) from their linear convolutive mixtures. The problem consists of recovering the sources up to shaping filters from the observations of MIMO system output. The various signals are assumed to be linear but not necessarily i.i.d. (independent and identically distributed). An iterative, normalized higher-order cumulant maximization based approach is developed using the third-order and/or fourth-order normalized cumulants of the "beamformed" data. The approach is source-iterative, i.e., the sources are extracted (at each sensor) and cancelled one-by-one. The proposed solution provides a decomposition of the given data at each sensor into its independent signal components. The proposed approach is an extension/application of a recently proposed approach for MIMO system identification where the system is driven by unobserved i.i.d. inputs.

1 Introduction

Given measurements $y_i(k)$, $(i = 1, 2, \cdots, N)$, at time k at N sensors, let these measurements be a linear convolutive mixture of M source signals $x_j(k)$, $(j = 1, 2, \cdots, M)$:

$$y_i(k) = \sum_{j=1}^{M} G_{ij}(q^{-1})x_j(k), \quad i = 1, 2, \cdots, N, \qquad (1-1)$$

$$\Rightarrow \quad \mathbf{y}(k) = \mathbf{G}(q^{-1})\mathbf{x}(k), \qquad (1-2)$$

where ij−th element of $\mathbf{G}(q^{-1})$ is $G_{ij}(q^{-1})$, $\mathbf{y}(k) = [y_1(k) \vdots y_2(k) \vdots \cdots \vdots y_N(k)]^T$, similarly for $\mathbf{x}(k)$, q^{-1} is the backward-shift operator (i.e., $q^{-1}x(k) = x(k-1)$, etc.), $x_j(k)$ is the j-th input at sampling time k, $y_i(k)$ is the i-th output, and

$$G_{ij}(q^{-1}) := \sum_{l=-\infty}^{\infty} g_{ij}(l)q^{-l} \qquad (1-3)$$

is the scalar transfer function with $x_j(k)$ as the input and $y_i(k)$ as the output. We will also use the notation

$$\widetilde{G}_{ij}(z) := G_{ij}(q^{-1})|_{q=z} = \sum_{l} g_{ij}(l)z^{-l}, \qquad (1-4)$$

the $\mathcal{Z}$-transform of the sequence $\{g_{ij}(l)\}_{l=-\infty}^{\infty}$. We allow all of the above variables to be complex-valued.

We wish to design a MIMO dynamic system $\mathbf{E}(q^{-1})$ with N inputs and M outputs such that the overall $M \times M$ system

$$\mathbf{T}(q^{-1}) := \mathbf{E}(q^{-1})\mathbf{G}(q^{-1}) \qquad (1-5)$$

decouples the source signals. Following the 2×2 case considered in [7], this implies that we must have

$$T_{ij}(q^{-1}) \begin{array}{ll} = 0 & \text{for} \quad i \neq i_j \\ \neq 0 & \text{for} \quad i = i_j \end{array} \qquad (1-6)$$

where $i = 1, 2, \cdots, M$; $j = 1, 2, \cdots, M$ and $i_j \in \{1, 2, \cdots, M\}$ such that $i_j \neq i_l$ for $j \neq l$. That is, in every column and every row of $\mathbf{T}(q^{-1})$ there is exactly one non-zero entry. In a blind separation problem, the nonzero entries of $\mathbf{T}(q^{-1})$ are allowed to be a scalar linear system (shaping filter), unlike the equalization problems where they must be constant gains.

The problem considered above arises in a wide variety of signal processing and communications applications; see [1]-[8], and references therein. One obvious application is array signal processing where the array manifold may be unknown or imprecisely known [5]. Separation of sources differs from blind equalization [9],[10],[13],[14],[17] in that the source signals are not necessarily i.i.d. (independent and identically distributed). In this paper we allow $N \geq M$ (N = number of sensors, M= number of sources) with M arbitrary, whereas quite a few existing papers are restricted to $M = N = 2$ ([1],[7],[8]) or $M = N$ ([2],[3],[17]). Our proposed approach has aspects that follow from [14] (see also [16]), yet our approach is more general in that it applies to signals with nonzero normalized fourth cumulant whereas [14] and [16] are restricted to signals with negative normalized fourth cumulant. Moreover, [16] deals with instantaneous mixtures of a restricted type and [14] deals with blind equalization, not source separation.

2 Model Assumptions

We impose the following conditions:

(AS1) $N \geq M$, at least as many outputs as inputs.

(AS2) The various components of $\mathbf{x}(k)$ are mutually independent and the coupling system is stable.

This work was supported by the National Science Foundation under Grant MIP-9312559.

0-8186-7576-4/96 $5.00 © 1996 IEEE

(AS3) $\mathbf{x}(k)$ is linear, i.e.

$$\mathbf{x}(k) \;=\; \mathbf{F}(q^{-1})\mathbf{w}(k), \qquad (2-1)$$

where $\mathbf{w}(k)$ is a zero-mean, $M-$vector stationary non-Gaussian process, temporally i.i.d. and spatially independent, with nonzero fourth cumulants. Because of **(AS2)** we may take $\mathbf{F}(q^{-1})$ to be diagonal. Assume also that the composite system

$$\mathbf{y}(k) \;=\; \mathbf{G}(q^{-1})\mathbf{F}(q^{-1})\mathbf{w}(k) \;=:\; \mathbf{B}(q^{-1})\mathbf{w}(k), \qquad (2-2)$$

is stable. Let $\mathcal{B}(z)$ denote the transfer function $\mathbf{B}(q^{-1})$ in the $\mathcal{Z}$-transform notation. Assume that $\text{rank}\{\mathcal{B}(z)\} = M$ for any $|z| = 1$.

We will denote the $ij-$th element of $\mathbf{B}(q^{-1})$ is $B_{ij}(q^{-1})$.

3 A Solution

Let $\text{CUM}_4(w)$ denote the fourth-order cumulant of a complex-valued random variable w, defined as

$$\text{CUM}_4(w) \;:=\; \text{cum}_4\{w, w^*, w, w^*\} \qquad (3-1)$$

$$\;=\; E\{|w|^4\} - 2[E\{|w|^2\}]^2 - |E\{w^2\}|^2. \qquad (3-2)$$

We will use the notation $\gamma_{4wi} = \text{CUM}_4(w_i(k))$ and $\sigma_{wi}^2 = E\{|w_i(k)|^2\}$. Consider an $1 \times N$ row-vector polynomial equalizer $\mathbf{C}^T(q^{-1})$, with its j-th entry denoted by $C_j(q^{-1})$, operating on the data vector $\mathbf{y}(k)$. Let the equalizer output be denoted by $e(k)$. We then have

$$e(k) \;=\; \sum_{i=1}^{N} C_i(q^{-1})y_i(k)$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{M} C_i(q^{-1})B_{ij}(q^{-1})w_j(k) \;=\; \sum_{j=1}^{M} H_j(q^{-1})w_j(k) \qquad (3-3)$$

where

$$H_j(q^{-1}) \;:=\; \sum_{i=1}^{N} C_i(q^{-1})B_{ij}(q^{-1}) \qquad (3-4)$$

so that

$$\mathbf{H}(q^{-1}) \;:=\; \mathbf{C}^T(q^{-1})\mathbf{G}(q^{-1})\mathbf{F}(q^{-1}).$$

In general, we have

$$H_j(q^{-1}) \;=\; \sum_{k=-\infty}^{\infty} h_j(k)q^{-k}. \qquad (3-5)$$

Define $\overline{h}_j(k) = \sigma_{wj}h_j(k)$, $\overline{\gamma}_{4wj} = \gamma_{4wj}/\sigma_{wj}^4$ and $|\overline{\gamma}_{4max}| := \max_{1\le j\le M}|\overline{\gamma}_{4j}|$.

As in [13], we propose to consider maximization of the cost

$$J_{42} \;:=\; \frac{|\text{CUM}_4(e(k))|}{[E\{|e(k)|^2\}]^2} \qquad (3-6)$$

for designing a linear equalizer to recover one of the inputs. It can be shown [13] that

$$J_{42} \;\le\; |\overline{\gamma}_{4max}| \qquad (3-7)$$

with equality iff

$$h_j(k) \;=\; d\delta(k-k_0)\delta(j-j_0), \quad j_0 \in \{1, 2, \cdots, M\}, \quad (3-8)$$

where d is some complex constant, k_0 is some integer, j_0 indexes some input out of the given M inputs such that $|\overline{\gamma}_{4j_0} = |\overline{\gamma}_{4max}|$, and $\delta(k-k_0) = 1$ if $k = k_0$, $= 0$ otherwise. Thus, (3-3) reduces to

$$e(k) \;=\; dw_{j_0}(k-k_0), \qquad (3-9)$$

i.e., the equalizer output is a possibly scaled and shifted version of one of the system inputs. It has been established in [13] that under **(AS1)**-**(AS3)**, such a solution exists and all locally stable stationary points of the given cost w.r.t. the combined composite channel-equalizer impulse response $\{\overline{h}_j(k)\}$ are characterized by solutions such as (3-8) and (3-9). Moreover, if doubly-infinite equalizers are used then all locally stable stationary points of the given cost w.r.t. the equalizer coefficients are also characterized by solutions such as (3-8) and (3-9).

The above discussion suggests an iterative solution where we iterate on input sequences one-by-one. Maximization of (3-6) w.r.t. the equalizer $\mathbf{C}(q^{-1})$ leads to the solution (3-9) under the sufficient conditions **(AS1)**-**(AS3)**. Given (3-9) we can estimate and remove the contribution of $w_{j_0}(k)$ from (1-1). Then we have a MIMO system with N outputs but $M-1$ inputs (instead of M inputs as in (1-1)-(1-2)). Repeat the process, i.e., maximize (3-6) w.r.t. a new equalizer to get a solution $e(k) = d'w_{j_0'}(k-k_0')$ where $j_0' \in (\{1, 2, \cdots, M\} - \{j_0\})$. That is, we follow the following procedure.

Step 1. Maximize (3-6) w.r.t. the equalizer $\mathbf{C}(q^{-1})$ to obtain (3-9).

Step 2. Cross-correlate $\{e(k)\}$ (of (3-9)) with the given data (1-1)-(2-2) and define a possibly scaled and shifted estimate of $b_{ij_0}(\tau)$ as

$$\widehat{b}_{ij_0}(\tau) \;:=\; \frac{E\{y_i(k)e^*(k-\tau)\}}{E\{|e(k)|^2\}}. \qquad (3-10)$$

Consider now the reconstructed contribution of $e(k)$ to the data $y_i(k)$ $(i = 1, 2, \cdots, M)$, denoted by $\widehat{y}_{i,j_0}(k)$:

$$\widehat{y}_{i,j_0}(k) := \sum_{l} \widehat{b}_{ij_0}(l)e(k-l). \qquad (3-11)$$

Step 3. Remove the above contribution from the data to define the outputs of a MIMO system with N outputs and $M-1$ inputs. These are given by

$$y_i'(k) \;:=\; y_i(k) - \widehat{y}_{i,j_0}(k). \qquad (3-12)$$

Step 4. If $M > 1$, set $M \leftarrow M-1$, $y_i(k) \leftarrow y_i'(k)$, and go back to **Step 1**, else quit.

In practice, all the expectations in (3-10) are replaced with their sample averages over appropriate data records.

Analyzing the above algorithm we have

$$E\{y_i(k)e^*(k-\tau)\} = \sum_{j=1}^{M} B_{ij}(q^{-1})E\{w_j(k)e^*(k-\tau)\}$$

$$= b_{ij_0}(k_0+\tau)d^*\sigma^2_{wj_0}. \qquad (3-13)$$

Using (3-13) in (3-10) we have

$$\widehat{b}_{ij_0}(\tau) = \frac{b_{ij_0}(k_0+\tau)d^*\sigma^2_{wj_0}}{|d|^2\sigma^2_{wj_0}} = b_{ij_0}(k_0+\tau)/d\,. \qquad (3-14)$$

It follows from (3-11) and (3-14) that

$$\widehat{y}_{i,j_0}(k) = \sum_l b_{ij_0}(l)w_{j_0}(k-l). \qquad (3-15)$$

Now use (3-12) and (3-15) to deduce that

$$y'_i(k) = \sum_{j=1, j\neq j_0}^{M} B_{ij}(z)w_j(k), \quad i=1,2,\cdots,N. \quad (3-16)$$

It is seen that we have decomposed the observations at the various sensors into its independent components: $\widehat{y}_{i,j_0}(k)$ in (3-11) represents the contribution of $\{x_{j_0}(k)\}$ to the $i-$th sensor. Eqn. (3-11) represents an embarrassment of riches: we have a large class of solutions to the problem of blind separation of convolutive mixtures. We emphasize that in our solution knowledge of $\mathbf{F}(q^{-1})$, $\mathbf{G}(q^{-1})$ or $\mathbf{B}(q^{-1})$ has not been assumed. Our solution is guaranteed to converge unlike that of [7].

Remark 1. We may replace the cost (3-6) with [13]

$$J_{32} := \frac{|\mathrm{CUM}_3(e(k))|}{[E\{|e(k)|^2\}]^{1.5}} \qquad (3-17)$$

where

$$\mathrm{CUM}_3(w) := \mathrm{cum}_4\{w, w^*, w\} = E\{|w|^2 w\}. \qquad (3-18)$$

The preceding discussion pertaining to (3-6) holds in this case with obvious modifications provided we replace the phrase "nonzero fourth cumulants" in **(AS3)** with the phrase "nonzero third cumulants." □

Remark 2. It follows from the preceding developments that under the conditions **(AS1)**-**(AS3)**, the proposed iterative approach is capable of blind identification of a MIMO transfer function $\mathcal{B}(z)$ up to a time-shift, a scaling and a permutation matrix provided that we allow doubly-infinite equalizers. That is, given $\mathcal{B}(z)$, we end up with a $\mathcal{A}(z)$ where the two are related via

$$\mathcal{A}(z) = \mathcal{B}(z)\mathbf{DAP} \qquad (3-19)$$

where $\mathbf{D}$ is an $M \times M$ "time-shift" diagonal matrix (recall k_0 in (3-8)), $\mathbf{A}$ is an $M \times M$ diagonal scaling matrix (recall d in (3-8)), and $\mathbf{P}$ is an $M \times M$ permutation matrix (recall j_0 in (3-8), we don't "know" which input it refers to). See also [13]. □

4 Simulation Example

Here we consider a 2-input 3-output MA(6) system model resulting in N=3 and M=2 in (2-2). Its 3×2 transfer $\mathbf{B}(z)$ in (2-2) is given by ($\mathbf{B}_i(q^{-1})$ denotes the i-th column of $\mathbf{B}(q^{-1})$)

$$\mathbf{F}_1(q^{-1}) = \begin{bmatrix} 0.7426 + 0.7426q^{-2} \\ 0.4456q^{-1} + 0.7426q^{-2} \\ 0.8911q^{-2} + 0.5941q^{-3} \end{bmatrix}, \qquad (4-1)$$

$$\mathbf{F}_2(q^{-1}) = \begin{bmatrix} (0.5678 + 0.3407q^{-1}) \\ (-0.2385q^{-1} - 0.5678q^{-2} + 0.8176q^{-3} \\ \quad +0.4088q^{-4} + 0.2385q^{-6}) \\ (0.6814q^{-1} + 0.9085q^{-2}) \end{bmatrix}. \qquad (4-2)$$

The inputs $\{w_j(k)\}$ $(j=1,2)$ in (2-1) and (2-2) are mutually independent, zero-mean i.i.d., 4-QAM sequences taking values $\pm 1 \pm j$ with probability 0.25 each. The additive noise at the various sensors was complex (circularly symmetric) zero-mean white Gaussian with identical variance. The equalizer length was chosen to be 15 taps, i.e., with $N=3$, $C_i(q^{-1})$ $(i=1,2,3)$ have 15 taps each. The initial guess for optimization of J_{42} was always taken to be center-tap initialization, i.e. we took $c_1(7)=1$ with the remaining taps $c_i(k)$ $(i=1,2,3)$ set to zero. For the purpose of impulse response estimation and extracted signal cancellation (see steps 2 and 3 in Sec. 3), $\widehat{b}_{ij_0}(\tau)$ was estimated for $-20 \leq \tau \leq 20$ (see (3-10)).

It is clear from (3-11) that how well one estimates the channel impulse response strongly influences how well one can separate the given observations into their constituent independent components. Therefore, we will take accuracy in impulse response estimation as a performance measure. In order to assess the performance of the proposed approach, one first needs to remove the ambiguities associated with the matrices $\mathbf{D}$, $\mathbf{A}$ and $\mathbf{P}$ in (3-19). This was accomplished by aligning (via cross-correlation and shifting) the estimated impulse responses with their true counterparts and by scaling them to have a fixed norm. For instance, the true model (4-1)-(4-2) is such that

$$\sum_{i=1}^{3}\sum_{k=0}^{6} |b_{ij}(k)|^2 = 3 \quad \text{for } j=1,2. \qquad (4-3)$$

We chose to truncate the estimated impulse responses to 12 samples (after proper alignment with the true impulse responses); this is much longer than the true length of 7. The estimated impulse response $\widehat{b}_{ij}(k)$ after truncation was normalized in a manner similar to (4-3):

$$\sum_{i=1}^{3}\sum_{k=-1}^{10} |\widehat{b}_{ij}(k)|^2 = 3 \quad \text{for } j=1,2. \qquad (4-4)$$

We will use the normalized mean-square error (NMSE) in estimating the channel impulse responses as a performance index. The length of each subchannel ($B_{ij}(q^{-1})$ $(i,j \in \{1,2\})$) was restricted to 12. For M_c Monte Carlo runs,

the NMSE_{ij} for subchannel $B_{ij}(q^{-1})$ is defined as

$$\mathrm{NMSE}_{ij} = \frac{M_c^{-1}\sum_{l=1}^{M_c}\left[\sum_{\tau=-1}^{10}\left|\widehat{b}_{ij}^{(l)}(\tau) - f_{ij}(\tau)\right|^2\right]}{\sum_{\tau=-1}^{10}|b_{ij}(\tau)|^2} \qquad (4-5)$$

where $\widehat{b}_{ij}^{(l)}$ denote the estimate of the ij-th subchannel impulse response for the l-th Monte Carlo run. The overall NMSE (called ONMSE) is obtained by averaging over all subchannels:

$$\mathrm{ONMSE} = (MN)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{M}\mathrm{NMSE}_{ij}. \qquad (4-6)$$

Table I shows the various NMSE's for different SNR's and record lengths for using the cost J_{42}. It is seen that the proposed method works well even for rather low average SNR of 13 dB.

TABLE I. *Normalized mean-square error (4-6) in estimating the system matrix channel impulse response. 4-QAM (complex-valued) inputs and cost J_{42}. 50 Monte Carlo runs, equalizer length = 15 taps (per subchannel)*

Table I				
Record Length	SNR 33 dB	23 dB	13 dB	3 dB
750	0.0161	0.0167	0.0271	0.5201
1500	0.0080	0.0082	0.0102	0.5091
3000	0.0039	0.0040	0.0053	0.1973

5 Conclusions

The problem of blind separation of independent linear signals (sources) from their linear convolutive mixtures was considered. An iterative, normalized higher-order cumulant maximization based approach was developed using the third-order and/or fourth-order normalized cumulants of the "beamformed" data. The approach is source-iterative, i.e., the sources are extracted (at each sensor) and cancelled one-by-one. The proposed solution provides a decomposition of the given data at each sensor into its independent signal components. The proposed approach is an extension/application of a recently proposed approach for MIMO system identification where the system is driven by unobserved i.i.d. inputs.

6 References

[1] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromorphic architecture," *Signal Processing*, vol. 24, pp. 1-10, 1991.

[2] J.L. Lacoume and P. Ruiz, "Separation of independent sources from correlated inputs," *IEEE Trans. Signal Processing*, vol. SP-40, pp. 3074-3078, Dec. 1992.

[3] E. Moreau and O. Macchi, "New self-adaptive algorithms for source separation based on contrast functions," in *Proc. IEEE Signal Proc. Workshop on Higher-Order Statistics*, South Lake Tahoe, CA, pp. 215-219, June 1993.

[4] L.Tong, Y. Inouye and R. Liu, "Waveform-preserving blind estimation of multiple independent sources," *IEEE Trans. Signal Processing*, vol. SP-41, pp. 2461-2470, July 1993.

[5] J.F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc.-F, Radar and Signal Processing*, vol. 140, pp. 362-370, Dec. 1993.

[6] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, No. 3, pp. 287-314, 1994.

[7] D. Yellin and E. Weinstein, "Criteria for multichannel signal separation," *IEEE Trans. Signal Processing*, vol. SP-42, pp. 2158-2168, Aug. 1994.

[8] A. Mansour and C. Jutten, "Fourth-order criteria for blind source separation," *IEEE Trans. Signal Processing*, vol. SP-43, pp. 2022-2025, Aug. 1995.

[9] A. Swami, G.B. Giannakis and S. Shamsunder, "Multichannel ARMA processes," *IEEE Trans. Signal Proc.*, vol. SP-42, pp. 898-914, April 1994.

[10] L. Tong, Y. Inouye and R. Liu, "A finite-step global convergence algorithm for parameter estimation of multichannel MA processes," *IEEE Trans. Signal Processing*, vol. SP-40, pp. 2547-2558, Oct. 1992.

[11] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis, II*, D.F. Findley, Ed. New York: Academic, 1981.

[12] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Info. Theory*, vol. IT-36, pp. 312-321, March 1990.

[13] J.K. Tugnait, "Identification of multichannel linear non-Gaussian processes using higher-order statistics," presented at *29th Annual Asilomar Conf. Signals Systems Computers*, Pacific Grove, CA, Oct. 29 - Nov. 1, 1995.

[14] J.R. Treichler and M.G. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-33, pp. 420-431, April 1985.

[15] M. Rosenblatt, *Stationary Sequences and Random Fields*. Birkhäuser: Boston, 1985.

[16] J.J. Shynk and R.P. Gooch, "Convergence properties of the multistage CMA adaptive beamformer," in *Proc. 27th Asilomar Conf. on Signals, Systems, Computers*, Pacific Grove, CA, pp. 622-626, Nov. 1993.

[17] P. Loubaton and P. Regalia, "Blind deconvolution of multivariate signals: a deflation approach," in *Proc. Intern. Conf. Commun.*, pp. 1160-1164, Geneva, Switzerland, June 1993.

Blind Source Separation of Convolutive Mixtures

C.Servière

CEPHAG-ENSIEG BP 46 38402 Saint-Martin d'Hères cedex FRANCE

Abstract

When a priori information about the propagation or the geometry of the array are not available, the model can be generalized to a blind source separation problem It supposes the statistical independence of the sources and their non-gaussianity. In this paper, the observed signals are supposed to be convolutive mixtures of wide-band sources. Several criteria of source separation are studied, which are based on the cancellation of different fourth-order cross-cumulants. For these criteria, we show in which conditions the separation is achieved. Results on real data illustrate the proposed methods.

1. Introduction

The problem of separating a mixture of several independent signals is encountered in many fields : in digital communication multipath channels, in speech enhancement (cocktail party problem), or in the diagnostic of rotating machines. Several methods have been recently proposed in [1] [2] [3] [4]. The problem, generally called "blind source separation", consists in identifying p independent and non-gaussian sources from M observed linear mixtures of these sources. These techniques are necessary when the propagation between sources and sensors cannot be modelled (unknown paths, unknown antenna deformation, complicated array geometry, or unavailable hypothesis of plane waves...).

Several methods [1] [2] [3] [4] [5] have been developed in time domain in the case of linear instantaneous mixtures, using higher-order statistics (usually fourth-order moments or cumulants, or non linear functions of the observations) or using a deflation approach. In the frequency domain, several methods based on the cross-bispectra or the trispectra of the estimated sources have been proposed in [6] [7] [8]. [9] and [10] propose an adaptive approach in the time-domain in the case of convolutive mixtures. [11] uses a priori information on the probability densities of the sources. In a general blind source separation problem, the observed data vector $\underline{r}(t)$ may be represented in frequency-domain by an instantaneous complex mixture for each frequency bin f. It leads to the following model:

(1) $\quad \underline{R}^k(f) = A(f)\,\underline{S}^k(f) + \underline{B}^k(f)$

where $\underline{R}^k(f)$ is the N-point Discrete Fourier Transform (DFT) of the kth data block of the observation $\underline{r}(t)$. $\underline{S}^k(f)$ represents the p sources vector and A(f) is an unknown matrix (M.p) which characterizes the linear propagation from sources to sensors. $\underline{B}^k(f)$ represents an additive M-dimensional gaussian noise. The problem consists in identifying the matrix A(f) as a product of three matrices :

(2) $\quad A(f)=V(f)\,\Delta(f)\,\Pi(f)$

The matrices V(f) (a unitary matrix) and $\Delta(f)$ (a diagonal matrix) are identified thanks to second-order criteria, by eigenvalue decomposition of the covariance matrix of $\underline{R}^k(f)$. After this first usual step using only second-order moments (developed in §2), we suppose that the components of the observations are normalized and uncorrelated, which is not a restrictive assumption.

We focus then in this paper on the identification of the matrices $\Pi(f)$ (which are unitary matrices) thanks to fourth-order criteria. In the case of instantaneous mixtures, two methods have been already proposed [1] [2]. We focus in this paper on the generalization of the source separation problem to convolutive mixtures of wide-band sources. C.Jutten proposes in [9] to cancel certain fourth-order cross-cumulants. The separation is only proved under the condition of independent, identically distributed (i.i.d.) processes with the same sign of kurtosis.

In this paper, we study several criteria based on the cancellation of other fourth-order cross-cumulants of the estimated sources and we show in which conditions the separation is achieved.

2. Modelization of the problem

The first step consists in the identification of the matrices V(f) and $\Delta(f)$ (which may be adaptively computed). The signals, noted $\underline{X}^k(f)$, issued from the projection of the observations $\underline{R}^k(f)$ in the signal subspace (which is spanned by the columns of V(f) associated to the dominant eigen values of the covariance matrix of the observations) are uncorrelated and normalized. The p normalized sources, called $\underline{NS}^k(f)$, are relied to the p new data, $\underline{X}^k(f)$ by :

(3) $\quad \underline{NS}^k(f) = (\Pi(f)\,P(f)\,D(f))^{+}\underline{X}^k(f)$

0-8186-7576-4/96 $5.00

where P(f) is a (p.p) permutation matrix and D(f) is a diagonal one. The notation '+' means transpose and conjugate.
The unitary matrix $\Pi(f)$ [2] can be decomposed into a product of Givens rotations. In the case of two sources, for example, $\Pi(f)$ is a function of two angles $\theta(f)$ and $\phi(f)$, at frequency bin f. The modelization of the normalized sources $NS_1(f)$ and $NS_2(f)$ after the second-order step is the following:

(4)$NS_1(f)=\cos(\theta(f))X_1(f)+\sin(\theta(f))\exp(-j\phi(f))X_2(f)$

(5)$NS_2(f)=\sin(\theta(f))\exp(j\phi(f))X_1(f) - \cos(\theta(f))X_2(f)$

In the time-domain, the p normalized sources $\underline{ns}(t)$ are expressed as convolutive mixtures of the p new data $\underline{x}(t)$:

$$(6) \qquad ns_j(t) = \sum_{i=1,p} \underline{h}_{ij}(t) * x_i(t) \qquad j=1, \ldots, p$$

where $x_i(t)$ is the i-th component of the vector $\underline{x}(t)$ and $ns_j(t)$ is the j-th component of the vector $\underline{ns}(t)$. The vectors $\underline{h}_{ij}(t)$ represent the finite impulse responses of the filters between the i-th component of the data $\underline{x}(t)$ and the j-th normalized source. They are exactly the N-point inverse Discrete Fourier Transforms of the filters, characterized in frequency-domain by the N matrices $(\Pi(f)\ P(f)\ D(f))^+$. In the case of two sources, the impulse responses of the filters $\underline{h}ij(t)$ are the inverse Discrete Fourier Transform of $[\cos(\theta(f))]$ and $[\sin(\theta(f))\exp(-j\phi(f))]$ for (f=0, ..., N-1).

3 Independence criterion

Thanks to the previous step (at second order), the components of $\underline{x}(t)$ are normalized and uncorrelated. As the information provided by the second-order statistics is not sufficient to identify the N matrix $(\Pi(f)\ P(f)\ D(f))^+$, we use an additive assumption : the statistical independence of the sources. The aim of the second step of blind source separation procedure is the identification of the vectors $\underline{h}_{ij}(t)$ (or the matrices $(\Pi(f)\ P(f)\ D(f))^+$) such that the estimated sources $ns_j(t)$ are independent. In that case, if the sources $ns_i(t)$ and $ns_j(t)$ are statistically independent, we must have the cancellation of each cross-cumulant for any delay k, l, m, n inferior to N,

(7)$C22(ns_i(t-k),ns_j(t-l),ns_j(t-m),ns_j(t-n))$,$C31(ns_i(t-k),ns_j(t-l),ns_j(t-m),ns_j(t-n))$,$C13(ns_i(t-k),ns_j(t-l),ns_j(t-m),ns_j(t-n))$. C represents the fourth-order cumulant as defined in [12]. It leads in the strict sense to a fourth-order independence.
In the case of instantaneous mixtures, P.Comon first proposed to maximize a contrast function based on the kurtosis of the estimated sources [2], such that the maxima are obtained for solutions which actually separate the sources. E.Moreau and O.Macchi proposed in [3] an approach based on the adaptive maximization of others contrast functions, using the kurtosis and the fourth-order cross-cumulants of the estimated sources.

In the case of convolutive mixtures of wide-band sources, the purpose of this paper is to study several criteria and to show in which conditions the separation is achieved. C.Jutten proposes in [9] to cancel the symmetrical fourth-order cross-cumulants $C22(ns_j(t),ns_i(t),ns_j(t-k),ns_j(t-k))$, functions of one delay k. He proves that it is a sufficient condition to separate two sources under the hypothesis of independent, identically distributed (i.i.d.) processes with the same sign of kurtosis. The cancellation of the two others cross-cumulants $C13(ns_j(t),ns_j(t-k),ns_j(t-k),ns_j(t-k))$ and $C31(ns_j(t),ns_j(t-k),ns_i(t-k),ns_i(t-k))$ are not applicable because spurious solutions exist.

3.1 Study of the criteria

In [9], the case of one delay is dealt with. We propose in this section to study the cancellation of two dissymmetrical cross-cumulants, functions of two delays k and l :

(8) $C13(ns_i(t-k), ns_j(t), ns_j(t), ns_j(t-l)) = 0$

$C31(ns_j(t), ns_i(t), ns_i(t-k), ns_j(t-l)) = 0$

As the estimated sources $ns_i(t)$ and $ns_j(t)$ are researched as uncorrelated signals, it leads to minimize the following cost function Ψ :

$$(9)\Psi = \sum_{i\neq j} \sum_{k,l=0}^{N-1} (E13(nsi(t-k)\ nsj(t)\ nsj(t)\ nsj(t-l)))^2$$

Ψ can also be computed as :

$$\Psi = N^2 \sum_{i\neq j} \sum_{f1,f2=0}^{N-1} \left| B_{i,j}(f1,f2) \right|^2$$

where $B_{i,j}$ (f1,f2) represents the two-dimensional Fourier Transform of E(nsi(t-k) nsj(t) nsj(t) nsj(t-l)), relative to the time variables k and l. $B_{i,j}(f1,f2)$ is equal to $E\{nsi(t)^2 NSi(f1) NSj(f2)\}$. The function Ψ is then minimized (and equal to zero) when $B_{i,j}$ (f1,f2) is canceled for each frequency bin f1 and f2. The cancellation of $B_{i,j}$ (f1,-f1) leads to two types of solutions : the first one achieves the separation while the second one consists in spurious solutions. These spurious solutions only exist in the case of sources with identical statistical properties at the fourth-order. However, we can show, computing the value of the function Ψ, that these spurious solutions do not cancel it. Consequently, the cancellation of Ψ (9) always assures the separation of the sources.
In the case of two sources, the estimated normalized sources $NS_1(f)$ and $NS_2(f)$ can be expressed in function of the normalized sources, called S1n(f) and S2n(f) with :

(10)$NS_1(f)$= H1'(f) S1n(f) + H2'(f) S2n(f)

(11)$NS_2(f)$= H1"(f) S1n(f) + H2"(f) S2n(f)

The source separation is achieved at frequency bin f when H1'(f) and H2"(f) (or H2'(f) and H1"(f)) are equal to zero.
The criterion Ψ can also be expressed in function of $\theta(f)$ and $\phi(f)$, or in function of the complex gains Hi'(f) and Hi"(f). The cancellation of $B_{1,2}$ (f1,-f1) and $B_{2,1}$ (-f1,f1)

leads to one or two types of solutions in function of the sources. If the sources not verify the following condition at each frequency bin f1, relative to the fourth-order statistics of the sources: (12)

$$\sum_{f=0}^{N-1} E\{|S1n(f)|^2|S1n(f1)|^2\} = \sum_{f=0}^{N-1} E\{|S2n(f)|^2|S2n(f1)|^2\}$$

the cancellation of ($B_{1,2}$ (f1,-f1) + $B_{2,1}$ (-f1,f1)) leads to (H1'(f1)H1"(f1)*=0) for each frequency bin f1. The separation is then achieved at each frequency bin. If the condition (12) is realized, several spurious solutions exist which verify :

$$(13)|H1'(f)|^2 = |H1''(f)|^2 = \frac{1}{2}$$

However, if we replace (13) in the function ψ, we remark that these spurious solutions do not cancel it. We also can show by studying the minima of $B_{1,2}$ (f1,f2) and $B_{2,1}$ (f1,f2) that the separation is not independently achieved for each frequency bin f. The sources associated to the identified signals are necessarily the same from one frequency bin to another. Consequently, the source separation is assured with the minimization of the criterion ψ, function of two delays k,l, or two frequency bins f1, f2.

3.2 Case of different sources

In the case of two different enough sources (relative to their fourth-order statistics), we deduce from §3.1 a simplificated criterion γ which achieved the separation at each frequency bin. It cancels the N following equations :

(14) $B_{1,2}$ (f1,-f1) + $B_{2,1}$ (-f1,f1) = 0

γ is deduced from (14) by :

$$(15)\gamma = \sum_{f=0}^{N-1} |B_{1,2}(f,-f) + B_{2,1}(-f,f)|^2$$

After some computations, we obtain that :

$$(16)B_{1,2}(f,-f) + B_{2,1}(-f,f) = H1'(f)H1''(f) * F(s(t),f)$$

where F(s(t),f) only depends on the fourth-order moments of the two sources. From the expression (16), we conclude that it only depend on the coefficients θ(f) and ϕ(f) at frequency bin f. In order to estimate them, it is then theoretically equivalent to minimize the function γ (15) or to find the solutions which cancel the equation (14) at frequency bin f. Call γ(f), the contribution of the cost function γ at the frequency bin f. The function γ(f) can be adaptively minimized. After some computations, we obtain three types of solutions which cancel the derivative of γ(f), relative to the variables θ(f) or ϕ(f). The first one leads to (H1'(f)=0) which separate the sources. The second one provides : ($|H1'(f)|^2$ =1/2) and we easily verify that these solutions correspond to maxima of γ(f) which are not stable points. The third type of solutions which verify : $\left(\frac{\partial|H1'(f)|}{\partial\theta(f)} = 0\right)$ or $\left(\frac{\partial|H1'(f)|^2}{\partial\theta(f)} = 0\right)$ contains all the previous points. The same conclusion is obtained with the derivatives relative to the variables ϕ(f). As a result, the proposed cost function has no local minima and it assumes that the proposed criterion may be adaptively minimized. We obtain the following adaptation laws for the estimation of θ(f) and ϕ(f) at time t+1, θ(f,t+1) and ϕ(f,t+1) :

$$(17)\theta(f,t+1) = \theta(f,t) - 2\mu \sum_{k,l=0}^{N-1} [E13(k,l)B + E31(k,l)C]$$

$$B = ns2(t)^2 ns2(t-1)\frac{\partial ns1(t-k)}{\partial\theta} + ns1(t-k)\frac{\partial ns2(t)^2 ns2(t-l)}{\partial\theta}$$

$$C = ns1(t)^2 ns1(t-1)\frac{\partial ns2(t-k)}{\partial\theta} + ns2(t-k)\frac{\partial ns1(t)^2 ns1(t-l)}{\partial\theta}$$

with :

$$\frac{\partial ns1(t)}{\partial\theta} = -\sin(\theta(f,t))X_1(f) + \cos(\theta(f,t))\exp(-j\phi(f,t))X_2(f)$$

$$\frac{\partial ns2(t)}{\partial\theta} = \cos(\theta(f,t))\exp(j\phi(f,t))X_1(f) + \sin(\theta(f,t))X_2(f)$$

We obtain similar adaptation laws for ϕ(f,t+1)

with : $\frac{\partial ns1(t)}{\partial\phi} = -j\exp(-j\phi(f,t))\sin(\theta(f,t))X_2(f)$

$$\frac{\partial ns2(t)}{\partial\phi} = j\exp(j\phi(f,t))\sin(\theta(f,t))X_1(f)$$

The unknown moments in (17) are adaptively estimated with the available data.

3.3 Case of similar sources

In the case of two similar sources (relative to their fourth-order statistics), the dissymmetrical cumulants $B_{1,2}$ (f1,-f1). and $B_{2,1}$ (f1,-f1) are zero for any estimated complex gains H'i(f) and H"i(f). We study here the cancellation of the two-dimensional Fourier Transform of the symmetrical cumulants C22($ns_i(t)$,$ns_i(t)$,$ns_j(t-k)$,$ns_j(t-l)$), funtions of two delays. The cancellation at frequency bins (f1,-f1) is equal to :

$$(17)\ \sum_f \left[|H''1(f1)|^2|H'1(f)|^2 + |H''2(f1)|^2|H'2(f)|^2\right].$$

$$\left[E\left\{|S(f)|^2|S(f1)|^2\right\} - 1\right] - 4|H''1(f1)|^2|H'1(f1)|^2 = 0$$

From (17), we can deduce that the separation is achieved in several conditions. If $\left[E\left\{|S(f)|^2|S(f1)|^2\right\} - 1\right]$=0 (which is the case of sinusoids, rotating machines noises [13]), the separation is obtained at each frequency bin f1. If $\left[E\left\{|S(f)|^2|S(f1)|^2\right\} - 1\right]$<0, it can be shown that the separation is jointly obtained for each frequency bin. Consequently, the estimated temporal sources are actually independent.

4. Results

The simulation here after illustrates the behavior of the proposed method in §3.3, in the case of convolutive mixtures of two sources. The two processes are rotating machine signals. The filters of the mixtures are MA filters of order 100. We compared the method to existing algorithms [9] [10] which in that case converge to local minima. We present the spectral densities of the observations in Fig1-2, of the right sources in Fig3-4, and of the estimated sources in Fig5-6. We remark the good correspondence between the spectra Fig3-4 and Fig5-6, which reveals the convergence of the proposed method towards a good solution.

5 Conclusion

We focus in this paper on the generalization of the blind source separation problem to convolutive mixtures of wide-band sources.Several criteria of source separation are studied, which are based on the cancellation of different fourth-order cross-cumulants. For these criteria, we show in which conditions the separation is achieved. Results on real data illustrate the proposed methods (separation of rotating machine noises).

References

[1] J. F. CARDOSO, A.BELOUCHRANI, B.LAHELD, "A new composite criterion for adaptive and iterative blind source separator", Proc ICASSP, vol 4, pp 273-276, 1994.

[2] P. COMON , "Independent component analysis" in Proc. Int. Workshop on Higher Order Statistics, Chamrousse, France, pp 111-120, 1991.

[3] E.MOREAU, O.MACCHI, "Complex self-adaptive algorithms for source separation based on higher order contrasts", pp1157-1160, in Proc. Signal ProcessingVII, Edinburgh, September 1994

[4]A.DINC,Y.BAR-NESS, "Bootstrap:A fast blind adaptive signal separator", Proc.ICASSP, vol 2, pp. 325-328, 1992

[5] N.DELFOSSE, P.LOUBATON, "Adaptive blind separation of independent sources: A deflation approach", pp59-83, Signal Processing, vol 45, 1995

[6] D.YELLIN, E.WEINSTEIN, "Multi-Channel Signal Separation based on Cross-Bispectra", Proc. Int. Workshop on Higher Order Statistics, Lake Tahoe, pp 270-274, 1993

[7] M. GAETA, J.L. LACOUME, "Source separation versus hypothesis" in Proc. Int. Workshop on Higher Order Statistics, Chamrousse, France, pp 271-274, 1991.

[8] P.COMON, B.EMILE, "Estimation of Time Delays in the Blind Mixture Problem", pp.482-485, in Proc. Signal Processing VII, Edinburgh, September 1994

[9] H.L.NGUYEN, C.JUTTEN, "Blind source separation for convolutive mixtures", pp209-229, Signal Processing, vol 45, 1995

[10] M.NAJAR, M.A.LAGUNAS, I.BONET, "Blind wideband source separation", Proc.ICASSP94, vol4, pp 65-68

[11] R.LAMBERT, "A new method for source separation", Proc.ICASSP95,pp2116-2119

[12] D.R.BRILLINGER, "Time series, data analysis and theory". San Francisco : Holden Day, 1981 p94.

[13] V.CAPDEVIELLE, C.SERVIERE, J-L.LACOUME, "Blind separation of wide-band sources in the frequency-domain", Proc.ICASSP95, pp2080-2083

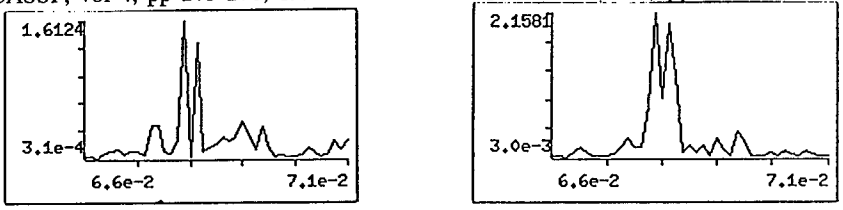


Fig.1-2 Spectral densities of the observations

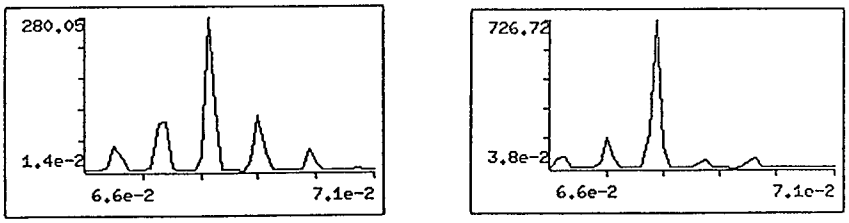


Fig.3-4 Spectral densities of the sources

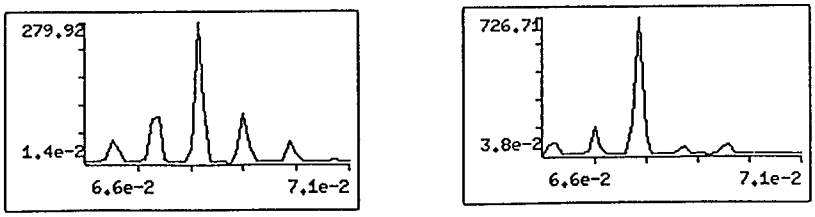


Fig.5-6 Spectral densities of the estimated sources

Unconstrained Optimization Criteria for Blind Equalization of Multichannel Linear Systems

Yujiro Inouye and Takehito Sato

Department of Systems Engineering
Faculty of Engineering Science
Osaka University
Toyonaka,Osaka 560, Japan
inouye@sys.es.osaka-u.ac.jp

Abstract

Blind equalization and blind deconvolution have been an important interesting topic in diverse fields including data communication, image processing and geophysical data processing. Recently, Inouye and Habe proposed a multistage maximization criterion and a single-stage maximization criterion for attaining the blind equalization of multichannel linear time-invariant systems. However, their maximization criteria should be subjected to several constraints of equations. In this paper, we present unconstrained new maximization criteria for accomplishing the blind equalization of multichannel linear time-invariant systems. Stochastic gradient algorithms are proposed for solving the unconstrained maximization problems. Simulation examples are included to examine the performance of the proposed algorithms.

1 Introduction

Blind equalization and blind deconvolution have been an important interesting topic in diverse fields including data communication, image processing and geophysical data processing [1]-[3]. Recently, Shalvi and Weinstein presented several new criteria for blind equalization of single-channel linear time-invariant systems [2]. Inouye and Habe extended the Shalvi-Weinstein approach to the multichannel case [3]. They proposed a multistage maximization criterion and a single-stage maximization criterion for attaining the blind equalization of multichannel linear time-invariant systems. However, their maximization criteria should be subjected to several constraints of equations. In general, unconstrained optimization criteria are generally better than constrained optimization criteria for the purpose of achieving the optimization.

In this paper, we present unconstrained new maximization criteria for accomplishing the blind equalization of multichannel linear time-invariant systems. Stochastic gradient algorithms are proposed for solving the unconstrained maximization problems. Simulation examples are included to examine the performance of the proposed algorithms.

2 Problem Formulation

Let us consider the system shown in Fig. 1. It is a cascade connection of an unknown multichannel system preceding a multichannel equalizer.

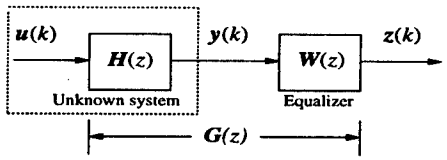


Figure 1. Unknown system and equalizer

We make the following assumptions on the system and the signals involved.

(A1) The unknown system $\boldsymbol{H}(z)$ is described by

$$\boldsymbol{y}(t) = \sum_{k=-\infty}^{\infty} \boldsymbol{H}(k)\boldsymbol{u}(t-k) \qquad (1)$$

where $\boldsymbol{y}(t)$ is a real/complex n-column output vector, $\boldsymbol{u}(t)$ is a real/complex n-column input vector, and $\{\boldsymbol{H}(k)\}$ is a real/complex $n \times n$ matrix sequence called the **impulse response**. The system is stable, that is, the impulse response

0-8186-7576-4/96 $5.00 © 1996 IEEE

satisfies the absolute summability condition

$$\sum_{k=-\infty}^{\infty} ||\boldsymbol{H}(k)|| < \infty \tag{2}$$

(A2) The transfer function defined by

$$\boldsymbol{H}(z) := \sum_{k=-\infty}^{\infty} \boldsymbol{H}(k) z^k \tag{3}$$

is of full rank on the unit circle $|z| = 1$ (this implies it has no zero on the unit circle).

(A3) The input process $\{\boldsymbol{u}(t)\}$ is a zero-mean, non-Gaussian vector process, whose component processes $\{u_i(t)\}$, $i = 1, \cdots, n$, are mutually independent. Moreover, each component process $\{u_i(t)\}$ is an independently and identically distributed (i.i.d.) process with variance $\sigma_{u_i}^2 \neq 0$ and fourth-order cumulant $\kappa_{4,u_i} \neq 0$.

(A4) The equalizer $\boldsymbol{W}(z)$ is described by

$$\boldsymbol{z}(t) = \sum_{k=-\infty}^{\infty} \boldsymbol{W}(k) \boldsymbol{y}(t-k) \tag{4}$$

where $\boldsymbol{z}(t)$ is a real/complex n-column vector, called the **equalizer output**, and $\{\boldsymbol{W}(k)\}$ is a real/complex $n \times n$ matrix sequence. It is assumed that the equalizer $\boldsymbol{W}$ is also stable.

For the blind equalization of the unknown system, we cannot observe the input, but can observe only the output. This implies there are inherent ambiguities in the solution to the multichannel equalization problem as follows: In general, we cannot identify the order of the arrangement of the components $u_1(t), \cdots, u_n(t)$ of input vector $\boldsymbol{u}(t)$, the time origin of each component $u_i(t)$, and the magnitude of each component $u_i(t)$.

Taking these ambiguities into account, the multichannel blind equalization problem is formulated such that it is to find an equalizer $\boldsymbol{W}$ so that the transfer function $\boldsymbol{G}(z)$ of the combined system takes the form of

$$\boldsymbol{G}(z) = \boldsymbol{P}\boldsymbol{\Lambda}(z)\boldsymbol{D} \tag{5}$$

where $\boldsymbol{P}$ is a permutation matrix, $\boldsymbol{\Lambda}(z)$ is a diagonal matrix with diagonal entries $\lambda_{ii}(z) = z^{l_i}$, $i = 1, \cdots, n$ (where l_i is an integer), and $\boldsymbol{D}$ is a constant diagonal matrix. Moreover, if we know all the magnitudes of the variances of the components of the input process ahead, we can constrain to make the diagonal matrix $\boldsymbol{D}$ in (5) be equal to a diagonal matrix with the diagonal entries all being unit magnitude.

It is said that a stationary random process $\{\boldsymbol{u}(t)\}$ satisfies the **normalized whitening condition** if the all the component processes $\{u_i(t)\}, i = 1, \cdots, n$, of $\{\boldsymbol{u}(t)\}$ are white random processes with unit variance and they are mutually uncorrelated. When the random process is zero-mean, this condition is equivalent to $E\{\boldsymbol{u}(t+k)\boldsymbol{u}^*(t)\} = \boldsymbol{I}\delta(k)$, where $\boldsymbol{I}$ denotes the identity matrix and $\delta(k)$ denotes the Kronecker delta.

By the multilinearity property of cumulants, we can derive the following formula for the components of the equalizer output vector $\boldsymbol{z}(t)$ from (1) and (4) with (A1)-(A4). Let $\{\boldsymbol{G}(t)\}$ be the impulse response of the cascade system in Fig. 1. Then for any $i_1, i_2 \in \{1, 2, \cdots, n\}$,

$$r_{z_{i_1}, z_{i_2}^*}(\tau_1) = \sum_{j=1}^{n} \sum_{\tau=-\infty}^{\infty} g_{i_1 j}(\tau + \tau_1) g_{i_2 j}^*(\tau) \sigma_{u_j}^2 \tag{6}$$

For any $i_1, i_2, i_3, i_4 \in \{1, 2, \cdots, n\}$, we have

$$\begin{aligned} &c_{4, z_{i_1}, z_{i_2}^*, z_{i_3}, z_{i_4}^*}(\tau_1, \tau_2, \tau_3) \\ &= \sum_{j=1}^{n} \sum_{\tau=-\infty}^{\infty} g_{i_1 j}(\tau + \tau_1) g_{i_2 j}^*(\tau + \tau_2) \\ &\qquad g_{i_3 j}(\tau + \tau_3) g_{i_4 j}^*(\tau) \kappa_{4, u_j}. \end{aligned} \tag{7}$$

3 Blind Equalization

To begin with, let us assume that the input process $\{\boldsymbol{u}(t)\}$ satisfies the normalized whitening condition by dividing each component $\{u_i(t)\}$ by the square root of variance $\sigma_{u_i}^2$ to eliminate the magnitude ambiguity. Let Z denote the set of all integers.

3.1 Constrained Criteria

In the previous work [3], the following two maximization criteria, the multistage maximization criterion (A) and the single-stage maximization criterion (B), were proposed and analyzed.

The multistage maximization criterion (A):

(Stage 1): Maximize $|\kappa_{4, z_1}|$ subject to $\sigma_{z_1}^2 = 1$.

(Stage k): Maximize $|\kappa_{4, z_k}|$ subject to $\sigma_{z_k}^2 = 1$ and $r_{z_i, z_k^*}(\tau) = 0$ for all $\tau \in Z$ and all $i = 1, 2, \cdots, k-1$. Here k moves successively from 2 to n.

The single-stage maximization criterion (B):

Maximize $\sum_{i=1}^{n} |\kappa_{4, z_i}|$ subject to $r_{z_i, z_i^*}(\tau) = \delta(\tau)$ for all $i = 1, \cdots, n$ and $r_{z_i, z_j^*}(\tau) = 0$ for all $\tau \in Z$ and all distinct $i, j = 1, \cdots, n$.

Theorem 1: Under the normalized whitening condition of the input process $\{\boldsymbol{u}(t)\}$, the multistage maximization criterion (A) and the single-stage maximization criterion(B), both yield a solution to the multichannel blind equalization.

3.2 Unconstrained Criteria

It is generally more difficult to solve a maximization problem with constraints than to solve a constraint-free maximization problem equivalent to the original one. In the sequel, we develop constraint-free criteria for solving the multichannel blind equalization.

Let us assume that we know all the magnitudes of the fourth-order auto-cumulants of the components of the vector process ahead and that they satisfies the following decreasing sequence condition

$$|\gamma_1| \geq |\gamma_2| \geq \cdots \geq |\gamma_n| \tag{8}$$

where $\gamma_i := \kappa_{4,u_i}$ for $i = 1, \cdots, n$. Consider the following potential function [2] defined by

$$\phi_i(z_i) := |\kappa_{4,z_i}| + |\gamma_i| f(\sigma_{z_i}^2) \tag{9}$$

where $f(\cdot)$ is a continuous real-valued function over $[0, \infty)$ such that

$$p(x) := x^2 + f(x) \tag{10}$$

monotonically increasing in $0 \leq x < 1$, monotonically decreasing $x > 1$, and has a unique maximum at $x = 1$. Such a function, for example, is given by $p(x) = 2\alpha x - \alpha x^2, \alpha > 0$.

Corresponding to the multistage maximization criterion (A), we consider the following unconstrained criterion.

The unconstrained multistage maximization criterion (C):

(Stage 1): Maximize

$$J_1 := |\kappa_{4,z_1}| + |\gamma_1| f(\sigma_{z_1}^2) \tag{11}$$

(Stage k): Maximize

$$J_k := |\kappa_{4,z_k}| + |\gamma_k| f(\sigma_{z_k}^2) - \lambda_0 \Big(\sum_{i=1}^{k-1} \sum_{\tau \in Z} |r_{z_i, z_k^*}(\tau)|^2\Big)^2 \tag{12}$$

where λ_0 is a positive constant greater than $|\gamma_1|, i.e., \lambda_0 \geq |\gamma_1|$.

Based on Theorem 1, we have the following theorem.

Theorem 2: Under the normalized whitening condition of the input process $\{\boldsymbol{u}(t)\}$, the unconstrained multistage maximization criterion (C) gives a solution to the multichannel equalization problem.

Corresponding to the single-stage maximization criterion (B), we need another assumption for the time being that all the magnitudes of the fourth-order cumulants are identical, *i.e.*,

$$|\gamma_1| = |\gamma_2| = \cdots = |\gamma_n| \tag{13}$$

Under this condition, we consider the following unconstrained criterion.

The unconstrained single-stage maximization criterion (D):

Maximize

$$J := \sum_{k=1}^{n} \{|\kappa_{4,z_k}| + |\gamma_k| f(\sigma_{z_k}^2)\} - \lambda_0 \Big(\sum_{k=2}^{n} \sum_{i=1}^{k-1} \sum_{\tau \in Z} |r_{z_i, z_k^*}(\tau)|^2\Big)^2, \tag{14}$$

where λ_0 is a positive constant.

Based on Theorem 1, we can obtain the following theorem.

Theorem 3: Under the normalized whitening condition of the input process $\{\boldsymbol{u}(t)\}$ and the condition (13), the unconstrained single-stage maximization criterion (D) gives a solution to the multichannel blind equalization problem.

Remark 1: When all the magnitudes of the fourth-order auto-cumulants of the components of the input vector process are not the same, the criterion function (14) with λ_0 being a small positive constant can not be generally applied for achieving the multichannel blind equalization. In such a general case, it is not clear at the present how to choose a large number for λ_0 in the criterion function (14) to solve the multichannel blind equalization problem.

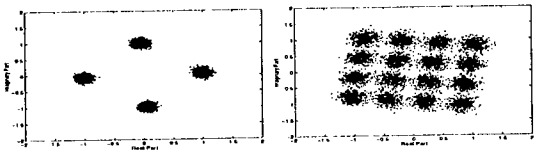
4 Simulation Examples

In order to see the effectiveness of the proposed criteria, we developed two stochastic gradient algorithms for solving the problem of the multistage maximization criterion (A) and the unconstrained multistage maximization criterion (C). They are omitted for page limit. The algorithm for criterion (A) requires (multichannel) spectral prewhiting of the output process of the unknown system. We used a finite impulse response (FIR) system to approximate the equalizer.

We took following system that is a 2-input and 2-output all-pass system described by

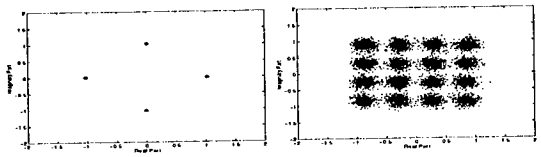
$$\boldsymbol{H}(z) = \begin{pmatrix} \frac{0.5+z^{-1}}{1+0.5z^{-1}} & 0 \\ 0 & \frac{0.2+z^{-1}}{1+0.2z^{-1}} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}. \tag{15}$$

We note that $\boldsymbol{H}(z)$ satisfies the all-pass condition $\boldsymbol{H}(e^{j\omega})\boldsymbol{H}^*(e^{j\omega}) = \boldsymbol{I}$. Hence we need not perform prewhiting in this case. The first channel input signal $u_1(t)$ was 16-QAM with unit variance, and the second channel input signal $u_2(t)$ was 4-PSK (phase-shift keying) with unit variance. We used a 2-input, 2-output and 24-tap equalizer $\boldsymbol{W}(z)$. The both algorithms contain stochastic expectation. Therefore, we used 50 data points to calculate expectation. The step size was chosen to be 0.02. The positive constant λ_0 in (12) was set to be 1. As a measure of performance we used the **multichannel intersymbol interference** denoted by M_{ISI}, defined in [3]. The initial M_{ISI} in the logarithmic (dB) scale was 8.0411 dB.



(a) **Equalized output of channel 1 using criterion (A)**

(b) **Equalized output of channel 2 using criterion (A)**



(c) **Equalized output of channel 1 using criterion (C)**

(d) **Equalized output of channel 2 using criterion (C)**

Figure 2. Signal constellations after equalization.

The both algorithms were tested in 10 Monte Carlo runs using 20,000 data samples at each of the two channel outputs. Fig. 2 shows the equalized signal constellations obtained by using the constrained criterion (A) and the unconstrained criterion (C) with $\alpha = 10$, respectively. Since the magnitude of fourth-order cumulant of the 4-PSK signal is greater than that of the 16-QAM signal, the 4-PSK signal was recovered as the first channel output $z_1(t)$ at Stage 1 and 16-QAM as second channel output $z_2(t)$ at Stage 2. We see from Fig. 2 that the equalized output of channel 1 using the unconstrained criterion (C) converges better than that using the constrained criterion (A), though there is no clear difference between the two equalized outputs of channel 2 using the constrained criterion (A) and using the unconstrained criterion (C).

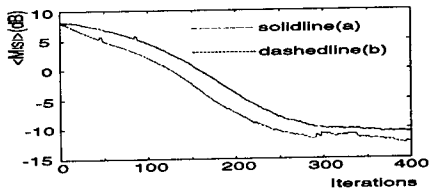


Figure 3. Performances of the algorithms of the constrained criterion (A) and the unconstrained criterion (C). The solid line (a) denotes <M_{ISI}> using the constrained criterion (A), and the dashed line (b) denotes <M_{ISI}> using the unconstrained criterion (C) with α=10.

In Fig. 3, we plotted the averaged M_{ISI}, denoted by $< M_{ISI} >$ over 10 Monte Carlo runs. By comparing the constrained criterion (A) with the unconstrained criterion (C) we found through simulations that the unconstrained criterion (C) exhibits better convergence behavior than the constrained criterion (A) except for the case of $\alpha = 1$. Therefore, we had better choose the value of α greater than 1.

5 Conclusions

We have proposed the unconstrained multistage maximization criterion and the unconstrained single-stage maximization criterion. Simulation examples have shown to illustrate the performance of the algorithm of the constrained multistage criterion (A) and the performance of the algorithm of the unconstrained multistage criterion (C). We have not yet developed two stochastic gradient algorithms for the problems of the single-stage maximization criterion (B) and the unconstrained single-stage maximization criterion (D).

References

[1] D. L. Donoho, "On minimum entropy deconvolution", in D. F. Findley, Ed., *Applied Times Series Analysis, II*, pp. 565-608, New York; Academic Press, 1981.

[2] O. Shalvi, and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems(channels)", *IEEE Trans. Inform. Theory*, vol. 36, pp. 312-321, Mar. 1990.

[3] Y. Inouye and T. Habe, "Multichannel blind equalization using second- and fourth-order cumulants", *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics*, pp. 96-100, 1995.

Source separation based on second order statistics — an algebraic approach

Ulf Lindgren
Department of Applied Electronics
Chalmers University of Technology
S-412 96 Göteborg, Sweden
lindgren@ae.chalmers.se

Alle-Jan van der Veen
Department of Electrical Engineering
Delft University of Technology
2628 CD Delft, The Netherlands
allejan@cas.et.tudelft.nl

Two unknown non-white stochastic sources (e.g. speech signals) are dynamically mixed by an unknown multipath channel and subsequently measured by two sensors. The objective is to construct an inverse filter that separates the two signals, based only on their independence. It is known that, under certain conditions, second-order statistics provide sufficient information to identify the filter. In contrast to the usual cost function optimization techniques, we propose an algorithm that computes the filter coefficients algebraically, *using linear algebra techniques such as the singular value decomposition.*

Keywords: *stochastic signal separation*

1. Introduction

We consider the problem of separating two mutually uncorrelated non-white stochastic sources jointly received over two unknown multipath channels. A number of papers have been published in this context, under various assumptions on the signals or the channels, and using various techniques; see [2–5, 7–9, 12, 13] and the references therein. Techniques may broadly be classified as (*a*) block-methods based on high-order statistics (second and fourth-order cumulants), (*b*) adaptive methods based on optimization of a blind cost function (or nonlinear contrast function), (*c*) maximum-likelihood estimation, presuming the source distributions are known. In many cases, a limited scenario with only scalar mixtures is considered.

The algorithm proposed in this paper is a block-method based on second-order statistics of the measurement data only. The parameters of the inverse filter are to be found such that the resulting filtered output signals $y_1(t)$ and $y_2(t)$ have zero cross-covariance function. Assuming a certain filter structure, the resulting conditions take the form of bilinear equations.

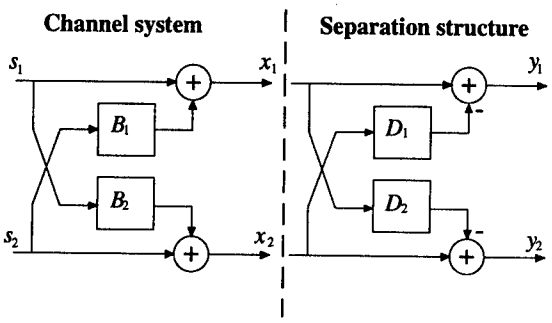


Figure 1. Separation scenario

The usual approach at this point is to set up a cost function whose minimum coincides with the solution of the equations, and to apply a stochastic gradient or Newton-type search algorithm to find the minimum [5]. Our main point is the observation that the equations can also be solved *algebraically*, via a singular value decomposition (SVD). This gives an exact solution to the problem in case the covariance data is exact. With estimated covariances, a subsequent step is needed, in which we have to find a linear combination of a collection of matrices such that the result has rank 1. A similar problem arose in the context of separation of constant-modulus signals [11].

2. Problem formulation

The data model that we consider in this paper is depicted in figure 1. The source signals are $s_1(n)$ and $s_2(n)$, which are linearly filtered white noise processes $\xi_1(n)$, $\xi_2(n)$. We make the following assumptions:

C1: $\xi_1(n)$ and $\xi_2(n)$ are realizations of mutually uncorrelated identically distributed sequences with non-zero variance and zero mean.

C2: $s_1(n)$ and $s_2(n)$ are generated by convolving $\xi_1(n)$ and $\xi_2(n)$ with two different asymptotically stable rational filters.

The source signals are measured via an unknown multichannel, with outputs $x_1(n)$ and $x_2(n)$. The structure of the channel is supposed to consist of a single direct path for the transfer of s_1 to x_1 and s_2 to x_2, and short FIR multipaths $B_1(q^{-1})$ and $B_2(q^{-1})$ for the crosscoupling s_1 to x_2 and s_2 to x_1. The objective is to retrieve $s_1(n)$, $s_2(n)$ from $x_1(n)$, $x_2(n)$. This can be done in a two step procedure where step one is a separation and step two a post-filtering: (1) from x_1, x_2, find $y_1(n) = Hs_1(n)$ and $y_2(n) = Hs_2(n)$, where $H(q^{-1})$ is some FIR filter; (2) inverse filter the sequences $Hs_1(n)$ and $Hs_2(n)$ with $H^{-1}(q^{-1})$ to retrieve s_1 and s_2. Here, we focus on the first step: the actual signal separation.

The separation structure is a direct feedforward filter as depicted in figure 1, where $D_1(q^{-1})$ and $D_2(q^{-1})$ are adaptive FIR filters. When $D_1 = B_1$, $D_2 = B_2$, the separation structure is equal to the channel inverse times the filter $H(q^{-1}) = 1 - D_1(q^{-1})D_2(q^{-1})$, in which case the filter outputs y_1, y_2 are equal to Hs_1, Hs_2. More generally,

$$\begin{bmatrix} y_1^{(\theta)} \\ y_2^{(\theta)} \end{bmatrix} = \begin{bmatrix} 1 & -D_1 \\ -D_2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1-B_2D_1 & B_1-D_1 \\ B_2-D_2 & 1-B_1D_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \quad (1)$$

where $\theta = [d_{10} \ \cdots \ d_{1U-1} \ d_{20} \ \cdots \ d_{2V-1}]^T = [\mathbf{d}_1^T \ \mathbf{d}_2^T]^T$ is the parameter vector of the separation structure. To enable separation, the filter lengths U, V of D_1 and D_2 should be at least as large as the channel lengths, L_1 and L_2. This is only possible if the natural assumption

C3: $L_1 \leq U$ and $L_2 \leq V$

is introduced. Condition C3 assures that the separation structure is in the model class.

In order to recover the sources we require that

C4: $H(q^{-1})$ is minimum phase.

This is natural since $H(q^{-1})$ has a stable inverse only if it is minimum phase. The condition C4 is fulfilled if $|B_1(e^{j\omega})B_2(e^{j\omega})| < 1$ for all $\omega \in [0, 2\pi]$, cf. [1].

3. An algebraic separation algorithm

The proposed algorithm is based on finding the coefficients θ of the separation filter such that the filter outputs y_1 and y_2 are mutually uncorrelated. Let $R^{\theta}_{y_1y_2}(l) = E(y_1(n)y_2(n-l))$ be the cross-correlation between the filtered signals. We will only force independence with respect to second order statistics, i.e. the cross-correlation of y_1 and y_2 is equal to zero for a selected number of $(2L+1)$ lags [7]:

$$R^{\theta}_{y_1y_2}(l) = 0\,, \qquad -L \leq l \leq L\,. \quad (2)$$

The cross-correlation $R^{(\theta)}_{y_1y_2}(l)$ is, under assumption C1 and C2, given in terms of the measured data x_1, x_2 as

$$R^{\theta}_{y_1y_2}(l) = R_{x_1x_2}(l) - \mathbf{d}_1^T\mathbf{r}_{x_2x_2}(l) - \mathbf{d}_2^T\mathbf{r}_{x_1x_1}(l) + \mathbf{d}_1^T\mathbf{R}_{x_2x_1}(l)\mathbf{d}_2, \quad (3)$$

where

$$\mathbf{r}_{x_1x_1}(l) = [R_{x_1x_1}(l) \ \ldots \ R_{x_1x_1}(l+V-1)]^T \quad (4)$$

$$\mathbf{r}_{x_2x_2}(l) = [R_{x_2x_2}(l) \ \ldots \ R_{x_2x_2}(l-U+1)]^T \quad (5)$$

$$\mathbf{R}_{x_2x_1}(l) = [\mathbf{r}_{x_2x_1}(l) \ \ldots \ \mathbf{r}_{x_2x_1}(l+V-1)] \quad (6)$$

$$\mathbf{r}_{x_2x_1}(l) = [R_{x_2x_1}(l) \ \ldots \ R_{x_2x_1}(l-U+1)]^T \quad (7)$$

Thus, the separation problem reduces to solving a system of bilinear equations. In [5] it is proven that there are at least as many equations as unknowns under conditions C1-3 with the exception of the static channel and white sources. By adding the condition C4, this identification problem becomes parameter identifiable (apart from static channels), cf. [6].

The equations (3) with left hand side equal to zero can be solved iteratively, in conjunction with a criterion, by means of gradient minimization techniques, cf. [5]. However, since such techniques are usually bothered by local minima and require accurate initial points, it is interesting also to consider an exact solution of the equations, as follows.

The idea is to rewrite the bilinear equations (3) in matrix form, using Kronecker products to collect all unknowns into a single (structured) parameter vector. This produces

$$\begin{bmatrix} R_{y_1y_2}(-L) \\ \vdots \\ R_{y_1y_2}(L) \end{bmatrix} = \mathbf{P} \begin{bmatrix} \mathbf{d}_2 \otimes \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_1 \\ 1 \end{bmatrix} = \mathbf{0}\,, \quad (8)$$

where '$\otimes$' is the Kronecker product, and

$$\mathbf{P} = \begin{bmatrix} \mathrm{vec}(\mathbf{R}_{x_2x_1}(-L))^T & \mathbf{r}^T_{x_1x_1}(-L) & \mathbf{r}^T_{x_2x_2}(-L) & R_{x_1x_2}(-L) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{vec}(\mathbf{R}_{x_2x_1}(L))^T & \mathbf{r}^T_{x_1x_1}(L) & \mathbf{r}^T_{x_2x_2}(L) & R_{x_1x_2}(L) \end{bmatrix}.$$

'vec' denotes the vectoring operation which stacks all columns of a matrix into a single column. Thus, the problem is equivalent to finding a vector with a certain structure in the null space of the data matrix $\mathbf{P}$. This null space can be determined, or estimated, by a singular value decomposition of $\mathbf{P}$. Thus let a basis for the null space be given by $\{\mathbf{v}_1, \cdots, \mathbf{v}_\delta\}$, where δ is the dimension of the null space. Since the precise basis is arbitrary, the problem is to find a linear combination

of these vectors such that we obtain a vector with the required structure, i.e. to find $\lambda_1, \cdots, \lambda_\delta$ such that

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_\delta \mathbf{v}_\delta = \begin{bmatrix} \mathbf{d}_2 \otimes \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_1 \\ 1 \end{bmatrix}. \quad (9)$$

To make this equivalent problem more tractable, we move from vectors to matrices. For a vector $\mathbf{x}$ partitioned as

$$\begin{aligned} \mathbf{x} &= [\mathbf{x}_1^T \quad \mathbf{x}_2^T \quad \mathbf{x}_3^T \quad x_4]^T \quad (10) \\ &= [x_{11} \cdots x_{1,UV} \; x_{21} \cdots x_{2V} \; x_{31} \cdots x_{3U} \; x_4]^T \end{aligned}$$

define the operator

$$\mathrm{mat}(\mathbf{x}) := \begin{bmatrix} \mathrm{vec}^{-1}(\mathbf{x}_1) & \mathbf{x}_3 \\ \mathbf{x}_2^T & x_4 \end{bmatrix} \quad (11)$$

where $\mathrm{vec}(\mathbf{M})$ is a vectorization of the matrix $\mathbf{M}$ and $\mathrm{vec}(\mathrm{vec}^{-1}(\mathbf{x}_1)) = \mathbf{x}_1$. Note that

$$\mathrm{mat}\left(\begin{bmatrix} \mathbf{d}_2 \otimes \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} \mathbf{d}_1\mathbf{d}_2^T & \mathbf{d}_1 \\ \mathbf{d}_2^T & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{d}_1 \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{d}_2^T & 1 \end{bmatrix}.$$

Denote $\mathbf{V}_1 = \mathrm{mat}(\mathbf{v}_1), \cdots, \mathbf{V}_\delta = \mathrm{mat}(\mathbf{v}_\delta)$. Equation (9) is equivalent to finding $\lambda_1, \cdots, \lambda_\delta$ such that

$$\mathbf{V}_\lambda := \lambda_1 \mathbf{V}_1 + \cdots + \lambda_\delta \mathbf{V}_\delta = \begin{bmatrix} \mathbf{d}_1 \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{d}_2^T & 1 \end{bmatrix}. \quad (12)$$

Basically, we have to select λ_k's such that the resulting linear combination of matrices $\mathbf{V}_\lambda$ is rank 1, in which case it can always be scaled and factored into the required structure.

What is the value of δ? At first sight, given enough conditions (lags) we would expect $\delta = 1$, since the solution to the separation problem is usually unique. However, the Toeplitz structure of $\mathbf{R}_{x2x1}$ adds extra vectors to the null space of $\mathbf{P}$: certain columns of $\mathbf{P}$ are duplicated, which reduces its rank. The number of repeated entries in the Toeplitz matrix is $UV - (U+V-1) = (U-1)(V-1)$, so that we expect $\delta = 1 + (U-1)(V-1)$. The resulting null space basis also has structure: e.g. for $U = 3, V = 3$, a possible matrix basis is of the form

$$\{\mathbf{V}_1, \cdots, \mathbf{V}_5\} = \left\{ \left[\begin{array}{ccc|c} 1 & & & \\ & -1 & & \\ & & 0 & \\ \hline & & & 0 \end{array}\right], \left[\begin{array}{ccc|c} 0 & & & \\ & 1 & & \\ & & -1 & \\ \hline & & & 0 \end{array}\right], \right.$$
$$\left[\begin{array}{ccc|c} 0 & 1 & & \\ & 0 & -1 & \\ & & 0 & \\ \hline & & & 0 \end{array}\right], \left[\begin{array}{ccc|c} 0 & & & \\ 1 & 0 & & \\ & -1 & 0 & \\ \hline & & & 0 \end{array}\right], \left.\left[\begin{array}{ccc|c} * & * & * & \\ * & 0 & 0 & \mathbf{d}_1 \\ * & 0 & 0 & \\ \hline & \mathbf{d}_2^T & & 1 \end{array}\right]\right\}. \quad (13)$$

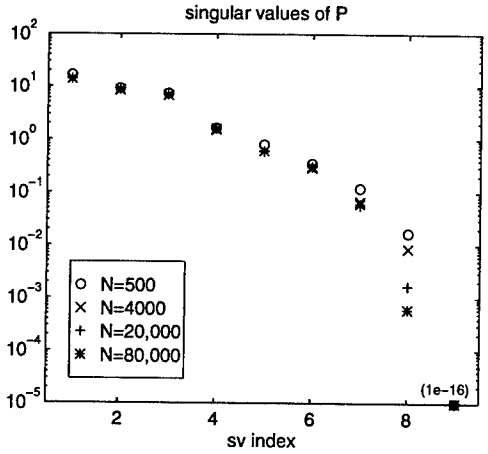


Figure 2. Singular values of P

If we simply remove the duplicate columns of $\mathbf{P}$, then the 'trivial' null space solutions $(\mathbf{V}_1, \cdots, \mathbf{V}_4)$ are suppressed. Only one vector in the null space is left, corresponding to $\mathbf{V}_5$ in (13). Hence, estimates of $\mathbf{d}_1$, $\mathbf{d}_2$ are immediately available, even without solving the rank-1 problem.

The above is true only for perfect knowledge of the covariance lags, i.e. for an infinite amount of data. In actuality, the estimates of these lags converge only slowly to their true values, and the null space is not well-determined. For accuracy reasons it is usually necessary to overestimate the value of δ, and actually search for $\lambda_1, \cdots, \lambda_\delta$ that produces $\mathbf{V}_\lambda$ in (12) that is as close to rank 1 as possible. This is reminiscent of the problem considered (and solved) in [11], where it is shown how a simultaneous diagonalization of (square) matrices $\mathbf{V}_1, \cdots, \mathbf{V}_\delta$ provides good estimates of the λ_k's. The simulation results reported in section 4 are based on a blunt application of this diagonalization algorithm, followed by a few steps of an optimization routine to improve the λ_k's. Although the results are reasonable, it should be remarked that the diagonalization algorithm is theoretically not well motivated for this application, because unlike the case in [11], we now expect only one solution $[\lambda_1 \cdots \lambda_\delta]$, rather than δ independent solutions. This means that the $\mathbf{V}_k$ need not be simultaneously diagonalizable.

4. Simulations

We investigate the performance of the algorithm by simulation. In accordance with conditions C1 and C2, the source signals $s_1(n)$ and $s_2(n)$ are generated by filtering two mutually uncorrelated white Gaussian noise sequences through two autoregressive filters. One filter has a complex pole pair at radius 0.8 and angle $\pm\pi/4$; the other filter has a radius of 0.8 and angle $\pm 3\pi/4$. The channel in this simula-

Method	N	*Mean*				*Variance* $\times 10^{-3}$			
		d_{10}	d_{11}	d_{20}	d_{21}	d_{10}	d_{11}	d_{20}	d_{21}
True/CRB	500	0.5	−0.1	0.7	0.3	0.134	0.130	0.104	0.104
	2000					0.034	0.033	0.026	0.026
	4000					0.017	0.016	0.013	0.013
Algebraic	500	0.499	−0.098	0.701	0.299	0.691	0.719	2.15	1.88
	2000	0.502	−0.098	0.697	0.297	0.162	0.157	0.480	0.481
	4000	0.500	−0.100	0.700	0.299	0.086	0.086	0.232	0.229
Recursive	500	0.500	−0.100	0.697	0.302	0.718	0.582	1.21	2.57
	2000	0.500	−0.100	0.700	0.300	0.043	0.048	0.032	0.035
	4000	0.500	−0.100	0.700	0.300	0.017	0.016	0.016	0.016
Weinstein	500	0.748	0.074	0.748	0.419	1677	873	1677	254
	2000	0.651	0.024	0.651	0.480	62.3	276	623	47.5
	4000	0.668	0.055	0.668	0.493	1334	104	1334	101

Table 1. Mean value and variance of the estimated filter coefficients

tion consists of two filters $B_1(q^{-1}) = 0.5 - 0.1q^{-1}$ and $B_2(q^{-1}) = 0.7 + 0.3q^{-1}$. The correlation matrices (4)–(7) are estimated from $N = 500$, 2000, and 4000 samples of x_1, x_2. We took $L = 4$ lags into account, which gives a total of 9 equations for 4 unknowns. The Cramér-Rao Bound (CRB) for this scenario is derived as $N\text{Var}_\theta = [0.067, 0.065, 0.052, 0.052]$, cf. [10].

A total of 200 independent runs were performed for each sample size. The estimated mean value and parameter variance for the present algebraic algorithm are given in table 1, along with two other algorithms. The "recursive" algorithm is basically a stochastic Newton search algorithm based on [5], and the "Weinstein" algorithm is the one found in [13].

For the algebraic algorithm, theoretically $\delta = 2$, but we have used $\delta = 3$ because even for $N = 4000$ there is no clear gap between the large and small singular values, as is illustrated in figure 2. Even so, the algebraic algorithm performed less good than the recursive method and did not reach the CRB. Given exact (rather than estimated) covariance data, $\mathbf{P}$ does have precisely two zero singular values, and the algorithm did produce the exact solution.

It is known that the "Weinstein" algorithm cannot separate the sources unless $b_{10} = b_{20}$, and this shows up in the results. For a scenario where $b_{10} = b_{20}$ the algorithm works, but yields estimates with a higher variance than the other two algorithms. Note that the variance for "Weinstein" is larger for $N = 4000$ than for $N = 2000$. This is due to large deviations for some, typically two, parameter trajectories.

References

[1] H. Broman, U. Lindgren, H. Sahlin, and P. Stoica. Source separation: A TITO system identification approach. *Submitted to IEEE Trans. Signal Proc.*, 1995.

[2] J. Cardoso, A. Belouchrani, and B. Laheld. A new composite criterion for adaptive and iterative blind source separation. In *Proc. IEEE ICASSP*, pages IV:273–276, 1994.

[3] C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, June 1991.

[4] R. Lambert. A new method for source separation. In *Proc. IEEE ICASSP*, pages 2116–2119, 1995.

[5] U. Lindgren. Signal separation. Technical report, No. 201L (licentiate dissertation), Chalmers University of Technology, Göteborg, Sweden, 1995.

[6] U. Lindgren and H. Broman. Source separation: Using a criterion based on second order statistics. Technical report, Technical Report CTH-TE-36, Chalmers University of Technology, Göteborg, Sweden, 1995.

[7] U. Lindgren, T. Wigren, and H. Broman. On local convergence of a class of blind separation algorithms. *IEEE Trans. Signal Proc.*, 43:3054–3058, Dec. 1995.

[8] P. Loubaton and P. Regalia. Blind deconvolution of multivariate signals using adaptive FIR lossless filters. In J. V. e.a., editor, *Signal Processing VI: Proc. EUSIPCO-92*, pages 1061–1064 vol.2. Elsevier, 1992.

[9] M. Nájar, M. Lagunas, and I. Bonet. Blind wideband source separation. In *Proc. IEEE ICASSP*, pages IV:65–68, 1994.

[10] H. Sahlin and U. Lindgren. "The Asymptotic Cramér-Rao Lower Bound for Blind Signal Separation". To appear in "The Proceedings of the 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing", Corfu, Greece, 1996.

[11] A. J. van der Veen and A. Paulraj. An analytical constant modulus algorithm. *IEEE Trans. Signal Proc.*, 44(5), May 1996.

[12] S. Van Gerven and D. Van Compernolle. On the use of decorrelation in scalar signal separation. In *Proc. IEEE ICASSP*, pages III:57–60, 1994.

[13] E. Weinstein, M. Feder, and A. V. Oppenheim. Multichannel signal separation by decorrelation. *IEEE Trans. Speech Audio Processing*, 1:405–413, Oct. 1993.

The Asymptotic Cramér-Rao Lower Bound for Blind Signal Separation *

Henrik Sahlin and Ulf Lindgren
Chalmers University of Technology,
Department of Applied Electronics,
S-412 96 Gothenburg, Sweden,
salle@ae.chalmers.se,
Telephone +46 31 772 1845

Abstract

This paper considers some aspects of the source separation problem. Unmeasurable source signals are assumed to be mixed by means of a channel system resulting in measurable output signals. These output signals can be used to determine a separation structure in order to extract the sources. When solving the source separation problem the channel filter parameters have to be estimated. This paper presents a compact and computationally appealing formula for computing a lower bound for the variance of these parameters, in a general Many Inputs Many Outputs scenario. This lower bound is the asymptotic (assuming the number of data samples to be large) Cramér-Rao lower bound. The CRLB formula is developed further for the two-input two-output system and compared with the results from a Recursive Prediction Error Method.

1 Introduction

The problem of separating two signals that are mixed through an unknown dynamic channel is considered. Both the source generating filters and the mixing channels are modeled as ARMA-filters. This model is realistic in applications such as hands-free and handheld mobile telephony in the presence of acoustic interference. Noise reduction in hearing-aids is another application.

A lower bound for the covariance matrix of unbiased parameter estimates is given by the Cramér-Rao Lower Bound (CRLB) [3, 8]. The Prediction Error Method (PEM) is, for Gaussian distributed disturbances, asymptotically efficient, cf [5]. This means that the covariance matrix of the estimated parameters is asymptotically equal to the CRLB.

Source separation is an intensive area of research and in the past years many algorithms have been presented [2, 6, 7, 4]. It is of great interest to compute the CRLB for the source separation problem, since it provides a bench-mark to compare algorithms.

2 Problem formulation

Figure 1 depicts the scenario under consideration. Unmeasurable source signals, x_1 and x_2, are mixed by means of a channel system resulting in measurable output signals y_1 and y_2. These output signals can be used to determine a separation structure in order to extract the sources. A two-input two-output (TITO) system is used to specify the source separation problem.

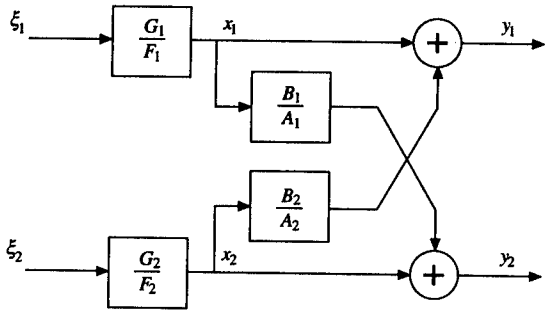


Figure 1. The data generating system.

It is assumed that the source signals, x_1 and x_2, can be modeled as finite order ARMA processes, and that

*This work was financially supported by the Swedish Research Council for Engineering Sciences (TFR) and the Swedish National Board for Industrial and Technical Development (NUTEK).

0-8186-7576-4/96 $5.00 © 1996 IEEE

the channel filters have rational transfer functions. The resulting equations can be put in a matrix form as

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} \frac{G_1}{F_1} & \frac{G_2 B_2}{F_2 A_2} \\ \frac{G_1 B_1}{F_1 A_1} & \frac{G_2}{F_2} \end{bmatrix} \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \end{bmatrix} \quad (1)$$

where A_1, A_2 etc. are polynomials in the unit delay operator q^{-1}, and the signals y_1, y_2 etc. are functions of the discrete time variable $t = 1, 2, ...$ (To simplify notation the dependence on q^{-1} and t is omitted whenever appropriate). Without any restriction F_1, F_2, A_1, and A_2 are assumed to be monic and minimum phase, and G_1 and G_2 to be minimum phase with nonzero direct terms. Also, the two driving sequences ξ_1 and ξ_2 are assumed to be zero mean, unit variance, mutually uncorrelated white Gaussian noises.

Denote $B_1(q^{-1}) = b_{10} + b_{11}q^{-1} + ... + b_{1n_1}q^{-n_1}$, $B_2(q^{-1}) = b_{20} + b_{21}q^{-1} + ... + b_{2n_2}q^{-n_2}$, etc. Let, for $k = 1, 2$, $\sigma_k = G_k(0)$, and introduce the following scaled versions of $\{G_k\}$ and $\{\xi_k\}$: $G_k \leftarrow G_k/\sigma_k, \tilde{\xi}_k = \sigma_k \xi_k$ (For convenience we use the same notation for the scaled and un-scaled $\{G_k\}$ polynomials).

The system in (1) can be rewritten as (assuming $b_{10}b_{20} \neq 1$) :

$$\mathbf{y(t)} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} 1 & \frac{B_2}{A_2} \\ \frac{B_1}{A_1} & 1 \end{bmatrix} \begin{bmatrix} \frac{G_1}{F_1} & 0 \\ 0 & \frac{G_2}{F_2} \end{bmatrix} \begin{bmatrix} 1 & b_{20} \\ b_{10} & 1 \end{bmatrix}^{-1}}_{\mathbf{H}}$$

$$\underbrace{\begin{bmatrix} \tilde{\xi}_1 + b_{20}\tilde{\xi}_2 \\ b_{10}\tilde{\xi}_1 + \tilde{\xi}_2 \end{bmatrix}}_{\mathbf{e}} = \mathbf{H}(q^{-1})\mathbf{e}. \quad (2)$$

The purpose of the latter manipulation is that $\mathbf{H}(0) = \mathbf{I}$, which simplifies the following derivations.

3 A compact expression of the CRLB for MIMO-systems

In this section a compact expression for the CRLB is derived. It should be noted that the derivation and formulae are independent of the number of source signals i.e. the results can be used for a general Many Inputs Many Outputs (MIMO) system.

Lemma 1
Consider a transfer matrix $\mathbf{H}(q^{-1})$ *with* $\mathbf{y} = \mathbf{He}$ and $\Lambda = E\{\mathbf{e}\mathbf{e}^T\}$. *Assume that* $\mathbf{H}$ and $\mathbf{H}^{-1}$ are stable, $\mathbf{H}(0) = \mathbf{I}$, and $\mathbf{e}(t)$ is white noise. Parameterize *the model with the vector* θ *which contains the unknown coefficients of* $\mathbf{H}$ and Λ.

Denote

$$\mathbf{H}_k = \frac{\partial \mathbf{H}}{\partial \theta_k}, \Lambda_k = \frac{\partial \Lambda}{\partial \theta_k}$$

and let the number of samples be N.
Then, for $N \gg 1$, *the CRLB is given by*

$$CRLB = \mathbf{J}^{-1}, \quad (3a)$$

where $\mathbf{J}$ *is the Fisher information matrix:*

$$[\mathbf{J}]_{k,l} = N \operatorname{Tr}\{0.5\Lambda_k\Lambda^{-1}\Lambda_l\Lambda^{-1} + \text{Real}\left\{\frac{1}{2\pi j}\oint \mathbf{H}_k\Lambda\mathbf{H}_l^*\mathbf{H}^{-*}\Lambda^{-1}\mathbf{H}^{-1}\frac{dz}{z}\right\}\} \quad (3b)$$

and where $*$ *denotes conjugate transpose.*

Proof Consider Whittle's formula [8] for the asymptotic information matrix

$$[\mathbf{J}]_{k,l} = \operatorname{Tr}\frac{N}{4\pi j}\oint \phi^{-1}(z)\frac{\partial\phi(z)}{\partial\theta_k}\phi^{-1}(z)\frac{\partial\phi(z)}{\partial\theta_l}\frac{dz}{z} \quad (4)$$

where $\phi(z)$ denotes the spectral density completely defined by the unknown vector θ. The spectral density can be written as $\phi(z) = \mathbf{H}\Lambda\mathbf{H}^*$. Taking the derivative of $\phi(z)$ and Λ with respect to an arbitrary element in θ and inserting into (4) yields

$$\begin{aligned}[\mathbf{J}]_{k,l} &= \operatorname{Tr}\left\{\frac{N}{4\pi j}\oint \mathbf{H}^{-*}\Lambda^{-1}\mathbf{H}^{-1} \right. \\ &\quad (\mathbf{H}_k\Lambda\mathbf{H}^* + \mathbf{H}\Lambda_k\mathbf{H}^* + \mathbf{H}\Lambda\mathbf{H}_k^*) \\ &\quad \mathbf{H}^{-*}\Lambda^{-1}\mathbf{H}^{-1} \\ &\quad \left.(\mathbf{H}_l\Lambda\mathbf{H}^* + \mathbf{H}\Lambda_l\mathbf{H}^* + \mathbf{H}\Lambda\mathbf{H}_l^*)\frac{dz}{z}\right\} \\ &= \operatorname{Tr}\left\{\frac{N}{4\pi j}\oint \mathbf{H}^{-1}\mathbf{H}_k\mathbf{H}^{-1}\mathbf{H}_l \right. \\ &\quad +(\mathbf{H}^{-1}\mathbf{H}_k\mathbf{H}^{-1}\mathbf{H}_l)^* \\ &\quad +\Lambda^{-1}\mathbf{H}^{-1}\mathbf{H}_k\Lambda_l + (\Lambda^{-1}\mathbf{H}^{-1}\mathbf{H}_k\Lambda_l)^* \\ &\quad +\Lambda_k\Lambda^{-1}\Lambda_l\Lambda^{-1} \\ &\quad +\Lambda^{-1}\mathbf{H}^{-1}\mathbf{H}_l\Lambda_k + (\Lambda^{-1}\mathbf{H}^{-1}\mathbf{H}_l\Lambda_k)^* \\ &\quad +\Lambda^{-1}\mathbf{H}^{-1}\mathbf{H}_k\Lambda\mathbf{H}_l^*\mathbf{H}^{-*} \\ &\quad \left.+(\Lambda^{-1}\mathbf{H}^{-1}\mathbf{H}_k\Lambda\mathbf{H}_l^*\mathbf{H}^{-*})^*\frac{dz}{z}\right\}. \end{aligned} \quad (5)$$

The integral above can be considered as the inverse z transform evaluated at $t = 0$. For the causal terms the initial value theorem can then be used, i.e. the value of the inverse z-transform evaluated at $t = 0$ is equal to the value at $z^{-1} = 0$. Since $\mathbf{H}(0) = \mathbf{I}$, the terms 1, 2, 3, 4, 6, 7 are all zero. Term 5 is independent of z and can be moved outside the integral and equation (3b) follows. □

4 Computational aspects

The expression (3) is still very tedious for computation. The computations can be reduced using some of the symmetry of the source separation problem. In this section a TITO system will be considered. Parameterize the problem with the vector θ which contains the coefficients of $B_1, A_1, G_1, F_1, \sigma_1, B_2, A_2, G_2, F_2$, and σ_2 in this order.
Consider the model in (2). The transfer matrix $\mathbf{H}$ can be divided into three parts

$$\begin{aligned}\mathbf{H} &= \begin{bmatrix} 1 & \frac{B_2}{A_2} \\ \frac{B_1}{A_1} & 1 \end{bmatrix} \begin{bmatrix} \frac{G_1}{F_1} & 0 \\ 0 & \frac{G_2}{F_2} \end{bmatrix} \begin{bmatrix} 1 & b_{20} \\ b_{10} & 1 \end{bmatrix}^{-1} \\ &= \mathbf{C}\mathbf{G}\mathbf{D}^{-1} \end{aligned} \quad (6)$$

and the covariance matrix of $\mathbf{e}$ in (2) can in a similar way be written as

$$\begin{aligned}\Lambda &= E\mathbf{e}\mathbf{e}^T = \begin{bmatrix} \sigma_1^2 + b_{20}^2\sigma_2^2 & b_{10}\sigma_1^2 + b_{20}\sigma_2^2 \\ b_{10}\sigma_1^2 + b_{20}\sigma_2^2 & \sigma_2^2 + b_{10}^2\sigma_1^2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & b_{20} \\ b_{10} & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 & b_{20} \\ b_{10} & 1 \end{bmatrix}^T \\ &= \mathbf{D}\bar{\Lambda}\mathbf{D}^T. \end{aligned} \quad (7)$$

Computing the derivatives of $\mathbf{H}$ w.r.t. the elements of θ and denoting $\mathbf{u}_k$ a column vector with unity in position k and zeros elsewhere, yields

$$\frac{\partial \mathbf{H}}{\partial a_{1k}} = -\frac{B_1}{A_1^2}\mathbf{u}_2\mathbf{u}_1^T z^{-k}\mathbf{G}\mathbf{D}^{-1} \quad (8a)$$

$$\frac{\partial \mathbf{H}}{\partial b_{1k}} = \frac{1}{A_1}\mathbf{u}_2\mathbf{u}_1^T z^{-k}\mathbf{G}\mathbf{D}^{-1} \quad (8b)$$

$$\frac{\partial \mathbf{H}}{\partial f_{1k}} = -\frac{G_1}{F_1^2}\mathbf{C}\mathbf{u}_1\mathbf{u}_1^T z^{-k}\mathbf{D}^{-1} \quad (8c)$$

$$\frac{\partial \mathbf{H}}{\partial g_{1k}} = \frac{1}{F_1}\mathbf{C}\mathbf{u}_1\mathbf{u}_1^T z^{-k}\mathbf{D}^{-1} \quad (8d)$$

$$\frac{\partial \mathbf{H}}{\partial \sigma_1^2} = 0 \quad (8e)$$

$$\frac{\partial \mathbf{H}}{\partial a_{2k}} = -\frac{B_2}{A_2^2}\mathbf{u}_1\mathbf{u}_2^T z^{-k}\mathbf{G}\mathbf{D}^{-1} \quad (9a)$$

$$\frac{\partial \mathbf{H}}{\partial b_{2k}} = \frac{1}{A_2}\mathbf{u}_1\mathbf{u}_2^T z^{-k}\mathbf{G}\mathbf{D}^{-1} \quad (9b)$$

$$\frac{\partial \mathbf{H}}{\partial f_{2k}} = -\frac{G_2}{F_2^2}\mathbf{C}\mathbf{u}_2\mathbf{u}_2^T z^{-k}\mathbf{D}^{-1} \quad (9c)$$

$$\frac{\partial \mathbf{H}}{\partial g_{2k}} = \frac{1}{F_2}\mathbf{C}\mathbf{u}_2\mathbf{u}_2^T z^{-k}\mathbf{D}^{-1} \quad (9d)$$

$$\frac{\partial \mathbf{H}}{\partial \sigma_2^2} = 0. \quad (9e)$$

Consider the second term in (3b). It is easily verified that $\mathbf{H}_k\Lambda\mathbf{H}_l^*$ is zero if $\mathbf{H}_k$ corresponds to an arbitrary derivative in (8) and $\mathbf{H}_l$ to an arbitrary derivative in (9), and vice versa. This follows from the property that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal vectors. Thus, the second term in (3b) contributes with a block diagonal matrix to $\mathbf{J}$. The first term of (3b) is zero except when θ_k and θ_l equals b_{10}, b_{20}, σ_1 or σ_2, due to the fact that Λ_k is zero otherwise.

In order to visualize the compact form of $\mathbf{J}$, denote the first and the second term in (3b) with $L_{k,l}$ and $S_{k,l}$ respectively. The Fisher Information matrix then becomes

$$\mathbf{J} = \begin{bmatrix} \mathbf{\Delta}_1 & \mathbf{\Delta}_2 \\ \mathbf{\Delta}_3 & \mathbf{\Delta}_4 \end{bmatrix} + \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \end{bmatrix} \quad (10)$$

where $\mathbf{\Delta}_1, \mathbf{\Delta}_2, \mathbf{\Delta}_3$ and $\mathbf{\Delta}_4$ have nonzero values only in the corners, whereas $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ vanish at the last row and column:

$$\mathbf{\Delta}_1 = \begin{bmatrix} L_{1,1} & 0 & \cdots & 0 & L_{1,m} \\ 0 & & & & 0 \\ \vdots & & & & \vdots \\ 0 & & & & 0 \\ L_{m,1} & 0 & \cdots & 0 & L_{m,m} \end{bmatrix} \quad (11)$$

$$\mathbf{\Sigma}_1 = \begin{bmatrix} S_{1,1} & \cdots & S_{1,m} & 0 \\ \vdots & & \vdots & \vdots \\ S_{m-1,1} & \cdots & S_{m-1,m-1} & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix}. \quad (12)$$

5 Simulations

Using formula (3) and the structure discussed in section 4 the CRLB was computed for a TITO system with FIR-channels and AR(2) source generating filters. The numerical values for the CRLB are presented in Table 1.

	value	CRLB		value	CRLB
b_{10}	0.6	$0.0626/N$	b_{11}	0.5	$0.0628/N$
b_{11}	-0.2	$0.0620/N$	b_{12}	0.1	$0.0622/N$
f_{11}	-1.3	$0.518/N$	f_{21}	1.1	$0.655/N$
f_{12}	0.7	$0.514/N$	f_{22}	0.6	$0.643/N$
σ_1	1	$2.13/N$	σ_2	1	$2.18/N$

Table 1. CRLB values for FIR-channel and AR(2) source filters

Simulations with a Recursive Prediction Error Method (RPEM) applied to the source separation problem are presented in figure 2. For a presentation and analysis of the algorithm, see [1].

From figure 2 it is seen that the variances of the

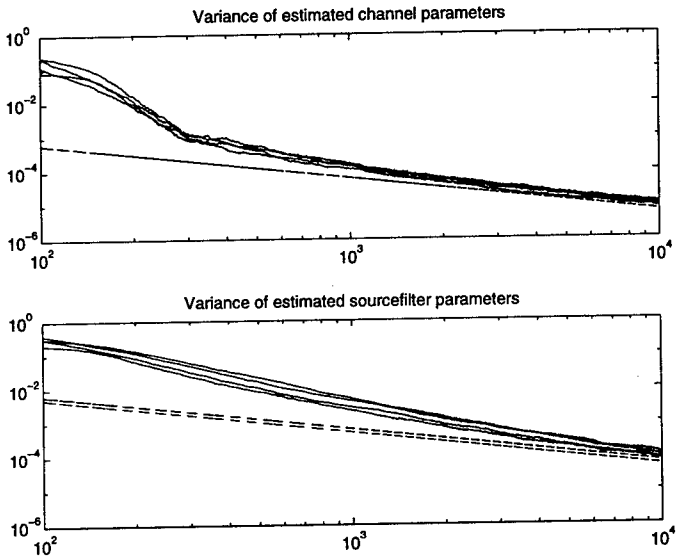


Figure 2. Asymptotic CRLB value (dashed) and parameter variance from an RPEM estimate (solid)

parameters approach the CRLB. The fact that the variances of the source filter parameters deviate from the theoretical CRLB can be explained by the parsimony principle. This is due to the fact that the model used in [1] is overparameterized in the estimation of the noise covariance matrix (7). The estimation of this matrix includes estimation of three parameters, whereas it only depends on two independent variables namly σ_1 and σ_2 in addition to the channel parameters b_{10} and b_{20}.

6 Conclusions

In this paper a formula for the CRLB for the source separation problem is derived. After a reformulation of the problem the CRLB formula was found to be compact and computationally appealing. Values of this bound are computed for a simple test scenario and compared with simulations of a recursive prediction error method.

References

[1] H. Broman, U. Lindgren, H. Sahlin, and P. Stoica. "Source Separation: A TITO System Identification Approach". Technical report, CTH-TE-33, Chalmers University of Technology, Göteborg, Sweden, 1995. Submitted to *IEEE Transactions on Signal Processing.*

[2] C. Jutten and J. Heuralt. "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture". *Signal Processing*, 24:1–10, Jun. 1991.

[3] S. Kay. *"Fundamentals of statistical signal processing, Esimation theory"*. Prentice-Hall, 1993.

[4] U. Lindgren, T. Wigren, and H. Broman. On the convergence of a class of blind separation algorithms. *IEEE Trans. Signal Processing*, 43:3054–3058, December 1995.

[5] T. Söderström and P. Stoica. *System Identification.* Prentice-Hall International, London, UK, 1989.

[6] L. Tong, Y. Inouye, and R. Liu. "Waveform-preserving blind estimation of multiple independent sources". *IEEE Trans. Signal Processing*, 41:2461–2470, Jul. 1993.

[7] E. Weinstein, M. Feder, and A. V. Oppenheim. "Multi-channel signal separation by decorrelation". *IEEE Trans. Speech Audio Processing*, 1:405–413, Oct. 1993.

[8] P. Whittle. "The Analysis of multiple time series". *Journal of Royal Statistical Society*, 1953.

Analysis and comparative evaluation of techniques for multichannel blind deconvolution.*

G. Harikumar and Yoram Bresler†

Coordinated Science Laboratory and the Beckman Institute
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
1308 West Main Street, Urbana, IL 61801, USA
G. Harikumar: hari@csl.uiuc.edu (217) 244 6384
Yoram Bresler: ybresler@uiuc.edu (217) 244 9660
FAX: (217) 244-1642

Abstract

Deterministic multichannel blind deconvolution is an important problem arising in numerous areas of engineering. Recently, two different approaches to solving this problem, Maximum Likelihood techniques (like IQML) and subspace techniques (like EVAM), have been proposed. These methods are theoretically elegant and computationally efficient, and questions arise as to what the properties of these estimators are. We attempt to answer some of these questions in this paper. We show that the subspace based EVAM estimator is a coarse approximation of the IQML estimator. We present a new iterative scheme to compute the M-L estimator, and Cramér-Rao bounds for the channel and input estimates. In addition, we present a Monte-Carlo comparison study of the two estimators and establish the superiority of ML based techniques.

1. Introduction

Formally, the multichannel blind deconvolution problem can be posed in the following manner. Given $\mathbf{y}_1, \ldots, \mathbf{y}_p$ where

$$\mathbf{a} \star \mathbf{x}_i = \mathbf{y}_i, \quad 1 \leq i \leq p \tag{1}$$

and $x_i(0) = 1, \ 1 \leq i \leq p$, it is required to recover the $\mathbf{x}_i \in \mathbb{R}^{n_{x_i}}$ and $\mathbf{a} \in \mathbb{R}^{n_a}$.

This problem arises in various areas of engineering. For example, in seismic signal processing, it is required to recover the seismic trace from its convolutions with different (unknown) acoustic inputs. The problems of blind channel identification and equalization in communication systems, and image restoration when different blurred versions of the same image are available can also be posed in this fashion.

*This work was supported in part by the National Science Foundation grant No: MIP 91-57377, and a Schlumberger-Doll research grant

†Yoram Bresler is on sabbatical leave at the Technion, Israel Institute of Technology during 1995-96

The single channel blind deconvolution problem has multiple solutions and prior knowledge is required for its solution. In contrast, the multichannel blind deconvolution problem in Equation (1) has a unique solution if the channels $\mathbf{x}_i$ are FIR, and their z-transforms have no common zeros [1, 4].

In this paper, we address the deterministic problem, where the input $\mathbf{a}$ and the channels are $\mathbf{x}_i$ treated as unknown, constant vectors. Recently, two different approaches to solving this problem, Maximum Likelihood techniques (like IQML) and subspace techniques (like EVAM), have been proposed. These methods are theoretically elegant and computationally efficient. However, to date their performance limitations in the presence of noise have not been satisfactorily explored. In this paper, we attempt to do this. For the case of additive white Gaussian noise, we

1. Present an efficient algorithm for the computation of the Maximum Likelihood estimate (MLE).
2. Show that the EVAM estimate is an approximation to the result obtained after the first iteration of an IQML-based strategy to compute the M-L estimate.
3. Present an asymptotic analysis of EVAM.
4. Present Cramér-Rao bounds for the channel and input estimates.
5. Present the results of Monte-Carlo comparison studies of the performance of EVAM and the MLE, and demonstrate that the MLE is the superior estimator.

From this point on, we consider the case of p = 2. (Note that both the techniques under study here can be easily extended address the case of more than 2 channels.)

2 The Maximum Likelihood Estimator

Consider the noisy version of the multichannel blind deconvolution problem, defined by

$$\mathbf{y}_i = \mathbf{a} \star \mathbf{x}_i + \sigma \boldsymbol{\eta}_i \quad i \in \{1, 2\} \tag{2}$$

0-8186-7576-4/96 $5.00 © 1996 IEEE

where the η_i are noise vectors whose components $\mathcal{N}(0,1)$ and uncorrelated. In this case, the ML estimator is given by

$$(\mathbf{a}, \mathbf{x}_1, \mathbf{x}_2)_{ML} = \arg\min \sum_{i=1}^{2} \|\mathbf{a} \star \mathbf{x}_i - \mathbf{y}_i\|^2. \quad (3)$$

Consider the case when the $\mathbf{x}_i$ have equal length N, and their z-transforms have no common zeros. (When $n_{x_1} \neq n_{x_2}$, the shorter one can be assumed to be zero-padded). Let $\mathbf{y}_1$ and $\mathbf{y}_2$ be according to Equation (2), and let $\mathbf{y} = [\mathbf{y}_1^T \ \mathbf{y}_2^T]^T$. For $\mathbf{x} \in \mathbb{R}^{n_x}$ and integer m, we define the $(n_x + m - 1 \times m)$ Toeplitz "convolution matrix" $\mathbf{C}_m(\mathbf{x})$ as

$$\mathbf{C}_m(\mathbf{x}) = \begin{bmatrix} x_0 & 0 & \cdots & 0 \\ x_1 & x_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ x_{n_x} & x_{n_x-1} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & x_{n_x-1} \end{bmatrix} \quad (4)$$

Slock et al [10] have shown that the MLE of the channels can be written as

$$(\mathbf{x}_1, \mathbf{x}_2)_{MLE} = \arg\max \mathbf{y}^T \left[\mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\right]\mathbf{y}; \quad (5)$$

$$\text{where } \mathbf{A} = \left[[\mathbf{C}_{n_a}(\mathbf{x}_1)]^T \mid [\mathbf{C}_{n_a}(\mathbf{x}_2)]^T\right]^T. \quad (6)$$

They have further shown that this can be re-written as $(\mathbf{x}_1, \mathbf{x}_2)_{MLE} = \arg\min J(\mathbf{x}_1, \mathbf{x}_2)$ where

$$J(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{y}^T \left[\mathbf{B}^T\left(\mathbf{B}\mathbf{B}^T\right)^{-1}\mathbf{B}\right]\mathbf{y}; \quad (7)$$

$$\mathbf{B} = [[\mathbf{C}_{n_a+N-1}(-\mathbf{x}_2)] \mid [\mathbf{C}_{n_a+N-1}(\mathbf{x}_1)]] \quad (8)$$

The latter step follows from the "minimal null-space parameterization" concept suggested by Slock [10], who have proposed an IQML-type [2] iterative strategy to solve this problem. We propose another scheme based on gradient minimization of the cost function $J()$. Following a development similar to the one in [9], it can be shown that the gradient $\bigtriangledown_J()$ of $J()$ can be written as $\mathbf{Q}(\mathbf{x})\mathbf{x}$, where $\mathbf{Q}$ is an appropriate matrix. A stationary point of $J(x)$ satisfies $\mathbf{Q}(\mathbf{x})\mathbf{x} = \mathbf{0}\mathbf{x}$.This nonlinear eigenvalue problem is solved using the following iterative algorithm.

1. Choose a starting point $\mathbf{x}_0$.
2. For each $\mathbf{x}_k$, construct $\mathbf{Q}(\mathbf{x}_k)$.
3. Choose $\mathbf{x}_{k+1}$ to be the eigenvector corresponding to the smallest absolute eigenvalue of $\mathbf{Q}(\mathbf{x}_k)$.
4. Repeat until convergence.

The computation can be performed efficiently, by using the inverse power method to find the smallest eigenvalue and corresponding eigenvector of $\mathbf{Q}(\mathbf{x})$, and exploiting the structure of $\mathbf{Q}$. Simulations show the algorithm to converge rapidly to the correct solution for moderate and high signal to noise ratios (SNRs).

Simulations indicate that in terms of convergence characteristics and breakdown thresholds, this algorithm is very similar to the IQML-based strategy. Osborne et al [9] show that, when $\mathbf{B}$ is a single Toeplitz matrix matrix with the structure $\mathbf{C}_m$, the eigenvector-iteration algorithm has linear convergence. It is conjectured that a similar result holds for our application.

3 Subspace techniques

Consider Equation (1) for the case when $p = 2$. If Y_1, Y_2, G_1 and G_2 are the z-transforms of the sequences $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{g}_1$ and $\mathbf{g}_2$, respectively and length$(\mathbf{g}_1)$ = length$(\mathbf{g}_2)$ = $\max(n_{x_1}, n_{x_2}) \triangleq N$, all solutions $(G_1(z), G_2(z))$ to the equation

$$Y_1(z)G_1(z) + Y_2(z)G_2(z) = 0 \ \forall \ z \in \mathbb{C} \quad (9)$$

have the form $G_1(z) = \alpha X_2(z)$ and $G_2(z) = -\alpha X_1(z)$ for some $\alpha \in \mathbb{C}$ (ref [4]). This equation (and variants thereof for $p > 2$) forms the basis for the efficient and elegant subspace techniques that seem to have been developed simultaneously and independently by Liu et al [8] and Gurelli et al [4]. Gurelli et al call this technique EVAM, and we shall continue to do so here. Although EVAM can be used successively to solve for $\mathbf{x}_1$ and $\mathbf{x}_2$ even if only upper limits for their lengths n_{x_1} and n_{x_2} are available [4], we limit ourselves to the case when these lengths are known exactly. The EVAM estimator can be used to recover the channels up to a scale factor, as

$$(\mathbf{x}_1, \ \mathbf{x}_2)_E = \arg\min_{\mathbf{x}_i} [\mathbf{x}_2^T \ -\mathbf{x}_1^T]\mathbf{R}_Y^o[\mathbf{x}_2^T \ -\mathbf{x}_1^T]^T; \quad (10)$$

$$\mathbf{R}_Y^o = \left[\mathbf{C}_N^T(\mathbf{y}_1) \mid \mathbf{C}_N^T(\mathbf{y}_2)\right]\left[\mathbf{C}_N^T(\mathbf{y}_1) \mid \mathbf{C}_N^T(\mathbf{y}_2)\right]^T \quad (11)$$

and $\mathbf{C}$ is the convolution matrix defined earlier. (The appropriate scale factor has to come from prior knowledge, say $\|\mathbf{x}_i\| = 1$ or $x_1(0) = 1$.) Equation (11) can be re-written as

$$(\mathbf{x}_1, \ \mathbf{x}_2)_E = \arg\min_{\mathbf{x}_i} \mathbf{y}^T \left[\mathbf{B}^T\mathbf{B}\right]\mathbf{y} \quad (12)$$

where $\mathbf{B}$ is defined as in Equation (8). Thus it can be seen that the subspace technique is an approximation of the first iterate of the IQML technique (with $\mathbf{B}\mathbf{B}^T$ in Equation (7) replaced by the identity).

The EVAM estimate is exact when there is no noise and $\mathbf{a}$ is "persistently exciting" [4], a condition that is satisfied with probability one when the elements of $\mathbf{a}$ are drawn from a continuous probability distribution. However, when there is noise present, the matrix $\mathbf{R}_Y$ formed by the EVAM procedure is a perturbed version of $\mathbf{R}_Y^o$. The eigenvalues of the matrix are also perturbed, and the perturbation is not independent of $\mathbf{a}$. The application of EVAM to noisy data is justified in [4] with the following argument: the perturbation $\mathbf{R}_Y - \mathbf{R}_Y^o$ is a random matrix with mean $\sigma^2\mathbf{I}$ and the variance of each element of the order of $n_a^{-1}\sigma^4$. When n_a tends to infinity, it can be seen that the $\mathbf{R}_Y \approx \mathbf{R}_Y^o + \sigma^2\mathbf{I}$, the eigenvectors of which are the same as those of $\mathbf{R}_Y^o$. So theoretically EVAM can work even for very large noise levels in the presence of sufficiently large data lengths.

In this paper, we present a theorem that provides further evidence of the reliability of EVAM at large data lengths. Using tools found in [3, 7], we have been able to show [5] the following.

Theorem 1. *Let $\hat{\mathbf{x}}_E^i$ be the EVAM estimate of the channel $\mathbf{x}_i$ obtained for the noisy multichannel blind deconvolution problem presented in Equation (2). Then $\hat{\mathbf{x}}_E^i$ is "asymptotically unbiased to second order". This means the following. Let $\sigma \ll 1$ and $n_a \gg n_{x_1} + n_{x_2}$. Then we can write $\hat{\mathbf{x}}_E^i$ as*

$$\hat{\mathbf{x}}_E^i = \mathbf{x}_i + \sum_{j=1}^{\infty} (\sigma n_a^{-0.5})^j \mathbf{t}_{i,j} \tag{13}$$

where the $\mathbf{t}_{i,j}$ do not depend on σ or n_a. (Refer [11]). Let $E_n[]$ denote expectation with respect to different noise realizations. Then $E_n[\mathbf{t}_{i,1}] = \mathbf{0}$, $E_n[\mathbf{t}_{i,2}] = \mathbf{0}$ for $i \in \{1, 2\}$. Thus $E_n[\hat{\mathbf{x}}_E^i] = \mathbf{x}_i + O(\sigma^3 n_a^{-1.5})$.

4 The Cramer-Rao bound

Given a family of distributions $p_{\boldsymbol{\theta}}(\mathbf{z})$ for a random variable $\mathbf{Z} \in \mathcal{Z}$, indexed by a parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the Cramer Rao bound is a lower bound on the covariance of any unbiased estimator $\hat{\boldsymbol{\theta}}(\mathbf{z})$ of $\boldsymbol{\theta}$. It is not always achievable, but is a commonly used benchmark against which the mean square errors of proposed estimators are compared. Let $\mathbf{C}'_m$ to be equal to the convolution matrix $\mathbf{C}_m$ without its first column. With these definitions, and defining $\boldsymbol{\theta} = [x_1(1), \ldots, x_1(n_{x_1} - 1), x_2(1), \ldots, x_2(n_{x_2} - 1), \mathbf{a}^T]^T$, it can be shown [5] that the Cramer-Rao bound for $\boldsymbol{\theta}$ in the scenario in Equation (2) is given by $\boldsymbol{\kappa}(\boldsymbol{\theta})$, where

$$\boldsymbol{\kappa}(\boldsymbol{\theta}) = \sigma^2 \begin{bmatrix} \mathbb{A} & \mathbb{B} \\ \mathbb{B}^T & \mathbb{D} \end{bmatrix}^{-1}; \tag{14}$$

$$\mathbb{A} = \begin{bmatrix} [\mathbf{C}'_{n_{x_1}}(\mathbf{a})]^T [\mathbf{C}'_{n_{x_1}}(\mathbf{a})] & \mathbf{O} \\ \mathbf{O} & [\mathbf{C}'_{n_{x_2}}(\mathbf{a})]^T [\mathbf{C}'_{n_{x_2}}(\mathbf{a})] \end{bmatrix}; \tag{15}$$

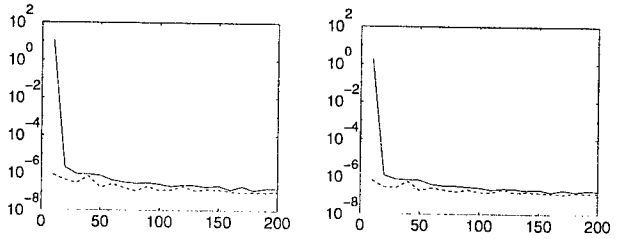
$$\mathbb{B} = \begin{bmatrix} [\mathbf{C}'_{n_{x_1}}(\mathbf{a})]^T [\mathbf{C}_{n_a}(\mathbf{x}_1)] \\ [\mathbf{C}'_{n_{x_2}}(\mathbf{a})]^T [\mathbf{C}_{n_a}(\mathbf{x}_2)] \end{bmatrix}; \tag{16}$$

$$\mathbb{D} = \sum_{i=1}^{2} [\mathbf{C}_{n_a}(\mathbf{x}_i)]^T [\mathbf{C}_{n_a}(\mathbf{x}_i)] \tag{17}$$

Note that $\sigma^2 \mathbb{A}^{-1}$ is the CRB for any unbiased estimator of the channels $\mathbf{x}_i$ when the input $\mathbf{a}$ is known, and $\sigma^2 \mathbb{D}^{-1}$ is the CRB for any unbiased estimator of $\mathbf{a}$ when the channels are known (non-blind case). Equation (14) can be further simplified (using standard results on matrix inversion [6]). The part of the CRB corresponding to the channels, $\boldsymbol{\kappa}(\mathbf{x}_1, \mathbf{x}_2)$ can be shown to be

$$\boldsymbol{\kappa}(\mathbf{x}_1, \mathbf{x}_2) = \sigma^2 \left[\mathbb{A}^{-1} + \mathbb{A}^{-1} \mathbb{B} (\mathbb{D} - \mathbb{B}^T \mathbb{A}^{-1} \mathbb{B})^{-1} \mathbb{B}^T \mathbb{A}^{-1} \right]. \tag{18}$$

It can be seen that the Cramer-Rao bounds for the blind case are greater than those for the non-blind case, as expected.



(a): channels (b): input

Figure 1. Plots of $e_E()$ (solid lines), $e_{MLE}()$ (dashed lines) and $\varsigma()$ (dotted lines) as a function of the input length n_a.

5 Monte Carlo Comparison Studies.

In the studies described in this section, we attempt to evaluate the performances of the EVAM and ML estimators for the 2-channel blind deconvolution scenario described by Equation (2), by comparing them to each other and to the Cramer Rao bound, under different conditions, In what follows, $\hat{\mathbf{x}}_E$ and $\hat{\mathbf{x}}_{MLE}$ are the EVAM and ML estimators for the channels, and $\hat{\mathbf{a}}_E$ and $\hat{\mathbf{a}}_{MLE}$ are similar estimators for the input. $e_E(\mathbf{x}) \triangleq E_n[\|\hat{\mathbf{x}}_E - \mathbf{x}\|^2]/(n_{x_1} + n_{x_2})$ and $e_E(\mathbf{a}) \triangleq E_n[\|\hat{\mathbf{a}}_E - \mathbf{a}\|^2]/(n_a)$, where the expectation E_n is over different realizations of the noise vectors $\boldsymbol{\eta}_i$. $e_{MLE}(\mathbf{x})$ and $e_{MLE}(\mathbf{a})$ are similarly defined. Also, we define the Cramer-Rao bounds $\varsigma(\mathbf{x}) \triangleq \text{trace}(\boldsymbol{\kappa}(\mathbf{x}))/(n_{x_1} + n_{x_2})$ and $\varsigma(\mathbf{a}) \triangleq \text{trace}(\boldsymbol{\kappa}(\mathbf{a}))/(n_a)$.

In the first study described here, we demonstrate the effect of the input length n_a on the 2 estimators. Here is a description of the study.

- Choose n_{x_1}, n_{x_2}, σ.
- Choose random $\mathbf{x}_1 \in \mathbb{R}^{n_{x_1}}, \mathbf{x}_2 \in \mathbb{R}^{n_{x_2}}$.
- Choose $\mathbf{u} \in \mathbb{R}^{200}$.
- For $n_u = 10 : 10 : 200 \mathbf{a} = \mathbf{u}(1 : nu)$.
- Compute $e_E(\mathbf{x}), e_E(\mathbf{a}), e_{MLE}(\mathbf{x}), e_{MLE}(\mathbf{a}), \varsigma(\mathbf{x}), \varsigma(\mathbf{a})$ for each n_u.

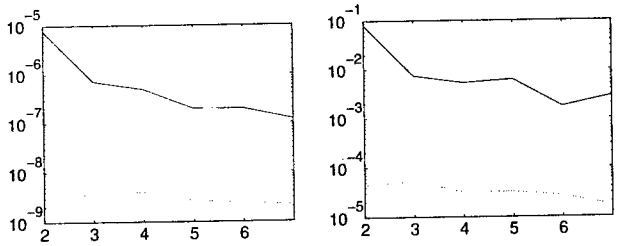
The results of one typical run (for a specific realization of the $\mathbf{x}_i$ and $\mathbf{u}$) is in Figure 1. In this case $n_{x_1} = n_{x_2} = 5$ and $\sigma = 0.001$, corresponding to an SNR of 60 decibels. Note that the σ was chosen to be rather small on purpose. In accordance to our analysis, it ensures that the EVAM estimates are essentially unbiased, thus justifying comparison with the CRB for unbiased estimation.

For all the plots in this chapter, the y-axis is calibrated on a logarithmic scale. Though ML has performed uniformly better than EVAM at all input lengths, the difference in performance becomes small for large n_a.

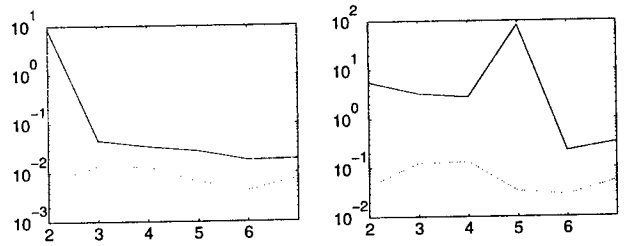
The next study compares the performance of the EVAM and ML estimators at different signal to noise ratios (SNRs). The methodology is the same as in the previous case, except

that the quantities e and ς are computed at different values of σ. The result of a typical run (for a specific choice of the $\mathbf{x}_i$ and $\mathbf{u}$) is shown in Figures 2 (a) through (d).

It can be seen that ML outperforms EVAM by a substantial margin at all SNRs. However, the iterative algorithms for ML do not converge for low SNRs. So the ML estimates used in the e_{MLE} plots in Figures 2(c) and (d) were computed using a descent algorithm. Note that the EVAM technique breaks down much sooner than ML. By SNR = 10 dB, both estimators have broken down.



(a): SNR = 80dB (b): SNR = 40dB

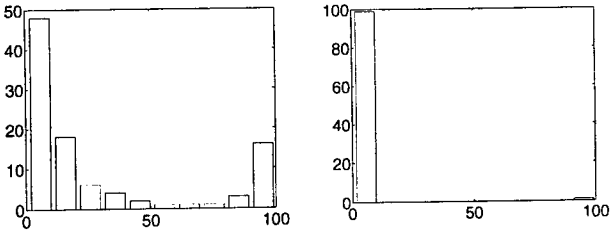


(c): SNR = 20dB (d): SNR = 10dB

Figure 2. Plots of $e_E(\mathbf{x})$ (solid lines) and $e_{MLE}(\mathbf{x})$ (broken lines). The x-axis is numbered in multiples of $\max(n_{x_1}, n_{x_2})$.

The last study compares the errors in the M-L estimates to those in the EVAM estimates and to the Cramer Rao Bound by Monte-Carlo studies that sample various realizations of the channels and the input. The numbers $n_a = 12$, $n_{x_1} = 5$, $n_{x_2} = 5$ and $\sigma = 0.01$ (corresponding to SNR = 40 dB) are fixed, and the vectors $\mathbf{a}$, $\mathbf{x}_1$ and $\mathbf{x}_2$ are generated randomly (according to some distribution) in $\mathbb{R}^{n_a}$, $\mathbb{R}^{n_{x_1}}$ and $\mathbb{R}^{n_{x_2}}$, respectively (100 times for this study). $\varsigma(\mathbf{x})$ and $\varsigma(\mathbf{a})$ are computed for each $(\mathbf{x}_i, \mathbf{a})$. $e_E()$ and e_{MLE} are also computed by an internal Monte-Carlo run, with random noise realizations. Figure 3(a) contains a histogram of the quantity $e_E(\mathbf{x})/e_{MLE}(\mathbf{x})$. It can be seen that the ratios are often very large, indicating that the EVAM estimates are far worse than the M-L estimates. Figure 3(b) contains a similar histogram of the ratio $e_{MLE}(\mathbf{x})/\varsigma(\mathbf{x})$. The ratio is nearly always one, showing essentially efficient performance of the M-L criterion with finite data at this SNR.

Recall that for the value of σ chosen, the inferior performance of EVAM is not due to threshold behavior.



(a): EVAM vs ML (b): ML vs CRB

Figure 3. (a): Histogram of the ratios $e_E(\mathbf{x})/e_{MLE}(\mathbf{x})$ (b): Histogram of the ratio $e_{MLE}(\mathbf{x})/\varsigma(\mathbf{x})$. In each case, the last bin has everything greater than 100.

References

[1] C. A. Berenstein and E. V. Patrick. Exact deconvolution for multiple convolution operators- an overview, plus performance characterizations for imaging sensors. *Proceedings of the IEEE*, 78(4):723–734, April 1990.

[2] Y. Bresler and A. Macovski. Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(5):1081–1089, October 1986.

[3] D. R. Brillinger. *Time Series Data Analysis and Theory*. Holt, Rinehart and Winston, 1974.

[4] M. I. Gurelli and C. L. Nikias. EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 43(1):134–149, January 1995.

[5] G. Harikumar. *Studies in inverse problems with applications to deconvolution*. Prelim proposal, University of Illinois at Urbana-Champaign, July 1995.

[6] H. V. Henderson and S. R. Searle. On deriving the inverse of a sum of matrices. *Siam Review*, 23(1):53–60, January 1981.

[7] M. Kaveh and A. Barabell. The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(2):331–340, April 1986.

[8] H. Liu, G. Xu, and L. Tong. A deterministic approach to blind channel identification of multichannel fir systems. *Proceedings of the ICASSP*, 4:581–584, April 1994.

[9] M. R. Osborne and G. K. Smythe. A modified Prony algorithm for exponential function fitting. *SIAM Journal on Scientific Computing*, 16(1):119–138, January 1995.

[10] D. T. M. Slock and C. B. Papadias. Further results on blind identification and equalization of multiple fir channels. In *Proceedings of the ICASSP*, pages 1964–1967. IEEE, 1995.

[11] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford, 1965.

MULTICHANNEL BLIND SIGNAL DECONVOLUTION USING HIGH ORDER STATISTICS

Eric MOREAU [1] *and Nadège THIRION* [2]

[1] GESSY, ISITV, BP 56, 83162 La Valette du Var Cedex, France
e-mail: moreau@isitv.univ-tln.fr
[2] CEPHAG, ENSIEG, BP46, 38402 Saint Martin d'Hères Cedex, France
e-mail: thirion@cephag.observ-gr.fr

ABSTRACT

The problem of multichannel blind signal deconvolution is considered. We show that input signals can be restored (or separated) using only the condition that they are statistically independent. Two main necessary and sufficient conditions involving high order cumulants are given and proved. Hence, a class of criteria for multichannel signal deconvolution are obtained. Self adaptive gradient based algorithms are derived in order to optimize the proposed criteria and computer simulations are presented in order to demonstrate that the proposed algorithm works.

1. INTRODUCTION

The problem of multichannel blind signal deconvolution (or blind equalization) of Linear Time Invariant (LTI) systems is currently receiving a lot of attention, see [1]-[8] and references therein. The problem finds numerous applications in diverse fields of engineering and applied sciences, e.g. data communication, sonar processing, seismic exploration, antenna processing, speech processing.

In the past ten years most of the proposed approaches consider a restrictive model known as source separation [9]-[12]. Indeed in that case the coupling channels are assumed (unknown) constant gains. Here we consider the more general model in which the coupling channels are unknown LTI systems. It can be simply formulated as follows. Several linear (temporal and spatial) mixtures of certain independent signals called sources are observed. We want to recover the unknown original sources without knowing the mixing filter. Hence, this must be realized from the only knowledge of the observations. This is the reason why this kind of approach is often qualified as "blind" or "unsupervised". In this paper the case of complex signals is considered.

2. PROBLEM FORMULATION

We consider the multichannel LTI and generally non-causal system described by

$$\mathbf{x}(t) = \sum \mathbf{G}(k)\mathbf{a}(t-k) \qquad (1)$$

where $\mathbf{a}(t)$ is the $(N,1)$ vector of *statistically independent* sources, $\mathbf{x}(t)$ is the $(N,1)$ vector of observations and $\{\mathbf{G}(.)\}$ is a sequence of (N,N) matrices which describes the impulse response of the LTI mixing filter.

The multichannel blind deconvolution problem consists in estimating a LTI filter (equalizer) $\{\mathbf{H}(.)\}$ thanks to the only observations $\mathbf{x}(t)$ of an unknown LTI system $\{\mathbf{G}(k)\}$ and such that the vector

$$\mathbf{y}(t) = \sum \mathbf{H}(k)\mathbf{x}(t-k) \qquad (2)$$

restores the N input signals a_i. We define the global LTI filter $\{\mathbf{S}(.)\}$ according to

$$\mathbf{y}(t) = \sum \mathbf{S}(k)\mathbf{a}(t-k)\ . \qquad (3)$$

It is necessary to make the two following assumptions.

A1 Each source a_i is a sequence of zero-mean complex independent and identically distributed (i.i.d.) continuous or discrete random variables. Without any loss of generality they are assumed unit power. Moreover we shall assume that non-zero cumulants of random variables exist and are finite whenever they are introduced. In particular, this implies that sources must be non-Gaussian. Finally we assume that the p-th order joint cumulant of the real and imaginary parts of each source are equal.

A2 The unknown LTI system $\{\mathbf{G}(.)\}$ is assumed stable and invertible.

Notice that assumption **A1** is not very restrictive e.g. in digital communication since most signals have a symmetric constellation, e.g. 4-QAM, 16-QAM, V27.

Because sources are assumed inobservable, there are some inherent indeterminations in their restitution. That is, in general, we cannot identify the order, the power and the time origin of each sources. Indeed this combines the inherent indeterminations of the source separation problem together with those of the classical blind scalar deconvolution problem. Hence signals are said separated if and only if (iff) the global LTI system $\{\mathbf{S}(.)\}$ reads

$$\mathbf{S}(z) \triangleq \sum_k \mathbf{S}(k)z^{-k} = \mathbf{D}(z)\mathbf{D}_1\mathbf{P} \qquad (4)$$

where $\mathbf{D}(z)$ is a diagonal matrix such that its entries are $d_{ii}(z) = z^{-n_i}$, $i = 1,\ldots,N$, n_i integers, $\mathbf{D}_1$ an invertible constant diagonal matrix and $\mathbf{P}$ a permutation matrix.

0-8186-7576-4/96 $5.00

3. DECONVOLUTION CRITERIA

Contrast functions as defined in [2] constitute blind deconvolution criteria in the sense that they are maximum iff the relation in (4) holds for $\mathbf{S}(z)$. In the following "white" vectors $\mathbf{y}$ are considered, i.e. vectors such that

$$\mathsf{E}[\mathbf{y}(t)\mathbf{y}^H(t-\tau)] = \mathbf{I}\delta(\tau) \tag{5}$$

where $\mathbf{I}$ is the (N,N) identity matrix, $\delta(.)$ the dirac distribution and E the mathematical expectation operator. White vectors $\mathbf{y}$ are deduced from sources $\mathbf{a}$ thanks to (3) if

$$\mathbf{S}(z)\mathbf{S}^H(\frac{1}{z^*}) = \mathbf{I}\,. \tag{6}$$

Let us define the two functions $\mathrm{I}_p^{\mathcal{R}}$ and $\mathrm{I}_p^{\mathcal{I}}$ according to

$$\mathrm{I}_p^{\mathcal{R}}(\mathbf{y}) \triangleq \sum_{i=1}^{N} |\mathsf{C}_p\mathcal{R}(y_i)|\,, \quad \mathrm{I}_p^{\mathcal{I}}(\mathbf{y}) \triangleq \sum_{i=1}^{N} |\mathsf{C}_p\mathcal{I}(y_i)| \tag{7}$$

where $\mathsf{C}_p u$ is the p-th order joint cumulant of real random variable u, p an integer greater or equal to 3 and $\mathcal{R}(y_i)$ (resp. $\mathcal{I}(y_i)$) stands for the real (resp. imaginary) part of complex random variable y_i. The following theorems are proved in the paper.

Theorem 1 *The function $\mathrm{I}_p^{\mathcal{R}}(.)$ (resp. $\mathrm{I}_p^{\mathcal{I}}(.)$) for $p \geq 3$ is a contrast over the set of white random vectors having at most one null cumulant of order p of its real part (resp. imaginary part) .*

Proof: We only consider $\mathrm{I}_p^{\mathcal{R}}(.)$ because the proof for $\mathrm{I}_p^{\mathcal{I}}(.)$ is completly similar. Clearly, $\mathrm{I}_p^{\mathcal{R}}(.)$ is symmetrical and invariant under scale change. Let us show that if $\{\mathbf{S}(.)\}$ is such that (6) holds then

$$\mathrm{I}_p^{\mathcal{R}}(\mathbf{Sa}) \leq \mathrm{I}_p^{\mathcal{R}}(\mathbf{a})\,. \tag{8}$$

From (3) one has

$$y_i(t) = \sum_{j,k} s_{ij}(k)a_j(t-k)\,. \tag{9}$$

Thus thanks to the independence of the sources

$$\mathsf{C}_p\mathcal{R}(y_i) = \sum_{j,k} \mathcal{R}^p(s_{ij}(k))\mathsf{C}_p\mathcal{R}(a_j) + (-1)^p\mathcal{I}^p(s_{ij}(k))\mathsf{C}_p\mathcal{I}(a_j)\,. \tag{10}$$

Since $\mathsf{C}_p\mathcal{R}(a_j) = \mathsf{C}_p\mathcal{I}(a_j)$ one has

$$\sum_{i=1}^{N} |\mathsf{C}_p\mathcal{R}(y_i)| \leq \sum_{j=1}^{N} |\mathsf{C}_p\mathcal{R}(a_j)|A_j \tag{11}$$

where

$$A_j = \sum_{i,k} (|\mathcal{R}(s_{ij}(k))|^p + |\mathcal{I}(s_{ij}(k))|^p)\,. \tag{12}$$

Now from (6), $\forall j$, $\sum_{i,k} |s_{ij}(k)|^2 = 1$, thus $\forall j$, $A_j \leq 1$ and (8) is realized.

Let us consider the equality in (8). If one source, say a_N, is such that $\mathsf{C}_p\mathcal{R}(a_N) = 0$ then equality in (8) requieres equality $A_j = 1$ for $j = 1,\ldots,N_1$ which holds if it exists one and only one (i,j), $i = 1,\ldots,N$; $j = 1,\ldots,N_1$ and $\forall k$ such that $|\mathcal{R}(s_{ij}(k))| = 1$ or $|\mathcal{I}(s_{ij}(k))|$. Because $\{\mathbf{S}(.)\}$ is such that (6) then $\mathbf{S}(z)$ is of the form (4) and $\mathrm{I}_p^{\mathcal{R}}(.)$ is a contrast over the set of white random vector having at most one null cumulant of order p of its real part. •

Hence by the theorem, for white random vectors $\mathbf{y}$ deduced from eq.(3), necessary and sufficient condition for blind deconvolution is

$$\sum_{i=1}^{N} |\mathsf{C}_p\mathcal{R}(y_i)| = \sum_{i=1}^{N} |\mathsf{C}_p\mathcal{R}(a_i)|\,, \tag{13}$$

or

$$\sum_{i=1}^{N} |\mathsf{C}_p\mathcal{I}(y_i)| = \sum_{i=1}^{N} |\mathsf{C}_p\mathcal{I}(a_i)|\,. \tag{14}$$

This leads to the two following constrained blind deconvolution criteria

$$\max \sum_{i=1}^{N} |\mathsf{C}_p\mathcal{R}(y_i)| \quad \text{subject to} \quad \mathbf{y} \text{ white} \tag{15}$$

$$\max \sum_{i=1}^{N} |\mathsf{C}_p\mathcal{I}(y_i)| \quad \text{subject to} \quad \mathbf{y} \text{ white} \tag{16}$$

Now in the specific case of sources a_i with identical sign ε_p of the p-th order cumulant of $\mathcal{R}(a_i)$ and $\mathcal{I}(a_i)$ for all i, we have the following theorem.

Theorem 2 *For even integer $p > 3$, the functions*

$$\mathrm{J}_p^{\mathcal{R}}(\mathbf{y}) \triangleq \varepsilon_p \sum_{i=1}^{N} \mathsf{C}_p\mathcal{R}(y_i) \quad \text{and} \quad \mathrm{J}_p^{\mathcal{I}}(\mathbf{y}) \triangleq \varepsilon_p \sum_{i=1}^{N} \mathsf{C}_p\mathcal{I}(y_i)$$

are contrasts over the set of white random vectors having non zero cumulant of order p of its real and imaginary part.

The proof is easily deduced from Theorem 1 and eq.(10) where if p is even then $\mathrm{sign}(\mathsf{C}_p\mathcal{R}(y_i)) = \mathrm{sign}(\mathsf{C}_p\mathcal{R}(a_i)) = \varepsilon_p$.

If we consider the value $p = 4$, we have the following simplified theorem.

Theorem 3 *The functions*

$$\mathrm{K}^{\mathcal{R}}(\mathbf{y}) \triangleq \varepsilon_4 \sum_{i=1}^{N} \mathsf{E}\mathcal{R}^4(y_i) \quad \text{and} \quad \mathrm{K}^{\mathcal{I}}(\mathbf{y}) \triangleq \varepsilon_4 \sum_{i=1}^{N} \mathsf{E}\mathcal{I}^4(y_i)$$

are contrasts over the set of white random vectors having non zero cumulant of order 4 of its real and imaginary part.

Proof: We only consider $\mathrm{K}^{\mathcal{R}}$. One has $\mathsf{C}_4\mathcal{R}(y_i) = \mathsf{E}\mathcal{R}^4(y_i) - 3\mathsf{E}^2\mathcal{R}^2(y_i)$. Since white vectors are considered $\mathsf{E}\mathcal{R}^2(y_i)$ is constant $\forall i$. Thus $\mathrm{K}^{\mathcal{R}}(\mathbf{y}) = \mathrm{J}_4^{\mathcal{R}}(\mathbf{y}) + \mathrm{cst}$ where cst is a certain constant. Then the theorem is proved. •

As previously we can deduce necessary and sufficient conditions for blind deconvolution and the corresponding maximization criteria.

4. SELF-ADAPTIVE ALGORITHM

In order to achieve the deconvolution, we have to find a filter $\{\mathbf{H}\}$ such that the proposed contrasts are maximum. A stochastic gradient based adaptive algorithm is proposed in this section. The set of definition of the proposed contrast is the set of white vectors. Hence in the following we consider that a first stage realize a multichannel spectral prewhitening of the observations. This "classical" stage will not be discussed here. In order to ensure the whiteness of $\mathbf{y}$, $\{\mathbf{H}\}$ must be such that

$$\mathbf{H}(z)\mathbf{H}^H(\frac{1}{z^*}) = \mathbf{I} \tag{17}$$

that is the filtering transfer matrix is lossless or all-pass. Such transfer admits a special parametrization thanks to planar (Givens) rotations, see e.g. [4]. In the simplest case ($N = 2$, $k = 0, 1$) one has

$$\mathbf{H}(z) = \mathbf{Q}_1(\theta_1, \phi_1) \begin{pmatrix} z^{-1} & 0 \\ 0 & 1 \end{pmatrix} \mathbf{Q}_2(\theta_2, \phi_2) \tag{18}$$

where

$$\mathbf{Q}_i(\theta_i, \phi_i) = \begin{pmatrix} e^{j\phi_i}\cos\theta_i & \sin\theta_i \\ -\sin\theta_i & e^{-j\phi_i}\cos\theta_i \end{pmatrix}. \tag{19}$$

Using this parametrization, we have now to find the angles θ_i and ϕ_i in order to maximize one contrast. Denoting p anyone of parameters (θ_i, ϕ_i), a deterministic procedure is to reach the maximum of a contrast $\mathcal{C}$ thanks to an iterative algorithm which updates p with the increment

$$\Delta p = \mu \frac{\partial \mathcal{C}}{\partial p} \tag{20}$$

where μ is a small positive constant. Hence the optimum is found as the limit of the sequence

$$p(n) = p(n-1) + \mu \left.\frac{\partial \mathcal{C}}{\partial p}\right|_{p=p(n-1)}. \tag{21}$$

In cases of the contrast in this paper, it is possible to express the criteria as the expectation of some random variable. We use a loss complex version of the gradient algorithm (21) by dropping the expectation. It will be called a "stochastic algorithm". For $N = 2$ and contrast $\mathrm{K}^{\mathcal{R}}(.)$, one easily has the stochastic increment

$$\Delta p = 4\mu\varepsilon_4(\mathcal{R}^3(y_1)\frac{\partial \mathcal{R}(y_1)}{\partial p} + \mathcal{R}^3(y_2)\frac{\partial \mathcal{R}(y_2)}{\partial p}) \tag{22}$$

where $\partial\mathcal{R}(y_1)/\partial p$ are deduced from (2) and (18).

Convergence analysis of the proposed algorithm is beyond the scope of this paper. However computer simulations are presented in order to demonstrate that the proposed algorithm works.

5. COMPUTER SIMULATIONS

The performances of the algorithm are associated to an index/measure of performance defined on the global filtering matrix $\{\mathbf{S}\}$ according to

$$\mathrm{ind}(\{\mathbf{S}\}) \triangleq \frac{1}{2}\left[\sum_i \left(\sum_{j,k} \frac{|s_{ij}(k)|^2}{\max_{\ell,m}|s_{i\ell}(m)|^2} - 1\right) + \sum_j \left(\sum_{i,k} \frac{|s_{ij}(k)|^2}{\max_{\ell,m}|s_{\ell j}(m)|^2} - 1\right)\right]$$

This positive index is indeed zero if $\{\mathbf{S}\}$ is such that $\mathbf{S}(z)$ satisfies (6) and a small value indicates the proximity to the desired solution. We present simulations in the case of two sources. Three kind of sources are considered: i) two 4-QAM communication sources; ii) two 16-QAM communication sources and iii) two constant modulus sources: $\exp(j\phi)$ where ϕ is a random variable with uniform probability density over $[0, 2\pi[$. The mixing filter is of the form (18) where $\theta_1 = \pi/6$, $\phi_1 = \pi/18$, $\theta_2 = \pi/9$ and $\phi_2 = \pi/36$. The algorithm (22) is tested via Monte Carlo simulations. In Fig.1, 2 and 3 we have plotted the sample average over 500 data realizations of the index as a function of iterations respectively in cases i), ii) and iii). The index decreases monotically and achieve the steady state level of -33dB, -27dB and -28dB respectively in the three cases. In Fig.4 and 5 we have plotted one realization of the performance index, the estimated parameters, the observed signals and the reconstructed signals at channel 1 when steady state is achieved.

6. REFERENCES

[1] V. Capdevielle, C. Serviere and J.L. Lacoume, "Separation of Wideband Sources", Proc. *HOS'95, IEEE Workshop on HOS*, Girona, Spain, pp 66-70, 1995.

[2] P. Comon, "Contrasts for Multichannel Blind Deconvolution", *I3S-CNRS Research Report No 95-44*, September 8, 1995, submitted to IEEE SPL.

[3] Y. Inouye and T. Habe, "Multichannel Blind Equalization Using Second and Fourth Order Cumulants", Proc. *HOS'95, IEEE Workshop on HOS*, Girona, Spain, pp 96-100, 1995.

[4] P. Loubaton and P. Regalia, "Blind Deconvolution of Multivariate signals: A Deflation Approach", Proc. *ICC'93, International Conference on Communication*, Geneva, Switzerland, Vol. 2, pp 1160-1164, May 1993.

[5] H.L. Nguyen Thi and C. Jutten, "Blind Source Separation for Convolutive Mixtures", *Signal Processing*, Vol. 45, pp 209-229, 1995.

[6] A. Swami, G. Giannakis, and S. Shamsunder, "Multichannel ARMA Processes", *IEEE Transactions on SP*, Vol. 42, No. 4, pp 898-913, April 1994.

[7] N. Thirion, "Séparation d'ondes en prospection sismique", PhD Thesis, INPG, September 1995.

[8] D. Yellin and E. Weinstein, "Criteria for Multichannel Signal Separation", *IEEE Transactions on SP*, Vol. 42, No. 8, pp 2158-2168, August 1994.

[9] J.F. Cardoso and A. Souloumiac, "Blind Beamforming for non Gaussian Signals", *IEE Proceedings F*, Vol. 40, pp 362-370, 1993.

[10] P. Comon, "Independent Component Analysis, a New Concept?", *Signal Processing*, Vol. 36, pp 287-314, 1994.

[11] C. Jutten and J. Herault, "Blind Separation of Sources, Part I: An Adaptative Algorithm Based on Neuromimetic Architecture", *Signal Processing*, Vol. 24, pp 1-10, 1991.

[12] E. Moreau and O. Macchi, "High Order Contrasts for Self-Adaptive Source Separation", *International Journal of Adaptive Control and Signal Processing*, Vol. 10, No. 1, pp 19-46, January 1996.

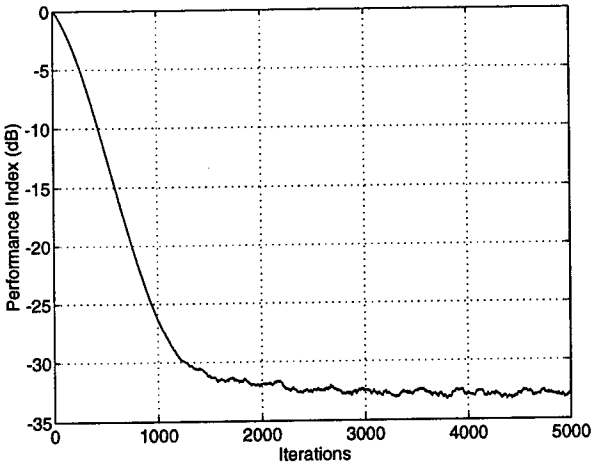


Figure 1: Sample average of the performance index for two 4-QAM sources.

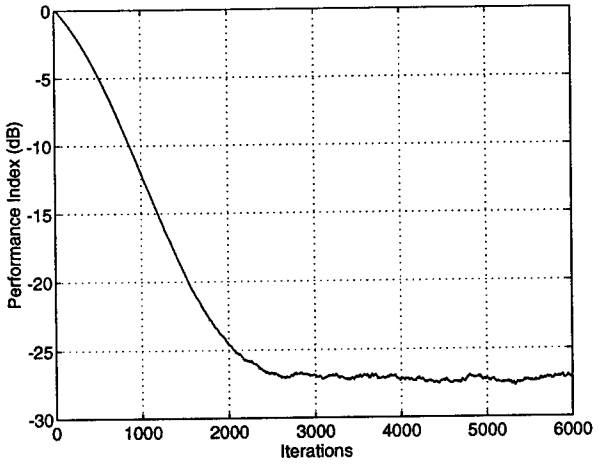


Figure 2: Sample average of the performance index for two 16-QAM sources.

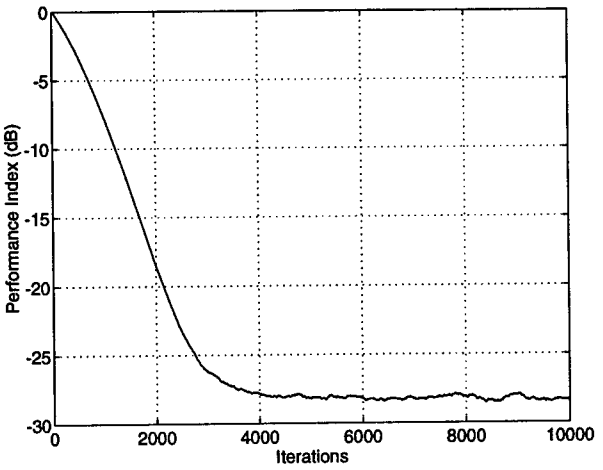


Figure 3: Sample average of the performance index for two constant modulus sources.

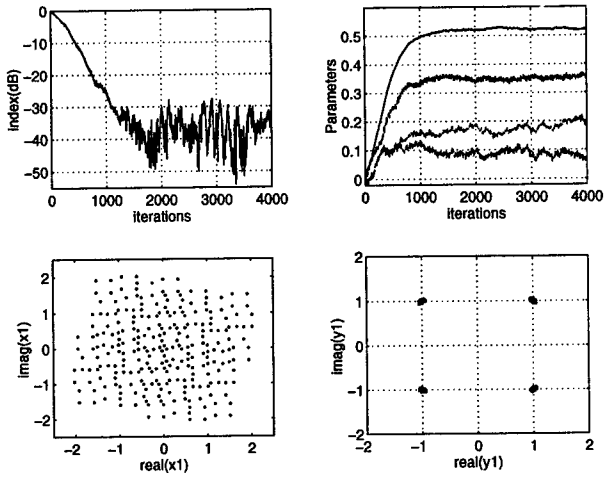


Figure 4: Performance index + parameters + observed signals + reconstructed signals for two 4-QAM sources.

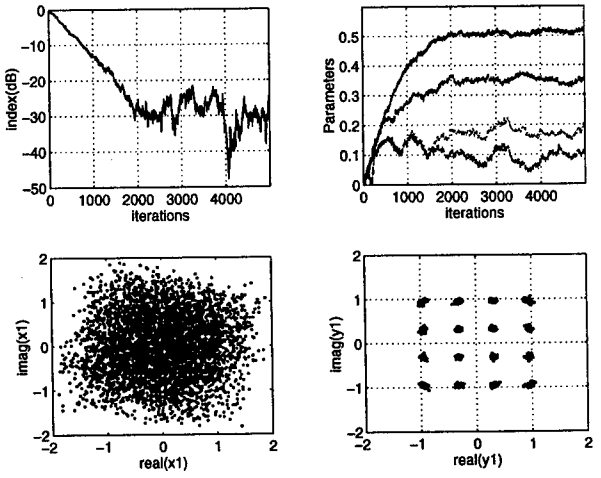


Figure 5: Performance index + parameters + observed signals + reconstructed signals for two 16-QAM sources.

Resolving Power of Spectral Matrix Filtering : A Discussion on the Links Steering Vectors / Eigenvectors

Nadège THIRION, Jean-Louis LACOUME, Jérôme MARS

Centre d'Etude des PHénomènes Alétoires et Géophysiques (CEPHAG),
URA CNRS 346, BP 46, rue de la Houille Blanche, 38402 ST MARTIN D'HÈRES Cedex France,
Tél/Fax : 76.82.64.21 / 76.82.63.84, e-mail : thirion@cephag.observ-gr.fr

Abstract- *Problems of separation of convolutive mixtures of wideband signals impinging on an antenna of sensors often arise in signal processing. In seismic signal processing, techniques have been developed to perform separation of seismic waves (f-k or median filters, spectral matrix filtering, Radon or Karhunen-Loeve transforms, Maximum Likelihood methods). They give good results in most cases, yet, their limits might occur in difficult contexts (waves of very close energies or/and near slowness). We analytically study the resolving power of spectral matrix filtering to theoretically explain why the method does not work any more for waves of close energies. This problem brings us to the question of links between two basis : eigenvectors and steering vectors.*

Keywords- *Array Processing, Spectral Matrix Filtering, Seismic Signal Separation, Blind Processing of Wideband Signals.*

I. Introduction

Problems of separation of convolutive mixtures of wideband signals impinging on an antenna of sensors are widely spread. Typical examples can be found in passive sonar, geophysics, etc... In geophysical operations, the aims of signal processing are the separation and the identification of waves to get a better understanding of the onshore. Techniques have been developed to achieve these purposes (Karhunen-Loeve transform [5], f-k filter, median filter [4], spectral matrix filtering [7,8], Radon transform [2], Maximum Likelihood Estimator [3]). They give good results in many cases but their limits might occur for waves of very close energies or too near slowness. Focusing on spectral matrix filtering, we determine its resolving power by analytically studying links between two basis : on the one hand the eigenvectors basis which is the mathematical object given by the eigendecomposition of the spectral matrix of observed signals and on the other hand the steering vectors basis which is the physical object we are interested in. We explain how these two basis fit together. This fitting depends on different parameters, yet, our choice was to express results versus a geometrical criteria (i.e. the spatial coherency of waves vectors) and the energy ratio of the sources.

II. Theoretical background

II.1. The model

We suppose that the antenna is linear and composed by N sensors. The signal $r_k(t)$ recorded on the k^{th} sensor is a linear combination of the p detected waves, plus an additive noise [9]. This noise is supposed to be spatially and spectrally white, gaussian and independent of the signals of interest. Its spectral density is noted σ_b^2. These assumptions are written in the time domain as follows :

$$r_k(t) = \sum_{i=1}^{p} s_{k,i}(t) * a_i(t) + b_k(t) = \sum_{j=1}^{p} o_{k,j}(t) + b_k(t) \quad (1)$$

where * is the convolution operator, $a_i(t)$ is a deterministic amplitude term (referred to as the i^{th} source or wave-front) : it does not convey information about the propagation, $s_{k,i}(t)$ describes the propagation of the i^{th} wave recorded on the k^{th} sensor and $b_k(t)$ stands for the noise. The transcription of (1) into matrix notations gives :

$$R(t) = \mathbf{S}(t) * A(t) + B(t) = \sum_{j=1}^{p} O_j(t) + B(t) \quad (2)$$

using following notations :

- $R(t) = \left[r_1(t), \cdots, r_N(t)\right]^T$ is the $(N,1)$ vector of the observations. ^{T} stands for the transposition operation.
- $\mathbf{S}(t) = \left[S_1(t), \cdots, S_p(t)\right]$ is a (N,p) matrix whose k^{th} columns is the so-called k^{th} *steering vector* expressed as : $S_k(t) = \left[s_{1,k}(t), \cdots, s_{N,k}(t)\right]^T$.

The phase of its first component is assumed to be null which implies that the first sensor is chosen as a reference. This convention ensures the unity of the sources. Besides, these steering vectors describe propagation on the antenna. Under the plane waves assumption with neither attenuation nor dispersion, the complex gain between two sensors reduces to a pure phase term. But in the general case, more complex phenomena have to be taken into account.

- $O_j = S_j(t) * a_j(t)$ is the j^{th} wave vector.

Equation (3) is obtained by Fourier transforming (2). The general problem is then divided into a set of problems of separation of instantaneous mixtures of signals. The calculus at a given frequency bin does not depend any more on those made at other frequency bins.

$$R(\nu) = \underbrace{\mathbf{S}(\nu).\mathbf{D}(\nu)}_{\mathbf{S}'(\nu)} . \underbrace{\mathbf{D}^{-1}(\nu)A(\nu)}_{A'(\nu)} + B(\nu) \qquad (3)$$

We will focus on the problem of separation of colored but uncorrelated sources A. The diagonal renormalisation matrix $\mathbf{D}$ we have introduced, ensures spectral whitening of the sources (i.e. the new sources $A'(\nu)$ have unit power). Whatever $\mathbf{S}'(\nu)$, matrix of the new steering vectors, its Singular Values Decomposition (*SVD*) is given by [12] :

$$\mathbf{S}'(\nu) = \mathbf{V}(\nu).\Delta^{1/2}(\nu).\Pi(\nu) \qquad (4)$$

where $\mathbf{V}$ is a unitary (N,N) matrix (i.e. $\mathbf{V}.\mathbf{V}^H = \mathbf{I}_N$; ^{H} denotes transconjugaison operator and $\mathbf{I}_N$ is the (N,N) identity matrix), Δ is a (N,p) diagonal matrix whose N-p last lines are null (it is obviously supposed that $N>p$), Π is a (p,p) unitary matrix, parametered in the general case as a product of Givens rotation matrices (Π'), multiplied by a diagonal matrix of pure phase terms ($\mathbf{P}$) [12]. In the most simple case which is the two waves case, its expression simplifies to :

$$\Pi(\theta,\kappa,\psi_1,\psi_2) = \Pi'.\mathbf{P} = \begin{pmatrix} \cos\theta(\nu) & \sin\theta(\nu).e^{j\kappa(\nu)} \\ -\sin\theta(\nu).e^{-j\kappa(\nu)} & \cos\theta(\nu) \end{pmatrix} . \begin{pmatrix} e^{j\psi_1(\nu)} & 0 \\ 0 & e^{j\psi_2(\nu)} \end{pmatrix} \qquad (5)$$

It only depends on four parameters which vary with the frequency.

II.2. Eigendecomposition of spectral matrix and estimation of matrices V and Δ

To analytically determine the two matrices $\mathbf{V}$ and Δ involved in the parametrisation of $\mathbf{S}'$, we build the spectral matrix $\Gamma(\nu)$, related to the observations $R(\nu)$ and defined as :

$$\Gamma(\nu) = \zeta\left[R.R^H\right] = \begin{cases} \zeta\left[\mathbf{S}'.A'.A'^H.\mathbf{S}'^H\right] + \zeta\left[B.B^H\right] \\ \mathbf{V}.\left(\Delta + \sigma_b^2.\mathbf{I}_N\right).\mathbf{V}^H \end{cases} \qquad (6)$$

where ζ stands for an averaging operator. Equation (6) is obtained by reintroducing the parametrisation of the propagation matrix $\mathbf{S}'$ that was given in equation (4). It can also be identified with the eigendecomposition of the spectral matrix because of the uniqueness of this one. Thus, eigendecomposition enables the determination of two of the matrices that are looked for : the p first columns of matrix $\mathbf{V}$ are the p first eigenvectors of matrix Γ (assuming that eigenvalues have been arranged in a descending way). In the same way, the p largest eigenvalues λ_k of Γ are related to Δ. In fact, we have :

$$\Delta = \begin{pmatrix} \sqrt{\lambda_1 - \hat{\sigma}_b^2} & 0 & \cdots \\ 0 & \sqrt{\lambda_2 - \hat{\sigma}_b^2} & \\ \vdots & \ddots & \ddots \end{pmatrix} \qquad (7)$$

The eigenvectors associated with the p largest eigenvalues belong to the same subspace (called the Signal Subspace (*SS*)) as the one spanned by the p steering vectors of the desired waves vectors. Yet, nothing guarantees the exact fitting between these two basis. This is obviously due to the fact that the unitary matrix Π also involved in the equation statement is not reachable by this own treatment. We can even notice that eigenvectors define an orthonormal basis whereas steering vectors are not necessarily orthogonal.

In next sections, we explain how the two basis fit together, and we quantify resolving power of the spectral matrix filtering. The analytical calculations prove that, in most cases, treatments based on exploitation of second order properties of received signals are not sufficient to separate waves but enable extraction of the most energetic one. To reach separation, treatments have to be completed, in other words matrix Π has to be estimated. In blind separation of wideband independent sources, this matrix is determined using the fact that it has to lead to most independent sources [6,1] in the sense of a higher order criteria. Blind separation of seismic waves has been performed replacing this criteria by a local distance stationarity criteria applied on the phases of the estimated wave vectors [10,11].

III. Spectral matrix filtering : resolving power

We now focus on the case of two uncorrelated plane waves. The two vectors $V_1(\nu)$ and $V_2(\nu)$ associated with the two largest eigenvalues λ_1 and λ_2 have to be analytically calculated. To reach this purpose, we exploit the two following properties : these vectors are eigenvectors of matrix $\Gamma(\nu)$ (equation 8) and they are linear combination

of steering vectors because of their belonging to the *SS* (equation 9) :

$$\begin{aligned}\Gamma(\nu).V_k(\nu) &= \lambda_k.V_k(\nu) \qquad \forall k=1..2 \\ &= \left(S_1'.S_1'^H + S_2'.S_2'^H + \sigma_b^2.\mathbf{I}_N\right).V_k(\nu)\end{aligned} \tag{8}$$

$$\begin{cases} V_1(\nu) = c_1 S_1' + c_2 S_2' \\ V_2(\nu) = d_1 S_1' + d_2 S_2' \end{cases} \tag{9}$$

where c_1, c_2, d_1, d_2 are complex numbers.
This set of hypothesis leads to the following system :

$$\begin{cases} c_1(\lambda_1 - \sigma_b^2 - Pa_1) = c_2 S_1'^H.S_2' \\ c_2(\lambda_1 - \sigma_b^2 - Pa_2) = c_1 S_2'^H.S_1' \end{cases} \tag{10}$$

where $Pa_i = S_i'^H.S_i' = \| S_i' \|^2$. To solve this system, different cases have to be distinguished :

(i) Waves are geometrically orthogonal (i.e. $S_1'.S_2'^H=0$) but sources have different energies, then the eigenvector which is associated with the largest eigenvalue is collinear to the steering vector of the most energetic wave, and the eigenvector associated with the second eigenvalue is collinear to the steering vector of the less energetic wave. This appears in eq. (11). The treatment is completed at the end of the second order stage to the extent that the found basis already coincides with the wanted basis :

$$\begin{cases} \lambda_1 = \sigma_b^2 + Pa_1 \;\; ; \;\; V_1 = (1/\sqrt{Pa_1}).S_1' \\ \lambda_2 = \sigma_b^2 + Pa_2 \;\; ; \;\; V_2 = (1/\sqrt{Pa_2}).S_2' \end{cases} \tag{11}$$

(ii) The case of orthogonal waves with the same energy is a singular one. Eigenvalues are found to be always identical. Whatever the vector belonging to the space spanned by steering vectors, it is an eigenvector. The system always remains undetermined...

(iii) We now suppose that the waves are not orthogonal. It can be easily established that the two largest eigenvalues of the spectral matrix are given by :

$$\lambda_{1/2} = \frac{1}{2}.\left(Pa_1 + Pa_2 \pm \sqrt{\left(Pa_1 - Pa_2\right)^2 + 4.\left|S_2'^H.S_1'\right|^2} \right) + \sigma_b^2$$

A condition about c_1, c_2 is deduced :

$$c_1/c_2 = \frac{2.S_1'^H.S_2'}{\left(Pa_2 - Pa_1\right) + \sqrt{\left(Pa_1 - Pa_2\right)^2 + 4.\left|S_2'^H.S_1'\right|^2}}$$

We obtain the same kind of relation for d_1, d_2. These two ratios are representative of the geometrical organization between the two considered basis. The transformation which ensures the passing from one basis to the other one is the multiplication by a compression matrix (Δ) and a unitary matrix expressed as a complex rotation matrix. In the two waves case, it becomes :

$$\mathbf{S}'(\nu) = \mathbf{V}.\Delta^{1/2}.\Pi =$$

$$\left(V_1\ V_2\right).\begin{pmatrix} \sqrt{\lambda_1 - \sigma_b^2} & 0 \\ 0 & \sqrt{\lambda_2 - \sigma_b^2} \end{pmatrix}.\begin{pmatrix} \cos\theta.e^{j.\psi_1} & \sin\theta.e^{j.(\psi_2+\kappa)} \\ -\sin\theta.e^{j.(\psi_1-\kappa)} & \cos\theta.e^{j.\psi_2} \end{pmatrix}$$

Conditions on coefficients c_1, c_2, d_1, d_2 are deduced from this last equality :

$$c_2/d_2 = \sqrt{(\lambda_2 - \sigma_b^2)/(\lambda_1 - \sigma_b^2)}.\tan\theta.e^{-j\kappa}$$

$$c_1/d_1 = \sqrt{(\lambda_2 - \sigma_b^2)/(\lambda_1 - \sigma_b^2)}.\frac{1}{\tan\theta}.e^{-j(\kappa+\pi)} \tag{12}$$

Thus we have to parameter the unknowns. Uniqueness of this parametrisation is ensured by the normalisation of the eigenvectors :

$$\arg(c_1) + \psi_1 = 0$$
$$\arg(d_2) + \psi_2 = 0 \qquad \kappa = \arg(d_2) - \arg(c_2)$$

$$\left.\begin{aligned} |c_1| &= \cos\theta / \sqrt{\lambda_1 - \sigma_b^2} \\ |c_2| &= \sin\theta / \sqrt{\lambda_1 - \sigma_b^2} \end{aligned}\right\} \Rightarrow \left|c_2/c_1\right| = \tan\theta$$

and :

$$\left.\begin{aligned} |d_1| &= \sin\theta / \sqrt{\lambda_2 - \sigma_b^2} \\ |d_2| &= \cos\theta / \sqrt{\lambda_2 - \sigma_b^2} \end{aligned}\right\} \Rightarrow \left|d_2/d_1\right| = 1/\tan\theta$$

We now quantify the dependency of angles θ and κ of the unitary matrix on parameters of interest. In our case, the two desired angles are expressed versus E the energy ratio of the sources ($E=Pa_2/Pa_1$) and the spatial coherency ρ between the two waves. ρ is the normalized scalar product between steering vectors (it is a geometrical criteria). In the case of plane waves, with equispaced sensors, we have :

$$\rho(\nu) = \frac{S_1'^H(\nu).S_2'(\nu)}{\sqrt{\|S_1'(\nu)\|^2.\|S_2'(\nu)\|^2}} = \frac{1}{N}.\frac{\sin(N.\Delta\Phi(\nu))}{\sin(\Delta\Phi(\nu))}.e^{j(N-1)\Delta\Phi(\nu)}$$

with : $\Delta\Phi = \frac{\Phi_1 - \Phi_2}{2}$

The steering vector of the k^{th} wave is defined in the frequency domain by :

$$S_k'(\nu) = \sqrt{Pa_k}.S_k(\nu) = \sqrt{\frac{Pa_k}{N}}.\begin{pmatrix} 1 & e^{-j.2\pi\nu.\tau_{1,k}} & \cdots & e^{-j.2\pi\nu.\tau_{N-1,k}} \end{pmatrix}^T$$

In the plane wave case with equispaced sensors, the propagation matrix has a special structure : it is a *Vandermonde* matrix. The time delay on m^{th} sensor is then given by :

$$\tau_{m,k} = \frac{m.d.\sin(\theta_k)}{c} = \frac{m.\Phi_k}{2.\pi.\nu}$$

where d is the distance between two sensors, c the propagation velocity of the sound, and θ_k the angle of arrival on the antenna of the k^{th} wave.
The module of the spatial coherency varies between 0 and 1 ; $\rho = 0$ for geometrically orthogonal waves. It becomes

true if the number of sensors is great and the angles of arrival are different ; $\rho = 1$ for collinear waves. Finally, we find that :

$$\left|\frac{c_2}{c_1}\right| = \tan\theta = \frac{(E-1)+\sqrt{(1-E)^2+4.|\rho|^2.E}}{2.|\rho|.\sqrt{E}}$$

and $\kappa = \psi_1 - \psi_2 - (N-1).\Delta\Phi_{2,1}$

It is also possible to get the expression of eigenvectors, which will make it possible to quantify the resolving power of the spectral matrix. We have established that :

$$\frac{\text{Power of } S_1 \text{ on } V_1}{\text{Power of } S_2 \text{ on } V_1} = \frac{1}{(\tan\theta)^2} = \frac{\text{Power of } S_2 \text{ on } V_2}{\text{Power of } S_1 \text{ on } V_2}$$

Waves of identical energy characterize a singular case because angle θ does not depend on spatial coherency any more. It remains equal to 45° (figure 1). Moreover it is the less favorable one in terms of separation to the extent that, after the second order stage, sources still remain totally mixed (the same proportion of each source on both whitened signals (figure 2)). In the case of orthogonal waves (spatial coherency coefficient equals 0), angle θ remains equal to 0° (separation is achieved after simple projection onto eigenvectors). In all other cases, the separation is still not performed after the second order stage, but on the first eigenvector : proportion of the most energetic source is widely superior to the proportion of the least energetic source. In spite of the fact that second source is less energetic, its proportion remains superior to the proportion of most energetic source, as far as the second eigenvetor is concerned.

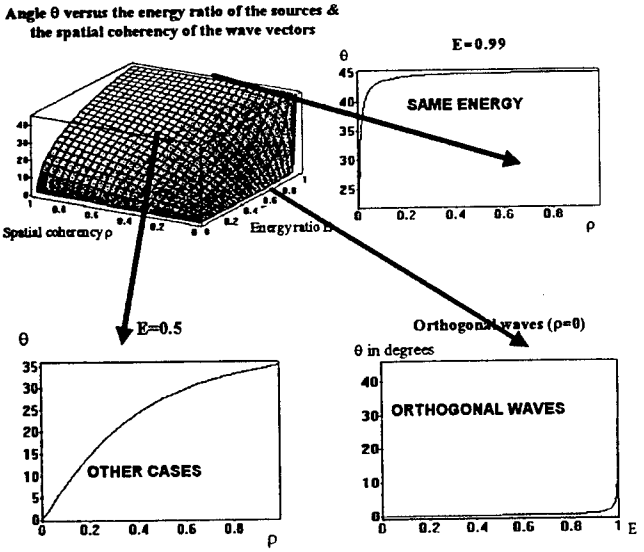


Figure 1 : Variations of the angle θ versus energy ratio and spatial coherency of the waves

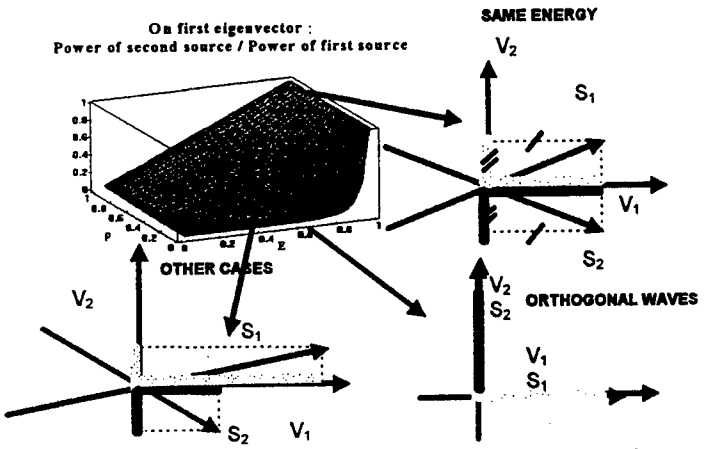


Figure 2 : Fluctuations versus spatial coherency and energy ratio

IV- Conclusion

In this work we explain how the basis of steering vectors and eigenvectors fit together and how this fitting depends on different parameters such as the energy ratio of waves and their spatial correlation degree.

V- References

[1] **Comon, P.**, 1989, "Separation of sources using higher order cumulants" : *SPIE Conf. On Advanced Algorithms and Architectures for Signal Processing, Real-Time Signal Processing*, vol. XII, pp. 170-181, San Diego, CA.

[2] **Durrani T. S., Bisset D.** 1984, "The radon transform and its properties" : *Geophysics*, n°49, pp. 1180-1187.

[3] **Esmersoy, C.**, 1990, "Inversion of P and SV waves from multicomponents offset vertical seismic profiles" : *Geophysics*, vol. 55, pp. 39-50.

[4] **Hardage, B. A.**, 1985 "Vertical seismic profiling, Part A : Principles" : *Handbook of geophysical exploration, Sect. I : Seismic exploration, K. Helbig and S. Treitel Editors, Geophysical Press*, London.

[5] **Hemon, C., Mace, D.**, 1978, "Essai d'application de la transformée de Karhunen-Loève au traitement sismique" : *Geophysical Prospecting*, vol. 26, pp. 600-626.

[6] **Lacoume, J. L., Ruiz, P.**, 1988, "Sources identification : a solution based on the cumulants" : *IEEE ASSP, Workshop on Spectrum Estimation*, Minneapolis, pp. 199-203.

[7] **Mars, J., Glangeaud, F., Lacoume, J. L., Fourmann, J. M., Spitz, S.** 1987, "Séparation of seismic waves" : *55th Annual SEG Meeting*, New Orleans, pp. 489-492.

[8] **Mermoz, H.**, 1969, "Elimination des brouilleurs par traitement optimal d'antenne" : *Annales des Télécommunications*, tome 24, n°7-8, pp. 282-293.

[9] **Robinson, E. A.**, 1967, "Predictive decomposition of time series with application to seismic exploration" : *Geophysics*, vol. 32, pp. 418-484.

[10] **Thirion, N., Mars, J., Lacoume, J. L.**, 1995, "Séparation aveugle de signaux large bande : un nouveau challenge en prospection sismique" , *GRETSI*, Juan-les-Pins.

[11] **Thirion, N.**, 1995, "Séparation d'ondes en prospection sismique" : *thèse de doctorat l'I.N.P.G.*

[12] **Wilkinson, J. H.**, 1965, "The algebraic eigenvalue problem" : *Oxford University Press.*

Multichannel Equalization Lower Bound: a function of Channel Noise and Disparity

I. Fijalkow
ENSEA / ETIS, 6 av. du Ponceau, 95014 Cergy-Pontoise Cdx, France
e-mail: fijalkow@ensea.fr, fax: (33-1) 30 73 66 27 *

Abstract

Recents studies have shown that provided spatial or temporal diversity, blind identification / equalization is perfectly achievable under some conditions on the channel transfer function and amount of data considered. However, in the presence of channel noise, equalization can no longer be achieved perfectly. We study the best achievable linear equalizer performances in terms of the input / output minimum mean square error (MMSE), defining the **channel equalizability** *as a function of the multichannel transfer function roots and the signal to noise ratio (SNR). We show that a* **channel disparity lower bound** *can be deduced as a function of the SNR in order to achieve a given amount of MMSE.*

Keywords: **Fractionally spaced / multichannel equalization, channel disparity.**

1. Introduction

Equalization is a crucial part of digital communication systems [1]. The way equalization is implemented is a trade-off between reaching high performances and computation cost. In particular, the equalizer length determines the computation need. However, it must be chosen carefully so to guaranty the performances required by the remaining parts of the system. The usual one-input / one-output channel equalization problem is known to require an equalizer length proportional to the inverted channel impulse response, the value of which is prohibitive for short FIR channels. One-input / multiple-outputs equalization induced by **channel diversity** was recently shown to be perfectly achievable within a finite length equalizer, ([2], [3], [5], [6]....) under some **Zero-Forcing** (ZF) conditions to be recalled later. First, we quantify equalization best achievable performances when the propagation is disturbed by additive channel noise. Performances are measured by the input / output minimum mean square error (MMSE). We investigate then the links between this lower bound and a measure of channel diversity. This should provide, for a given amount of channel noise, a measure of constraint on the channel so to equalize to a given amount of MMSE. It should help in evaluating the benefit of additional diversity with respect to the equalizer length to be used. Such a bound should also be very useful in order to compare algorithms and criterions performances.

2. Spatio-Temporal Equalization

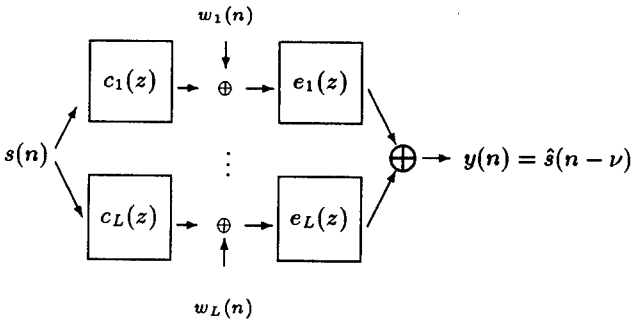


Figure 1. Noisy Fractionally-Spaced Equalization scheme

The one-input / multiple-outputs channel model ($\mathbf{c}(z) = (c_1(z), ..., c_L(z))^T$) is a well suited formalism for spatial diversity (i.e., a sensors array) as well as temporal diversity (i.e., sampling the received signal at an higher rate than the emitted sequence), see [4]. The induced equalization problem consists of choosing a L-variate equalizer transfer function $\mathbf{e}(z) = (e_1(z), ..., e_L(z))^T$ such that

$$y(n) = [\mathbf{e}(z)^T]\boldsymbol{r}(n) = \boldsymbol{e}^T\boldsymbol{R}(n) \qquad (1)$$

estimates at best $s(n-\nu)$, as in Figure 1. $\boldsymbol{R}(n)$ is the observation regression vector containing the N past

observations of $\boldsymbol{r}(n)$. ν is the channel-equalizer delay. Assuming $\mathbf{c}(z)$ is a polynomial vector of degree Q, equation (1) can be rewritten as

$$y(n) = \boldsymbol{e}^{\mathsf{T}}\mathbf{C}\boldsymbol{S}(n) + \boldsymbol{e}^{\mathsf{T}}\boldsymbol{W}(n)$$

where $\boldsymbol{S}(n) = (s(n), s(n-1), ..., s(n-Q-N+1))^{\mathsf{T}}$ contains the $N+Q$ past observations of $s(n)$. $\boldsymbol{W}(n)$ is defined alike $\boldsymbol{R}(n)$. $\mathbf{C}$ is a $NL \times (N+Q)$ (Sylvester) **channel convolution matrix** defined by the taps of the degree Q multivariate transfer function, $\mathbf{c}(z)$, as in [4].

ZF is defined here in terms of the channel transfer function invertability as:

$$\mathbf{e}^{\mathsf{T}}(z)\mathbf{c}(z) = z^{-\nu}$$

and is guaranteed under the following conditions expressed in terms of $\mathbf{C}$. Under the **ZF conditions** (i.e., $N-1 \geq Q$ and no common roots to all subchannels transfer functions), $\mathbf{C}$ is full-column rank so that any channel-equalizer global impulse response $h = \mathbf{C}^{\mathsf{T}}e$ is achievable, in particular the ZF h_ν corresponding to $z^{-\nu}$. So that in the absence of noise, these conditions allow perfect identification of the channel transfer function and perfect equalization, i.e., $y(n) = s(n-\nu)$.

However, as soon as there is a "non-negligible amount" of **channel noise**, perfect equalization can no longer be performed even if the channel impulse response is exactly identified. In particular, if the noise is filtered by a ZF equalizer, (the transfer function of which will be calculated as function of the channel transfer function poles and zeros locations), it may be enhanced so that the signal to noise ratio (SNR) at the equalizer output is reduced.

Example 1: the roots of $c_1(z)$ and $c_2(z)$ are respectively 0.59, -0.1 and, 0.6, 1.3, ZF may be achieved for $N = 2$ and the equalizers corresponding to the different possible delays have their norm displayed in the following table:

ν	0	1	2	3
$\|\|e\|\|$	250.0	250.0	326.9	25.8

So that the noise may be enhanced by a factor of more than 250 at the equalizer output, reducing all the more the output SNR and performances. Of course, some algorithms do not try to achieve ZF but a trade-off in terms of performances between ZF and noise enhancement, see [8] for instance.

In order to design robust algorithms, i.e., so to balance noise enhancement, we need to better understand what induces it. This goal motivates the following study of the channel roots locations effect on the equalization performances.

3. MMSE

When the channel is affected by additive white noise (independent from the source sequence), a measure of achievable direct linear equalization performance is given by the minimization of the input-output normalized MSE,

$$\mathrm{E}\left[(y(n) - s(n-\nu))^2\right] / \mathrm{E}\left[s^2\right] = \|h - h_\nu\|^2 + \gamma\|\mathrm{e}\|^2 \quad (2)$$

under the constraint $h = \mathbf{C}^{\mathsf{T}}e$, where $\gamma = \mathrm{E}\left[w^2\right] / \mathrm{E}\left[s^2\right]$ is the noise to signal ratio.

Under the ZF conditions, any value of the NL long vector h is achievable so that the minimization of (2) is proved ([8]) to correspond to $\mathrm{e}_\gamma = \mathbf{C}(\mathbf{C}^{\mathsf{T}}\mathbf{C})^{-1}h_\gamma$ with

$$h_\gamma = (I + \gamma(\mathbf{C}^{\mathsf{T}}\mathbf{C})^{-1})^{-1}h_\nu$$

which is all the more distinct from ZF than γ is high. The resulting MMSE is expanded in terms of γ as:

$$\gamma h_\nu^{\mathsf{T}}(\mathbf{C}^{\mathsf{T}}\mathbf{C})^{-1}h_\nu + \gamma^2 \sum_{m=0}^{\infty}(2m+1)(-\gamma)^m h_\nu^{\mathsf{T}}(\mathbf{C}^{\mathsf{T}}\mathbf{C})^{-(m+2)}h_\nu$$

Such an expansion is valid for small enough values of γ, precisely when $\|\gamma(\mathbf{C}^{\mathsf{T}}\mathbf{C})^{-1}\| \ll 1$.

Note that when a channel identification method is used, the maximum likelihood estimator of the input sequence in the presence of white gaussian noise induces an input / output MSE equal to $\gamma h_\nu^{\mathsf{T}}(\mathbf{C}^{\mathsf{T}}\mathbf{C})^{-1}h_\nu$ when the channel is perfectly estimated. Namely, it is equal to the minimal value of $\gamma\|e\|^2$ constrained to ZF, i.e., $h_\nu = \mathbf{C}^{\mathsf{T}}e$. It appears this expression of MSE is the first order approximation of the preceding MMSE full expression. It will be denoted as,

$$\mathrm{MMSE}(\gamma) = \gamma\|e\|^2 = \gamma h_\nu^{\mathsf{T}}(\mathbf{C}^{\mathsf{T}}\mathbf{C})^{-1}h_\nu \quad (3)$$

The optimal delay ν is therefore chosen so to minimize $h_\nu^{\mathsf{T}}(\mathbf{C}^{\mathsf{T}}\mathbf{C})^{-1}h_\nu$. This value, which is bounded by the inverses of the extremal eigenvalues of $\mathbf{C}^{\mathsf{T}}\mathbf{C}$, is the one of interest here.

4. A Measure of Channel Disparity

Under the ZF conditions, the invertability of $\mathbf{C}^{\mathsf{T}}\mathbf{C}$ is given, for $L = 2$, by:

$$\det\left(\mathbf{C}^{\mathsf{T}}\mathbf{C}\right) = K\Pi_{i,j}|z_1^i - z_2^j|^2 \quad (4)$$

where Π stands for product, K is a polynomial function of $(\|c_1\|^2 + \|c_2\|^2)^2$ with $\|c_k\|^2 = \sum_{i=0,Q} c_k(i)^2$, and z_k^i is the root i of subchannel k, $k = 1, 2$. Note that it is quite difficult to fully express K, even using Sylvester resultant results ([9]) and symbolic calculus.

(4) allows to connect the ZF condition "no common roots" to the identification / equalization performances given by (3). In particular, (4) shows also how close subchannels roots can create important noise enhancement as in Example 1 where one root difference is 0.01. From there, we can define a measure of channel disparity by $\gamma/K\Pi_{i,j}|z_k^i - z_l^j|^2$ which is all the more important than zeros of different subchannels get close to each other. However, one should note that a small value for $\det(\mathbf{C}^\top\mathbf{C})$ does not necessarily imply a large value for $\mathrm{MMSE}(\gamma)$, see Example 1 for $\nu = 3$.

The question of interest here is:
When are two "numerically close" roots so close that there is "lack of disparity" ?

We propose to study the minimum distance between the two closest roots allowing "disparity". Thus, we define there is **channel lack of disparity** when the actual equalization MMSE is better approximated by the MMSE value obtained by considering the two roots as equal than by the value deduced from (3). Otherwise, i.e., when the actual MMSE is better approximated by (3), we say that the channel presents **spatio-temporal diversity**.

4.1. Non-achievable ZF

To be able to quantify a bound of disparity, we need to look at the extreme case when there are mathematically equal roots.

In a previous contribution [7], we have shown that when there are common roots to all subchannels (referred to as lack of disparity), $\mathbf{C}$ is no longer full column-rank but can be factored as a product of two Sylvester matrices as $\mathbf{C} = \underline{\mathbf{C}}\mathbf{C}_0$ where $\underline{\mathbf{C}}$ is full column-rank, and $\mathbf{C}_0$ full row-rank. $\mathbf{C}_0$ is the convolution matrix associated to $c_0(z)$ which is formed by the Z_0 common roots. $\underline{\mathbf{C}}$ is associated to the remaining multichannel transfer function, $\underline{\mathbf{c}}(z)$. In that case, ZF is no longer achievable and the closest achievable equalizer to the ZF h_ν is $h = \mathbf{C}_0^\top(\mathbf{C}_0\mathbf{C}_0^\top)^{-1}\mathbf{C}_0 h_\nu = \Pi_0 h_\nu$ which is the projection of h_ν on the range of the non-full column-rank $\mathbf{C}_0^\top$.

The MMSE must then to be calculated using the previous factorization of $\mathbf{C}$. In the contrary of the case of ZF conditions, h is no longer an unconstraint parameter since it has to lay in the range of $\mathbf{C}_0^\top$. The only unconstrained parameter to be used for the minimization is also $\underline{e}$ such as $h = \mathbf{C}_0^\top\underline{e}$ and $\underline{e} = \underline{\mathbf{C}}^\top e$. The optimal value of e for a given $\underline{e}$ is also $e(\underline{e}) = \underline{\mathbf{C}}(\underline{\mathbf{C}}^\top\underline{\mathbf{C}})^{-1}\underline{e}$. So that the MMSE is obtained by minimizing over $\underline{e}$:

$$\|\mathbf{C}_0^\top\underline{e} - h_\nu\|^2 + \gamma\|\underline{\mathbf{C}}(\underline{\mathbf{C}}^\top\underline{\mathbf{C}})^{-1}\underline{e}\|^2$$

We can thus deduce a first order approximation in terms γ of the MMSE, $\mathrm{MMSE}_0(\gamma) =$

$$\|(I - \Pi_0)h_\nu\|^2 + 3\gamma h_\nu^\top \mathbf{C}_0^{-L}(\underline{\mathbf{C}}^\top\underline{\mathbf{C}})^{-1}(\mathbf{C}_0^{-L})^\top h_\nu \quad (5)$$

with $\mathbf{C}_0^{-L} = \mathbf{C}_0^\top(\mathbf{C}_0\mathbf{C}_0^\top)^{-1}$.

To define the disparity bound, we want to consider the conditions where $\|\gamma(\underline{\mathbf{C}}^\top\underline{\mathbf{C}})^{-1}\| \ll 1$ and $\|(\underline{\mathbf{C}}^\top\underline{\mathbf{C}})^{-1}\| > 1$. Thus, $\mathrm{MMSE}(\gamma)$ in (3) must be compared to the 0 order approximation in terms of γ of $\mathrm{MMSE}_0(\gamma)$. Therefore, we can define the disparity bound as:

$$\gamma h_\nu^\top(\underline{\mathbf{C}}^\top\underline{\mathbf{C}})^{-1}h_\nu = h_\mu^\top(I - \Pi_0)h_\mu \quad (6)$$

where ν and μ are the value minimizing the expression in which they are involved.

Example 2: In order to look at the lack of diversity bound, let us first consider the simplest case of two subchannels given by $c_k(z) = 1 - \xi_k z^{-1}$, $k = 1, 2$. Denoting $\xi_2 = \xi_1 + \epsilon$, we want to check what happens as $|\epsilon|$ decreases towards 0. The "numerical border of disparity" occurs when the MMSE calculated for supposedly distinct roots in (3) becomes larger than the MMSE calculated for these roots taken as equal in (5). With $N = 2$ (which is a large enough equalizer length when the roots are distinct), a simple formal calculus leads to the two expressions,

(3): $\mathrm{MMSE}(\gamma) = \gamma f_\nu(\xi_1)/\epsilon^2 + o(\gamma/\epsilon^2)$,
(5): $\mathrm{MMSE}_0(\gamma) = g_\mu(\xi_1) + o(1)$.

where f_ν and g_μ are rational bounded functions, the expressions of which are omitted for sake of space and clarity.

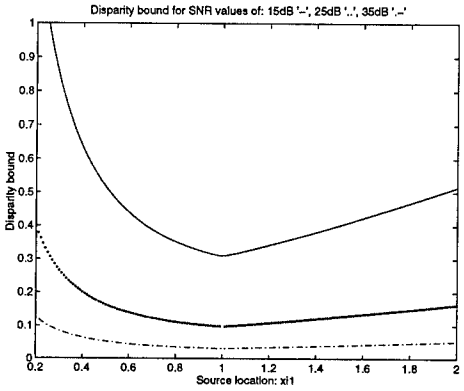


Figure 2. Disparity bound v.s root location and SNR

Next, we propose a disparity bound:

$$\epsilon^2 = \gamma f_\nu(\xi_1)/g_\mu(\xi_1)$$

depending on the signal to noise ratio, $\mathrm{SNR} = -10\log(\gamma)$ and the location of the root ξ_1. We display

in Figure 2 the disparity bound, ϵ, i.e., the distance one should ensure between two "close roots" to provide disparity, versus their location ξ and for several values of SNR.

Indeed, the greater the SNR is, the closer roots can be and still provide disparity. Note that the bound values can become so large that disparity is not possible. Furthermore, it appears the closer ξ_1 is to 1 (i.e., the unit circle when generalizing to complex numbers), the smaller the disparity bound is. It means that diversity is all the more important that the channel roots are close to the unit circle. This result is crucial since roots close to the unit circle is a very difficult condition for equalization when there is no diversity.

4.2. Additional Diversity

Another important question is whether additional channel diversity improves significantly the equalization performance.

We also want to extend the expression of $\det(\mathbf{C}^\top\mathbf{C})$ to the case of $L \geq 3$ to fully understand the condition "no comon roots to all subchannels". Generalized Sylvester resultant calculus (see [9] for classic results) lead to the fact that the rank of $\mathbf{C}^\top\mathbf{C}$ is equal to $N + Q - Z_0$, each common zero reducing the rank by its multiplicity. So that the determinant must be expressed by a weighted sum of products. Symbolic calculus and simple examples lead us to suggest the following expression:

$$\det(\mathbf{C}^\top\mathbf{C}) = \sum_{k<l} K_{k,l}\Pi_{i,j}|z_k^i - z_l^j|^2$$

where $K_{k,l}$ is some polynomial bounded function of the subchannels k and l. This measure shows the possible gain of disparity when one increases the disverity, i.e., when the number of subchannels is increased by either spatial or temporal diversity.

Example 3: To explicit this gain, let consider the two subchannels in Example 2 to which we add a third one, $c_3(z) = 1 - \xi_3 z^{-1}$. Thus,

$$\det(\mathbf{C}^\top\mathbf{C}) = K\left((\xi_1 - \xi_2)^2 + (\xi_2 - \xi_3)^2 + (\xi_3 - \xi_1)^2\right)$$

It appears clearly that the additional subchannel induces disparity. Note that the additional terms in the determinant result in a value greater than this for the "best" two subchannels combination.

5. Conclusion

We have proposed a channel disparity bound based on the channel noise power and multichannel transfer function. It explains how spatio-temporal diversity may induce enough disparity to allow equalization with a finite length equalizer. However, ZF equalization may have very poor performances when some subchannels are closer than the disparity bound depending on the SNR and on the other channel roots. The effect of close roots location versus the unit circle was also studied and results in diversity being all the more important that roots are close to the unit circle, which is the difficult case in monovariate channel equalization. Further study of the improvement of additional diversity is undertaken.

We hope these simple theoretical results will help in understanding the contribution of spatio-temporal diversity in more realistic channel conditions.

Acknowledgements: The author would like to thank J.R. Treichler for fruitful conversations, A. Touzni for his daily collaboration and D. Declercq for his help in symbolic calculus.

References

[1] J.R. Treichler, I. Fijalkow, C.R. Johnson Jr., *How long should an equalizer be ?*, submitted to IEEE DSP magazine, november 1995.

[2] L. Tong, G. Xu, and T. Kailath, *Fast blind equalization via antenna arrays*, in Proc. ICASSP'93, 1993.

[3] Z. Ding, Y. Li, *On channel identification based on second-order cyclic spectra*, IEEE Tr. on SP, May 1994.

[4] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, *Subspace methods for the blind identification of multichannel FIR filters*, IEEE Tr. on SP, February 1995.

[5] D.T.M. Slock, C.B. Papadias *Further results on blind identification and equalization of multiple FIR channels*, in Proc. ICASSP'95, 1995.

[6] G.B. Giannakis, S.D. Halford , *Blind fractionally-spaced equalization of noise FIR channels: Adaptive and optimal solutions*, in Proc. ICASSP'95, 1995.

[7] I. Fijalkow, J.R. Treichler, C.R. Johnson Jr., *Fractionally Spaced Blind Equalization: Loss of Channel Disparity*, in Proc. ICASSP'95, 1995.

[8] A. Touzni, I. Fijalkow, J.R. Treichler, *Robustness of fractionally-spaced equalization by CMA to lack of channel disparity and channel noise*, in Proc. SSAP'96. A full version has been submitted to the IEEE Tr. on SP special issue on advances in communications.

[9] T. Kailath, *Linear Systems*, Prentice-Hall, 1980.

Multiple Input Multiple Output ARMA systems: second order blind identification for signal extraction

A. Gorokhov[†]
Télécom Paris, Dept. Signal
46 rue Barrault
75634 Paris Cedex 13 FRANCE

P. Loubaton
UF SPI (EEA) Université de Marne la Vallée
2 rue de la Butte Verte
93166 Noisy-le-Grand Cedex FRANCE

Abstract

This paper addresses the blind identification of multiple input multiple output (MIMO) systems with the number of inputs strictly less then the number of outputs. On the contrast to the standard FIR modelling we assume the overall channel with arbitrary finite order rational transfer function. Certain quite reasonable technical hypotheses allow to adapt the existing linear prediction and subspace based approach and implement a finite order zero-forcing equalizer in the noise-free case. The noise-free condition also yields a simple performance analysis which is quite accurate at low noise levels and provides a meaningful comparison of the proposed estimators. The robustness to additive noise is studied here by computer simulations for both techniques.

Keywords: blind identification, equalization, performance analysis.

1. Introduction

Convolutive properties of the propagation media is a typical shortcoming in various applications. In digital communications it leads to severe inter symbol interference (ISI) dramatically reducing the channel capacity. Many recent publications consider the problem of blind dentification *i.e.* channel evaluation analyzing the output observation and further extraction of the input signals. Classical approaches to single input single output (SISO) identification [1, 2] usually exploit higher than two order statistics (HOS) and are demanding in sample volume. A noticeable improvement has been achieved due to either multiple antennas or the observation oversampling, see [3, 4]. Both of them allow to recast the problem into the single input multiple output (SIMO) identification where the second order estimation as well as finite length zero-forcing equalizer (ZFE) are available. A certain part of these results has been recently generalized for the MIMO case, [5, 6]. Most of them still treat the finite order polynomial transfer functions originating from multipath propagation environment. However the precise analysis of some secondary phenomena *e.g.* mutual coupling of sensors/receivers could enable certain improvement through a more sophisticated channel modelling especially at low noise levels.

We consider here rational transfer functions usually satisfying most of the applications. As a matter of fact a matrix-valued transfer function admits many different parametrizations (*i.e.* corresponding to various *canonical forms*, see [7]). One of the possible solutions is based on the AR factorization of the actual ARMA model obtained as a generalization of the similar results for the multivariate FIR case [6]. An alternative approach originates from the right matrix-fraction description (MFD) which implies a two step identification procedure with the MIMO subspace based estimator [5] followed by the linear prediction of a reduced size. Both methods yield finite order causal ZFE providing the instantaneous mixture of source signals. Such a mixture can be conventionally treated by any of the existing source separation techniques. We focus here on the deconvolution performance *i.e.* the residual ISI at the output of ZFE. Further asymptotic analysis given in section 5 allows to establish the residue ISI variances which appear to be invariant to a particular channel realization. We further compare the above mentioned estimation approaches and simulate their behaviour in the noisy case.

2. Data model and hypotheses

Let us consider M-variate time series $\{x(t)\}_{t\in\mathbb{Z}}$ being the output of some $M \times m$ linear system with rational transfer function $H(z) \triangleq \sum_{\tau=0}^{\infty} \mathbf{H}(\tau) z^{-\tau}$ and m-variate input series $\{s(t)\}_{t\in\mathbb{Z}}$. The sequences $\{x(t)\}_{t\in\mathbb{Z}}$ and $\{s(t)\}_{t\in\mathbb{Z}}$ addressed in the sequel as *observation* and *excitation* satisfy the linear equation written in the operator form:

$$x(t) = [H(z)]\, s(t), \qquad t \in \mathbb{Z}. \tag{1}$$

Here each entry $[H(z)]_{pq}$ may be interpreted as a transfer function between input q and output p. In the special case of constant transfer functions $H(z) \equiv H$

[†]This study is supported by CNET (France Télécom), ENST and partially by the SASPARC project of INTAS.

we have the instantaneous mixture separation problem. On the other hand, one deals with multichannel equalization if $m = 1$. In the most general case of linear processing, we look for a MIMO ZFE $E_H(z) \triangleq \sum_{\tau=0}^{\infty} \mathbf{E}_H(\tau)\, z^{-\tau}$ associated with channel $H(z)$ so that $E_H(z)\, x(t) = s(t)$ *i.e.* $E_H(z)\, H(z) = \mathrm{I}_m$. Moreover it is preferable to have a finite order polynomial ZFE for practical applications. Any consistent channel estimate $\hat{H}(z)$ calculated from the finite observation $\{x(t)\}_{t=1}^{T}$ yields a sample equalizer $\hat{E}_H(z)$ and consistent extraction of the input signals

$$\hat{s}(t) = [\hat{E}_H(z)]\, x(t), \qquad t \in \mathbb{Z}. \tag{2}$$

We study the conditions providing the existence of $E_H(z)$ and the extraction accuracy subject to various channel estimators. Some further results are available under the following hypotheses:

H1 The number of inputs m is strictly less than the number of outputs M.

H2 The emitted sequences $\{s_k(t)\}_{t\in\mathbb{Z}}$, $k = 1, \ldots, m$, are statistically independent and temporally non-correlated non-gaussian series.

H3 The rational matrix $H(z)$ is irreducible, see [7].

More involved analysis of **(H3)** provides quite clear interpretation in terms of inter-channel diversity and shows that this constraint is met in typical applications, [8]. Meanwhile this assumption plays a key role for channel identification and signals extraction since it ensures the existence of a finite order ZFE.

Lemma 1 *Let $H(z)$ be finite order rational function satisfying* **(H3)**. *Then there exists finite N_E and the associated $E_H(z)$ with* $\deg(\, E_H(z)\,) = N_E$ *such that* $E_H(z)\, H(z) = \mathrm{I}_m$.

One can presume that ZFE is not uniquely defined under **(H1)** since there exists a non-trivial left null-space of $H(z)$. Our reader will see how the choice of ZFE can be adapted to the particular factorization of $H(z)$ used at the preceeding channel evaluation stage.

3. Linear prediction approach

This kind of technique has been recently proposed in mono-source context [9] and later developed for the MIMO case of FIR channels, [6]. We propose a staightforward extension based on the following property.

Lemma 2 *Let $H(z)$ be a finite order rational function satisfying* **(H3)**. *Then there exists finite N_P and an associated $P(z) = \mathrm{I}_M + \sum_{\tau=1}^{N_P} \mathbf{P}(\tau)\, z^{-\tau}$ verifying* $P(z)\, H(z) = \mathbf{H}(0)$.

In the other words, any irreducible ARMA channel is also an AR channel of finite order. Consequently the equation (1) can be written as follows:

$$[P(z)]\, x(t) = \mathbf{H}(0)\, s(t), \qquad t \in \mathbb{Z}. \tag{3}$$

The prediction coefficients $\mathbf{P} = [\mathbf{P}(1), \ldots, \mathbf{P}(N_P)]$ and the innovation covariance matrix $\mathbf{D} = \mathbf{H}(0)\, \mathbf{H}(0)^H$ may be consistently estimated by solving a multivariate Yule-Walker equation with the empirical counterpart of the block-Toeplitz spatio-temporal covariance matrix:

$$\mathbf{R}_x = \{\mathbf{R}_x(\tau)\}_{\tau=0}^{N_P},\ \mathbf{R}_x(\tau) = \mathbb{E}\left\{x(t)\, x(t-\tau)^H\right\}. \tag{4}$$

Some more details on calculating $\hat{\mathbf{P}}$ and $\hat{\mathbf{D}}$ the consistent estimates of $\mathbf{P}$ and $\mathbf{D}$ can be found in [10]. Notice also that a similar estimation procedure is valid when the observation $\{x(t)\}_{t\in\mathbb{Z}}$ is corrupted by an additive temporally white noise with known spatial structure since a consistent estimate of $\mathbf{R}_x$ is still available.

Now any $M \times m$ square root $\mathbf{F}(0)$ of $\mathbf{D}$ such that $\mathbf{F}(0)\, \mathbf{F}(0)^H = \mathbf{D}$ verifies $\mathbf{H}(0) = \mathbf{F}(0)\, \Theta$ with some unitary $m \times m$ matrix Θ. Let us consider a FIR filter $E_F(z) \triangleq \mathbf{F}(0)^{\#} P(z)$. According to (4), we have $[E_F(z)]\, x(t) = \Theta\, s(t)$ *i.e.* $E_F(z)$ is a kind of ZFE providing the instantaneous mixture of source signals. Due to **(H2)**, further extraction of each source signal $s_k(t)$ can be completed by HOS source separation techniques. Let us denote by $\hat{Q}$ some consistent estimate of the $m \times m$ separator obviously verifying $\lim_{T\to\infty} \hat{Q} = \Theta^H$. We finally apply a finite order N_P FIR filter

$$\hat{E}_H(z) \triangleq \hat{Q}\, \hat{\mathbf{F}}(0)^{\#} \hat{P}(z) \tag{5}$$

to the observation series as indicated in (2). Due to the strict consistency of each empirical quantity, $\lim_{T\to\infty} \hat{s}(t) = s(t)$.

4. Generalized subspace approach

This method stamps from the canonical right MFD of rational functions. Let us denote by $\mathcal{S}(z)$ the column space of $H(z)$ *i.e.* $\mathcal{S}(z) = \mathrm{span}\{\, H(z)\,\}$. In the most general case one can deduce the following result, see [11] for definitions.

Lemma 3 *Let* **(H3)** *hold and let polynomial $M \times m$ matrix $B(z)$ be any minimal polynomial basis (MPB) of $\mathcal{S}(z)$ with invariant column degrees* $\deg(\, [B(z)]_k\,) = L_k$, $L_1 \leq, \ldots, \leq L_m$. *Than there exists $m \times m$ polynomial $C(z)$ of finite degree N_C and full rank almost everywhere in $\mathbb{C}$ such that* $H(z) = B(z)\, C(z)^{-1}$.

According to lemma 3, the identification procedure can be accomplished now in two steps: **(i)** identify any MPB of $\mathcal{S}(z)$; **(ii)** identify the associated $C(z)$. Let us focus on the first stage. We denote by $\{\tilde{B}(z), \tilde{C}(z)\}$ any arbitrary pair satisfying lemma 3 *i.e.* $H(z) = \tilde{B}(z)\, \tilde{C}(z)^{-1}$. Now the observation series can be rewritten as $x(t) = [\tilde{B}(z)]\, v(t)$, where $v(t) = [\tilde{C}(z)^{-1}]\, s(t)$. Notice that $\tilde{B}(z)$ is some MPB of $\mathcal{S}(z)$ and $v(t)$ has a full-rank covariance matrix $\mathbf{R}_v$ of any order (see definition (4)). As indicated in [5], one can perfectly identify some MPB of $\mathcal{S}(z)$ from a finite observation sample. Such a MPB may be consistently

estimated in the noisy case. For more details concerning the estimation of a particular MPB we address the reader to [12].

Let us denote by $\hat{B}(z)$ a consistent estimate of some MPB $B(z)$. As shown in [11] any MPB matches **(H3)** and according to lemma 1 there exists a finite order $E_B(z)$ verifying $E_B(z)\,B(z) = \mathrm{I}_m$. In practice such a left inverse of $B(z)$ can be calculated from the algebraic analog of (1): $[x(t)^T, \ldots, x(t-N_B)^T]^T = \mathcal{T}_{N_B}(B)\,[v(t)^T, \ldots, v(t-N_B-L_m)^T]^T$, where $\mathcal{T}_{N_B}(B)$ is a *generalized sylvester matrix* of order N_B associated with the polynomial $B(z)$, see [7]. The input signal $s(t)$ can be extracted by applying $m \times M(N_B+1)$ matrix $\mathbf{E}_B = [\mathbf{E}_B(0), \ldots, \mathbf{E}_B(N_B)]$ to $[x(t)^T, \ldots, x(t-N_B)^T]^T$ when $N_B \geq \sum_{k=1}^{m} L_k$. In this case $\mathbf{E}_B$ is calculated from the left pseudo-inverse of $\mathcal{T}_{N_B}(B)$ by taking its m upper rows. Its consistent estimate $\hat{\mathbf{E}}_B$ can be readily obtained using the empirical quantity $\hat{B}(z)$ instead of the true one. Further pre-filtering provides the intermediate output signal

$$v(t) = [\hat{E}_B(z)]\,x(t), \quad \lim_{T\to\infty} \hat{v}(t) = [C(z)^{-1}]\,s(t). \quad (6)$$

The described preliminary processing forms the kernel of the generalized subspace approach since it allows to reduce the initial problem to the identification of $m \times m$ polynomial matrix $C(z)$. One should notice that due to perfect identification of MPB $B(z)$ in the noise-free case the intermediate output $v(t) = [E_B(z)]\,x(t)$ safisfies $v(t) = [C(z)^{-1}]\,s(t)$ *i.e.* the generalized subspace estimator is *statistically equivalent* to the estimation of $C(z)$ from the series $\{v(t)\}_{t\in\mathbb{Z}}$.

Let us consider the identification of $C(z)$. Due to **(H3)**, its zero coefficient $\mathbf{C}(0) = \lim_{z\to\infty} C(z)$ is nonsingular and we can define matrix $A(z) = \mathrm{I}_M + \sum_{\tau=1}^{N_C} \mathbf{A}(\tau)\,z^{-\tau}$ such that $A(z) \triangleq \mathbf{C}(0)^{-1}C(z)$. It is easy to check that $v(t)$ and $s(t)$ verify the following equation

$$[A(z)]\,v(t) = \mathbf{C}(0)^{-1}s(t) \quad (7)$$

i.e. $\{v(t)\}_{t\in\mathbb{Z}}$ is the AR process with the prediction coefficients $\mathbf{A} = [\mathbf{A}(1), \ldots, \mathbf{A}(N_C)]$ and the the innovation covariance matrix $\mathbf{D}' = \mathbf{C}(0)^{-2}$. Obviously further identification of model (7) may be accomplished by means of the linear prediction approach described in section 3. Now $\hat{v}(t)$ is treated as the observation. Let us denote $\hat{\mathbf{A}}$ the esimate of $\mathbf{A}$, $\hat{\mathbf{F}}'(0)$ the empirical square root square root of $\mathbf{D}'$ and $\hat{Q}'$ some $m \times m$ separator estimate. Similarily to (5) we obtain $\hat{s}(t) = \hat{Q}'\,\hat{\mathbf{F}}'(0)^{-1}[\hat{A}(z)]\,\hat{v}(t)$. Finally

$$\hat{E}_H(z) \triangleq \hat{Q}'\,\hat{\mathbf{F}}'(0)^{-1}\hat{A}(z)\,\hat{E}_B(z), \quad (8)$$

the complete ZFE can be found by plugging $\hat{v}(t)$ from (6). This filter provides the consistent extraction of source signals from the observation according to (2) *i.e.* $\lim_{T\to\infty} \hat{s}(t) = s(t)$, similarily to (5).

5. Performance analysis

In this section we compare statistical efficiency of both identification techniques in the noise-free case. As it follows from section 4, the generalized subspace estimator allows perfect identification of factor $B(z)$ *i.e.* the *only* error is caused by the estimation of $C(z)$ *e.g.* linear prediction. Therefore we just need to compare: **(i)** extraction (2) via linear prediction according to (5); **(ii)** extraction of source signals from $v(t)$ via linear prediction according to the model (7). For this purpose we use some general asymptotic results concerning signals extraction via linear prediction in the noise-free case. Let $\{y(t)\}_{t\in\mathbb{Z}}$ be a M-variate time series satisfying the AR equation $[\mathcal{P}(z)]\,y(t) = \mathcal{H}s(t)$ with $\mathcal{P}(z)$ any prediction filter of order not more that $N_{\mathcal{P}}$ and some $M \times m$ full rank matrix $\mathcal{H}$. We assume further identification procedure described in section 3 and the associated estimate $\hat{E}(z) \triangleq \sum_{\tau=0}^{\infty} \hat{\mathbf{E}}(\tau)\,z^{-\tau}$ of ZFE defined according to (5) with $\hat{Q}$ providing the consistent signals estimate $\hat{s}(t) = [\hat{E}(z)]\,y(t)$. As a matter of fact consistent extraction yields that the global causal transfer function $\hat{\Gamma}(z) = \hat{E}(z)\,\mathcal{P}(z)^{-1}\mathcal{H}$, such that $\hat{s}(t) = [\hat{\Gamma}(z)]\,s(t)$, verifies $\lim_{T\to\infty} \hat{\Gamma}(z) = \mathrm{I}_m$ *i.e.* its Fourier coefficients match $\lim_{T\to\infty} \hat{\mathbf{\Gamma}}(\tau) = 0$ for $\tau > 0$. The *equalization errors* defined as the residue convolutive contribution to $s(t)$: ${}_{\triangle}\hat{s}(t) \triangleq \sum_{\tau=1}^{\infty} \hat{\mathbf{\Gamma}}(\tau)\,s(t-\tau)$ will be considered throughtout this paper as a performance index for the deconvolution techniques taking into account that the residual *separation error* essentially depends upon the source separation technique.

Theorem 1 *The asymptotic equalization errors* ${}_{\triangle}\hat{s}(t)$ *verify* $\lim_{T\to\infty} T\ \mathbb{E}\left\{{}_{\triangle}\hat{s}(t)\,{}_{\triangle}\hat{s}(t)^H\right\} = r\,\mathrm{I}_m$, *where* r *is the rank of* $\mathbf{R}_y$ *of order* $N_{\mathcal{P}} - 1$.

Notice that the equalization errors variances later addressed as the *equalization rates* are asymptotically invariant to the instantaneous separation performance as well as to the system parameters, they depend only upon the rank of $\mathbf{R}_y$. We further denote $r = r(\mathbf{R}_y, N)$ the rank of $\mathbf{R}_y$ having the order $N-1$, $N > 0$. Let $N_{\mathcal{P}min}$ be the minimum prediction order. Then for any $N_{\mathcal{P}} > N_{\mathcal{P}min}$ we have $r(\mathbf{R}_y, N_{\mathcal{P}}) > r(\mathbf{R}_y, N_{\mathcal{P}min})$ *i.e. order overestimation* always leads to performance degradation. To compare the potential efficiency of both methods one certainly needs to know minimum orders N_P and N_C, these latter non-trivially depending upon $H(z)$. More involved study of this quantities for typical channel realizations is being currently investigated. On the other hand, in the case of unknown N_P and N_C one can estimate channel via both of the designed methods if the observation window $N_{\mathcal{P}}$ is chosen sufficiently large. More precisely, we assume that the window of linear prediction approach and subspace method are chosen equal to $N_{\mathcal{P}}$. The corresponding number of the observable input samples including each source signal, equal to $N_{\mathcal{P}} + L_1$ (see section 4), provides us with the window for the secondary linear prediction *e.g.* identification of $C(z)$. It

is easy to show that $r(\mathbf{R}_x, N_\mathcal{P}) = mN_\mathcal{P} + \sum_{k=1}^{m} L_k$, $r(\mathbf{R}_v, N_\mathcal{P} + L_1) = m(N_\mathcal{P} + L_1)$ *i.e.* the performance of the generalized subspace technique is *not worse* than the performance of the pure linear prediction. The equality holds only if $L_1 = \ldots = L_m$, *i.e.* when $m = 1$.

6. Simulations

We present in this section some numerical examples validating theoretic results on both estimators in the noise-free case and their robustness to the additive noise. The overall propagation channel has been modelled as a left MFD form: $H(z) = \mathcal{A}(z)^{-1}\mathcal{B}(z)$ so that $[\mathcal{A}(z)]\, x(t) = [\mathcal{B}(z)]\, s(t)$. In digital multisensor communication, $\mathcal{B}(z)$ reflects the FIR propagation media between the user and the reception cite while $\mathcal{A}(z)$ might correspond to the mutual coupling of receivers. We compare linear prediction and generalized subspace approaches completed by joint diagonalization source separation procedure, see [13]. For system dimensions $M = 4$, $m = 2$ and observation sample size $T = 500$, we plot the residue equalization rates versus the degree of denominator $\mathcal{A}(z)$ for $\deg(\mathcal{B}(z)) = 2$, the degree of the first column of numerator $\deg([\mathcal{B}(z)]_1)$, $\deg([\mathcal{B}(z)]_2) = 0$, $\deg(\mathcal{A}(z)) = 2$ and versus the average signal-to-noise ratio, $\deg(\mathcal{A}(z)) = \deg(\mathcal{B}(z)) = 2$. Each simulated value is equiped with the confidence interval of ± 2 empirical standard deviation and the true value.

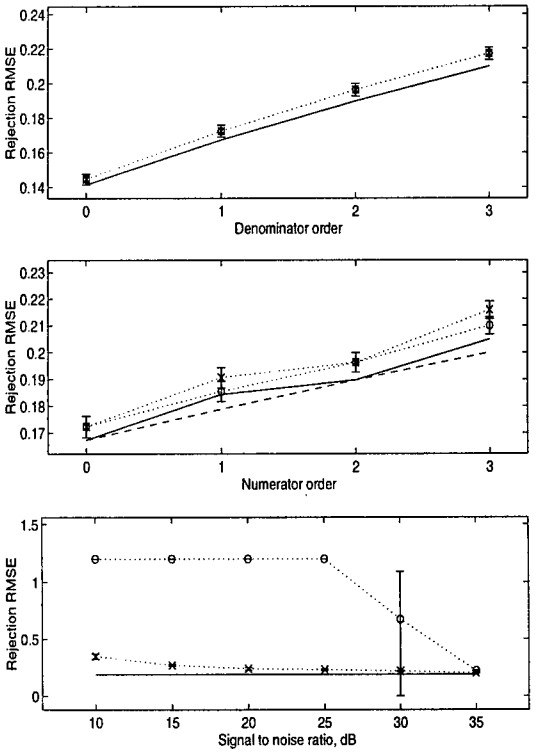


Fig.1. Equalization rates:
Linear prediction: — true, ..x.. numerical;
Generalized subspace: - - true, ..o.. numerical.

As it follows from Fig.1, the actual performance of two methods might be slightly different in favour of the subspace technique. On the other hand, this latter displays fatal degradation in the presence of noise, the values from $10dB$ to $25dB$ lead to the abnormal error in many cases even when the degrees L_k are *perfectly known*, identification of these quantities being a particular problem in the noisy case. Meanwhile the linear prediction approach appears to be robust at high and moderate signal-to-noise ratio levels.

References

[1] O. Shalvi, E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems", *IEEE Trans. on IT*, vol. 36, no. 2, 1990.

[2] C. Nikias, J. Mendel, "Signal processing with higher order spectra", *IEEE SP Magazine*, pp. 10-37, July 1993.

[3] W. Gardner, "A new method of channel identification", *IEEE Trans. on Comm.*, vol. 39, pp. 813–817, Aug. 1991.

[4] L. Tong, G. Xu, T. Kailath, "A new approach to blind identification and equalization of multipath channels", in *Proc. of the 25th Asilomar Conference, Pacific Grove, CA*, pp. 856–860, 1991.

[5] K. Abed Meraim, P. Loubaton, E. Moulines, "A subspace algorithm for certain blind identification problems", To appear in *IEEE IT*, 1996.

[6] D.T. Slock, C.B. Papadias, "Further results on blind identification and equalization of multiple FIR channels", *Proc. ICASSP*, pp. 1964–1967, 1995.

[7] T. Kailath, *Linear Systems*, Prentice-Hall, Inc., 1980.

[8] A. Gorokhov, P. Loubaton, E. Moulines, "Second order blind equalization in Multiple Input Multiple Output FIR systems. Part I: Identifiability and Estimation", Submitted to *IEEE SP*, 1995.

[9] K. Abed Meraim, P. Duhamel, D. Gesbert, P. Loubaton, S. Mayrargue, E. Moulines, D. Slock, "Prediction error methods for time-domain blind identification of multichannel FIR filters", *Proc. ICASSP*, pp. 1968–1971, 1995.

[10] K. Abed Meraim, E. Moulines, P. Loubaton, "Prediction error method for second order blind identification", To appear in *IEEE SP*, 1996.

[11] G. Forney, "Minimal bases of rational vector spaces, with applications to multivariable linear systems," *SIAM J. on Control*, vol. 13, pp. 493–520, May 1975.

[12] A. Gorokhov, P. Loubaton, "Subspace based techniques for second order blind separation of convolutive mixtures with temporally correlated sources", Submitted to *IEEE Trans. on Circuits and Systems*, 1995.

[13] J.-F. Cardoso, A. Souloumiac, "Blind beamforming for non Gaussian signals", *in Proc. IEE*, no. 6, vol. 140, Dec. 1993.

ON WEIGHTED SUBSPACE ESTIMATES IN SYSTEM IDENTIFICATION

J.-F. Cardoso, P. Loubaton, E. Moulines (cardoso,moulines@sig.enst.fr)
Ecole Nationale Supérieure des Télécommunications, Dept SIG & CNRS. 46 rue Barrault. 75634 Paris. France.

Abstract

Subspace based estimates, i.e. estimates obtained by exploiting the orthogonality between a set of vector statistics and a set of parameter-dependent vectors have gained much popularity in the signal processing litterature. The purpose of this contribution is to develop a general theory for such estimates. We in particular discuss the generalization of the optimal weighted subspace fitting approach, introduced by Viberg [2] in the DOA estimation context. We then establish that the optimally weighted estimate enjoy some invariance properties

1. SUBSPACE FITTING ESTIMATION

This section is concerned with general properties of subspace fitting estimates: we define a general framework and give the asymptotic performance of (optimal) subspace estimates.

1.1. Assumptions and notations

The common strand behind subspace estimation is to exploit the geometrical prope rty of a certain matrix-valued statistics for estimating unknown parameters. May be the most well-known example of such techniques is the so-called Pisarenko's method, which makes use of the eigen-subspace of a certain covariance matrix to estimate the frequency of the sine-waves in white noise. These methods have gained much popularity in the signal processing community in the last decade, and have been applied successfully to a variety of problems, such as the estimation of direction-of-arrivals in narrow-band array processing [1, 2], or more recently in system identification [3, 4, 5]. As seen below, this theory can be formulated in fairly general terms.

Consider a parametric statistical model where the distribution of n observations $y_1, \ldots, y_n$ depends on a parameter vector $\omega = (\theta, \mu) \in \Omega \subset \mathbb{R}^k$, $\theta \in \Theta \subset \mathbb{R}^l$, $\mu \in \mathbb{R}^{k-l}$, where Θ is a compact subset of $\mathbb{R}^l$. Here, θ is the *parameter of interest* and μ is a *nuisance parameter* (the values of μ are needed to make inferences about θ even through they have little informative import of their own).

A matrix-valued statistic $\hat{N}_n \in \mathbb{R}^{r\times q}$ is computed from $y_1, \ldots, y_n$. This statistic forms the basis for inferring the parameter of interest θ. Actually, we make no assumption on the distribution of the data themselves, but only on the asymptotic distribution of $\hat{N}_n$:

Assumption 1 (Asymptotic normality) *For all $\omega \in \Omega$, $\hat{N}_n$ is asymptotically normal with asymptotic mean $N(\omega)$ and asymptotic covariance matrix $\mathbf{\Sigma}_N(\omega)$.*

This assumption uses the following convention: A sequence Y_n of random $r \times q$ matrices is said to be *asymptotically normal* with asymptotic mean $\mathbf{Y}$ and asymptotic covariance matrix $\mathbf{C}$ if the $rq \times 1$ random vector $\sqrt{n}(\mathrm{Vec}(Y_n) - \mathrm{Vec}(\mathbf{Y}))$ tends in distribution to a zero-mean random vector with correlation matrix $\mathbf{C}$. We write: $Y_n \sim \mathcal{AN}(\mathbf{Y}, \mathbf{C})$.

Subspace fitting estimation is relevant when $\hat{N}_n$ converges to a rank deficient matrix $N(\omega)$ and when it exists a matrix-valued $S(\theta) \in \mathbb{R}^{r\times p}$, depending only on the parameter of interest and satisfying the following assumption.

Assumption 2 (Identifiability) *For any $\omega = (\theta, \mu) \in \Omega$,*

$$S^T(\theta')N(\omega) = 0 \;\Rightarrow\; \theta' = \theta. \tag{1}$$

Hence the basic mechanism of subspace fitting which consists in obtaining an estimate $\hat{\theta}$ of θ such that the columns $S(\hat{\theta})$ are 'as orthogonal as possible' to the columns of $\hat{N}_n$ as detailed in next section. Note that we do *not* require that $\mathrm{Span}(S(\theta, \mu))$ and $\mathrm{Span}(N(\theta))$ are orthogonal complements.

The following notational conventions hold throughout. First, bold face letters will denote values of functions of θ taken at the 'true value' of the parameters. In particular, we denote

$$\mathbf{S} \stackrel{\text{def}}{=} S(\theta), \quad \mathbf{N} \stackrel{\text{def}}{=} N(\theta, \mu). \tag{2}$$

It is often needed to collect derivatives of matrix valued functions w.r.t. θ into a unique larger matrix. A suggestive notation is needed for this construction. We will typically denote:

$$[\dot{\mathbf{S}}] \stackrel{\text{def}}{=} [\mathrm{Vec}(\frac{\partial S}{\partial \theta_1}), \ldots, \mathrm{Vec}(\frac{\partial S}{\partial \theta_l})], \tag{3}$$

$$[\dot{\mathbf{S}}^T\mathbf{N}] \stackrel{\text{def}}{=} [\mathrm{Vec}(\frac{\partial S}{\partial \theta_1}^T \mathbf{N}), \ldots, \mathrm{Vec}(\frac{\partial S}{\partial \theta_l}^T \mathbf{N})]. \tag{4}$$

where all the quantities at evaluated at point θ or (θ, μ). Since matrix $S(\theta)$ has size $r \times p$, matrix $[\dot{\mathbf{S}}]$ has size $rp \times l$. Finally, with several asymptotically normal matrix sequences appearing in the following, it will be convenient to note $\mathbf{C} = \mathrm{Cov}(Y)$ whenever $Y_n \sim \mathcal{AN}(\mathbf{Y}, \mathbf{C})$ under the distribution $\omega = (\theta, \mu)$.

This study being restricted to regular (root-convergence) estimation, we impose some regularity to functions $S(\cdot)$ and $N(\cdot)$ and also want to exclude cases where some linear combination of the parameters can be estimated at a super-efficient rate.

Assumption 3 (Regularity) *Functions $S(\cdot)$ and $N(\cdot)$ are differentiable with respect to θ at point $\omega = (\theta, \mu)$ and*

$$\mathrm{Span}([\dot{\mathbf{S}}^T\mathbf{N}]) \subset \mathrm{Span}(\mathrm{Cov}(\mathbf{S}^T\hat{N}_n)). \tag{5}$$

0-8186-7576-4/96 $5.00

Loosely speaking, this regularity assumption means that there is no direction in the parameter space which is not excited by the matrix errors $\mathbf{S}^T\hat{N}_n$.

Definition 1 *A pair* $(S(\cdot), \hat{N}_n)$ *is said to be* admissible for subspace fitting estimation *if it satisfies assumptions 1-3.*

Before proceeding, we stress that it is *not* assumed that $\text{Span}(\mathbf{N}) \oplus \text{Span}(\mathbf{S}) = \mathbb{R}^r$ neither that $\mathbf{N}$ or $\mathbf{S}$ have full column rank. If it holds that $\text{Span}(\mathbf{N}) \oplus \text{Span}(\mathbf{S}) = \mathbb{R}^r$, we say that a 'saturation condition' is fulfilled. In this case, additional properties of subspace estimates can be obtained (see section 2.4).

1.2. Subspace fitting estimates

Subspace fitting estimates are obtained as the minimizers of a criterion quantifying the orthogonality between the range space of $\hat{N}_n$ and the range space of $S(\eta)$:

$$\hat{\theta}_n^W = \arg\min_{\eta\in\Theta} \|S^T(\eta)\hat{N}_n\|_W^2 \tag{6}$$

where W is a (possibly rank deficient) symmetric non-negative matrix and $\|\cdot\|_W$ is the Weighted Euclidean norm, $\|M\|_W^2 = \text{Vec}(M)^T W \text{Vec}(M)$. For a given admissible pair $(S(\cdot), \hat{N}_n)$, weighting matrix W must be chosen to preserve the identifiability of assumption 2 : A matrix W is said to be *admissible* for the value (θ, μ) of the parameters if it preserves identifiability *i.e.* if $\|S^T(\theta')N(\theta,\mu)\|_W^2 = 0 \Rightarrow \theta' = \theta$ With this definition, we can state the following theorem.

Theorem 1 *If* $(S(\cdot), \hat{N}_n)$ *is an admissible pair and matrix* W *is admissible at point* (θ, μ) *for this pair, then* $\hat{\theta}_n^W$ *defined by* (6) is a consistent estimate of θ.

Of course, any fixed positive definite matrix W is admissible for any value of θ, the most straightforward choice being the identity matrix. As shown below, in the context of interest, 'optimal' weighting matrices are often rank deficient (in order to null out 'spurious' error terms) and the null space of the weighting matrix will generally depend on θ. This fact may appear problematic since the value of θ is unknown. However, one may show that a consistent estimate of the optimal weighting matrix can be used without affecting the asymptotic performance.

Weighted subspace estimates are asymptotically characterized as follows.

Theorem 2 *If* $(S(\cdot), \hat{N}_n)$ *is an admissible pair, then for any admissible weighting matrix* W*, the sequence* $\{\hat{\theta}_n^W\}$ *of estimates admits the stochastic expansion:* $\hat{\theta}_n^W = v_n^W + o_P(n^{-1/2})$ *with*

$$v_n^W = ([\dot{\mathbf{S}}^T\mathbf{N}]^T W [\dot{\mathbf{S}}^T\mathbf{N}])^{-1}[\dot{\mathbf{S}}^T\mathbf{N}]^T W \text{Vec}(\mathbf{S}^T\hat{N}_n). \tag{7}$$

It follows that $\{\hat{\theta}_n^W\} \sim \mathcal{AN}(\theta, C_W)$ with asymptotic covariance matrix:

$$C_W = ([\dot{\mathbf{S}}^T\mathbf{N}]^T W [\dot{\mathbf{S}}^T\mathbf{N}])^{-1}[\dot{\mathbf{S}}^T\mathbf{N}]^T\, W \times \text{Cov}(\mathbf{S}^T\hat{N}_n) W\, [\dot{\mathbf{S}}^T\mathbf{N}]\, ([\dot{\mathbf{S}}^T\mathbf{N}]^T W [\dot{\mathbf{S}}^T\mathbf{N}])^{-1}. \tag{8}$$

Covariance matrix C_W depends on the choice of the weighting matrix W. This raises the issue of an optimal choice of the weighting matrix, i.e. the existence of an optimal matrix $W_\star$ such that $C_{W_\star} \leq C_W$ for all admissible W, this inequality being understood in terms of the partial ordering of the Hermitian matrices. The following lemma allows to conclude easily about optimality.

Lemma 1 *Let* Q *and* Γ *be two matrices with the same number of rows;* Q^TQ *invertible and* Γ *a non negative symmetric matrix. If* $\text{Span}(Q) \subset \text{Span}(\Gamma)$*, then for any symmetric matrix* W *such that* Q^TWQ *is invertible, it holds that*

$$(Q^TWQ)^{-1}\, Q^TW\Gamma WQ\, (Q^TWQ)^{-1} \geq (Q^T\Gamma^{\#}Q)^{-1}. \tag{9}$$

This is a classic inequality, holding unconditionally when matrix Γ is full-rank. However, our purpose requires to deal with possibly singular matrices Γ; As shown by the lemma, the inequality stills holds provided the range of Γ is 'large enough'. Straightforward application of lemma 1 with $\Gamma = \text{Cov}(\mathbf{S}^T\hat{N}_n)$ and $Q = [\dot{\mathbf{S}}^T\mathbf{N}]$ yields the following optimality theorem.

Theorem 3 *For an admissible pair* $(S(\cdot), \hat{N}_n)$ *and any admissible weighting matrix* W $C_W \geq \left(J^{S,N}\right)^{-1}$ *where*

$$J^{S,N} \stackrel{\text{def}}{=} [\dot{\mathbf{S}}^T\mathbf{N}]^T \text{Cov}^{\#}(\mathbf{S}^T\hat{N}_n)[\dot{\mathbf{S}}^T\mathbf{N}] \tag{10}$$

and this lower bound to the asymptotic covariance is reached for $W_\star = \text{Cov}^{\#}(\mathbf{S}^T\hat{N}_n)$.

A few comments are in order. First, this result obviously parallels the theory of maximum likelihood estimation in regular statistical models. Here $J^{S,N}$ plays the role of the Fisher information matrix in the M.L. framework. It apparently depends not only on the statistics of $\hat{N}_n$ but also on the particular function S used to express the orthogonality between subspaces. Next section is devoted to establishing that this lower bound does not depend on functions S and N but only on 'subspace quantities'. Second, it must be stressed that it may exist many different weighting matrices attaining the lower bound. In other words, $C_W = C_{W_\star}$ does not necessarily imply that $W = W_\star$. Finally, the optimal weighting depends on the parameter θ but may be shown that its substitution by a consistent estimate does not affect the asymptotic distribution of $\hat{\theta}_n^{W_\star}$.

2. INVARIANCE OF SUBSPACE FITTING ESTIMATES

This section is devoted to establishing two invariance properties related to optimal subspace fitting. The basic intuition is that optimal procedures 'tend' to be independent of specific parameterizations. For instance, estimation based on optimal matching of a statistic is invariant under invertible transformation of the statistic. Since subspace fitting estimation is ultimately based on orthogonality between spaces, it may be expected that the behavior of optimal estimates is governed only by 'subspace quantities'.

2.1. Pseudo-scores

Our approach is based on first-order stochastic expansions of estimates: If a sequence $\{\hat{\theta}_n\}$ of estimates of a parameter θ can be written as

$$\hat{\theta}_n = \theta + \Gamma_u^{-1}u_n + o_P(n^{-1/2}), \tag{11}$$

where $\{u_n\} \sim \mathcal{AN}(0, \Gamma_u)$, then $\{u_n\}$ is said to be a sequence of *pseudo-scores* for $\{\hat{\theta}_n\}$. This terminology is clearly in analogy to classical M.L. estimation theory. It is easily found that if a sequence $\{\hat{\theta}_n\}$ admits a sequence of pseudo-scores $\{u_n\} \sim \mathcal{AN}(0, \Gamma_u)$, then $\{\hat{\theta}_n\} \sim \mathcal{AN}(\theta, \Gamma_u^{-1})$. In the

following, we shall establish invariance properties by comparing pseudo-scores. Clearly, if two estimators are associated to pseudo-scores differing only by a $o_P(n^{-1/2})$ term, they have the same asymptotic distribution.

Pseudo-scores of optimally weighted subspace fitting estimates have a characteristic form:

Theorem 4 *The optimally weighted subspace fitting estimate based on an admissible pair $(S(\cdot), \hat{N}_n)$ admits a sequence $\{u_n^{(S,N)}\}$ of pseudo-scores given by*

$$u_n^{(S,N)} = [\mathbf{S}^T\dot{\mathbf{N}}]^T \mathrm{Cov}^{\#}(\mathbf{S}^T\hat{N}_n)\mathrm{Vec}(\mathbf{S}^T\hat{N}_n). \quad (12)$$

This form of the pseudo-score suggests that matrix $\mathbf{S}$ may factor out, leaving an expression of the pseudo-score not depending explicitly on function $S(\cdot)$. A key device for this kind of manipulation is the following lemma.

Lemma 2 *Let $\{v_n\}$ denote an asymptotically normal sequence: $\{v_n\} \sim \mathcal{AN}(0, \mathrm{Cov}(v))$. For two matrices A and B with compatible dimensions, sequences $\{u_n^{(1)}\}$ and $\{u_n^{(2)}\}$:*

$$u_n^{(1)} = A^T\mathrm{Cov}^{\#}(v)v \quad (13)$$
$$u_n^{(2)} = A^TB(B^T\mathrm{Cov}(v)B)^{\#}B^Tv, \quad (14)$$

are equivalent, i.e. $u_n^{(1)} = u_n^{(2)} + o_P(n^{-1/2})$, if the following two conditions hold:

$$\mathrm{rank}(B^T\mathrm{Cov}(v)) = \mathrm{rank}(\mathrm{Cov}(v)), \quad (15)$$
$$\mathrm{Span}(A) \subset \mathrm{Span}(\mathrm{Cov}(v)). \quad (16)$$

2.2. Invariance with respect to the probe

We start by proving invariance with respect to the probing function $S(\cdot)$. Denote $P_S(\eta)$ the orthogonal projector onto $\mathrm{Span}(S(\eta))$ and denote $\mathbf{P}_S$ its value at the true parameter $\eta = \theta$:

$$P_S(\eta) \stackrel{\mathrm{def}}{=} S(\eta)\,(S(\eta)^TS(\eta))^{\#}S(\eta)^T, \qquad \mathbf{P}_S \stackrel{\mathrm{def}}{=} P_S(\theta). \quad (17)$$

We want to relate estimates based on the pair $(S(\cdot), \hat{N}_n)$ to estimates based on the pair $(P_S(\cdot), \hat{N}_n)$. The first property to be established is that estimates based on the pair $(P_S(\cdot), \hat{N}_n)$ do exist. This is guaranteed by the following theorem.

Theorem 5 *If $(S(\cdot), \hat{N}_n)$ is admissible, so is $(P_S(\cdot), \hat{N}_n)$.*

It follows that theorem 4 applies to the optimally weighted subspace estimate based on the pair $(P_S(\cdot), \hat{N}_n)$. According to theorem 4, a pseudo-score associated to the pair $(P_S(\cdot), \hat{N}_n)$ is:

$$u_n^{(P_S,N)} = [\mathbf{P}_S\dot{\mathbf{N}}]^T \mathrm{Cov}^{\#}(\mathbf{P}_S\hat{N}_n)\mathrm{Vec}(\mathbf{P}_S\hat{N}_n). \quad (18)$$

It is not difficult to establish that

$$[\mathbf{S}^T\dot{\mathbf{N}}] = (I \otimes \mathbf{S})^T[\mathbf{P}_S\dot{\mathbf{N}}], \quad (19)$$
$$\mathrm{Vec}(\mathbf{S}^T\hat{N}_n) = (I \otimes \mathbf{S})^T\mathrm{Vec}(\mathbf{P}_S\hat{N}_n). \quad (20)$$

Note that inserting identities (20) and (19) into expression (12) of $u_n^{(S,N)}$ results in

$$u_n^{(S,N)} = [\mathbf{P}_S\dot{\mathbf{N}}]^T(I \otimes \mathbf{S})\left((I \otimes \mathbf{S})^T\mathrm{Cov}(\mathbf{P}_S\hat{N}_n)(I \otimes \mathbf{S})\right)^{\#} \times (I \otimes \mathbf{S})^T\mathrm{Vec}(\mathbf{P}_S\hat{N}_n). \quad (21)$$

which would be identical to expression (18) of $u_n^{(P_S,N)}$, *if matrix $I \otimes \mathbf{S}$ canceled out in* (21). That such a cancelation occurs is not *a priori* granted because matrix $I \otimes \mathbf{S}$ is not invertible. However, one may prove that for an admissible pair $(S(\cdot), \hat{N}_n)$:

$$\mathrm{rank}((I \otimes \mathbf{S})^T\mathrm{Cov}(\mathbf{P}_S\hat{N}_n)) = \mathrm{rank}(\mathrm{Cov}(\mathbf{P}_S\hat{N}_n)) \quad (22)$$
$$\mathrm{Span}([\mathbf{P}_S\dot{\mathbf{N}}]) \subset \mathrm{Span}(\mathrm{Cov}(\mathbf{P}_S\hat{N}_n)). \quad (23)$$

Thus the technical conditions required to apply lemma 2 are fulfilled : matrix $I \otimes \mathbf{S}$ does cancel out in (21) and we can conclude with this theorem.

Theorem 6 (Invariance w.r.t. the probe) *Optimally weighted subspace fitting estimates based on the pairs $(S(\cdot), \hat{N}_n)$ and $(P_S(\cdot), \hat{N}_n)$ have equivalent pseudo-scores, i.e. $u_n^{(S,N)} = u_n^{(P_S,N)} + o_P(n^{-1/2})$.*

2.3. Invariance with respect to the statistic

We turn to invariance w.r.t. the statistic. We define

$$P_N(\eta) \stackrel{\mathrm{def}}{=} N(\eta)\,(N(\eta)^TN(\eta))^{\#}N(\eta)^T, \qquad \mathbf{P}_N \stackrel{\mathrm{def}}{=} P_N(\theta). \quad (24)$$

The 'sample projector' $\hat{P}_N$ must be defined via a S.V.D. because the rank of $\hat{N}_n$ is not necessarily equal to the rank of $\mathbf{P}_N$. Thus, if ρ_N is the column rank of $\mathbf{N}$, matrix $\hat{P}_N$ is defined as the orthogonal projector onto the space spanned by the ρ_N most significant left singular vectors of $\hat{N}_n$. Without strengthening our assumptions, we can establish admissibility of the pair $(S(\cdot), \hat{P}_N)$.

Theorem 7 *If $(S(\cdot), \hat{N}_n)$ is admissible, so is $(S(\cdot), \hat{P}_N)$.*

Thus theorem 4 applies to the optimally weighted subspace estimates based on the pair $(S(\cdot), \hat{P}_N)$: they are associated to a pseudo-score $u_n^{(S,P_N)}$ given by:

$$u_n^{(S,P_N)} = [\mathbf{S}^T\dot{\mathbf{P}}_N]^T \mathrm{Cov}^{\#}(\mathbf{S}^T\hat{P}_N)\mathrm{Vec}(\mathbf{S}^T\hat{P}_N). \quad (25)$$

Contrary to invariance w.r.t. the probe $S(\cdot)$, an additional 'rank condition' is required to actually obtain invariance w.r.t. $\hat{N}_n$.

Definition 2 *We say that the* rank condition *is fulfilled when it holds that*

$$\mathbf{S}^T\hat{N}_n = \mathbf{S}^T\hat{N}_n\mathbf{N}^{\#}\mathbf{N} + o_P(n^{-1/2}). \quad (26)$$

One can show that the rank condition holds in several familiar contexts. For instance, it is verified if matrix $\mathbf{N}$ has full column rank or if matrix $\hat{N}_n$ has (almost surely) the same rank as its limiting value $\mathbf{N}$. Under the rank condition, we can prove

$$[\mathbf{S}^T\dot{\mathbf{N}}] = (\mathbf{N} \otimes I)^T[\mathbf{S}^T\dot{\mathbf{P}}_N], \quad (27)$$
$$\mathrm{Vec}(\mathbf{S}^T\hat{N}_n) = (\mathbf{N} \otimes I)^T\mathrm{Vec}(\mathbf{S}^T\hat{P}_N) + o_P(n^{-1/2}) \quad (28)$$

Inserting relations (27) and (28) in (12) yields

$$u_n^{(S,N)} = [\mathbf{S}^T\dot{\mathbf{P}}_N]^T(\mathbf{N} \otimes I)\left((\mathbf{N} \otimes I)^T\mathrm{Cov}(\mathbf{S}^T\hat{P}_N)(\mathbf{N} \otimes I)\right)^{\#} \times (\mathbf{N} \otimes I)^T\mathrm{Vec}(\mathbf{S}^T\hat{P}_N) + o_P(n^{-1/2})$$

which reduces to (25) *if matrix* $\mathbf{N} \otimes I$ *cancels out in* (29). For an admissible pair $(S(\cdot), \hat{N}_n)$, we can prove under the rank condition that

$$\text{rank}((\mathbf{N} \otimes I)^T \text{Cov}(\mathbf{S}^T \hat{P}_N)) = \text{rank}(\text{Cov}(\mathbf{S}^T \hat{P}_N)) \quad (29)$$

$$\text{Span}([\mathbf{S}^T \dot{\mathbf{P}}_N]) \subset \text{Span}(\text{Cov}(\mathbf{S}^T \hat{P}_N)). \quad (30)$$

Thus the technical conditions required to apply lemma 2 are verified : matrix $\mathbf{N} \otimes I$ does cancel out. We obtain:

Theorem 8 *Under the rank condition, optimally weighted subspace fitting estimates based on the pairs* $(S(\cdot), \hat{N}_n)$ *and* $(S(\cdot), \hat{P}_N)$ *have equivalent pseudo-scores:* $u_n^{(S,N)} = u_n^{(S,P_N)} + o_P(n^{-1/2})$.

2.4. Discussion

We now combine the results of previous sections. In particular the pair $(P_S(\cdot), \hat{P}_N)$ is admissible and under the rank condition, the optimal subspace fitting estimate based on $(P_S(\cdot), \hat{P}_N)$ admits a pseudo-score $u_n^{(P_S,P_N)}$ which is equivalent to $u_n^{(S,P)}$, $u_n^{(P_S,N)}$ and $u_n^{(S,P_N)}$. It also follows that the asymptotic covariance matrix of subspace fitting estimates based on any of the pairs $(S(\cdot), \hat{N}_n)$, $(P_S(\cdot), \hat{N}_n)$, $(S(\cdot), \hat{P}_N)$ and $(P_S(\cdot), \hat{P}_N)$ is lower bounded by J_F^{-1} where:

$$J_F \stackrel{\text{def}}{=} [\mathbf{P}_S^T \dot{\mathbf{P}}_N]^T \text{Cov}^{\#}(\mathbf{P}_S \hat{P}_N)[\mathbf{P}_S^T \dot{\mathbf{P}}_N]. \quad (31)$$

We have thus completed a first part of our invariance program in showing that only subspace quantities are relevant to subspace fitting. The next step is to consider the case when the spaces spanned by $\mathbf{S}$ and $\mathbf{N}$ are complementary: if $S(\theta)$ and $N(\theta)$ taken altogether span the whole space *i.e.*

$$P_N(\theta) + P_S(\theta) = I, \quad (32)$$

we say that the *saturation condition* is met. One can then express pseudo-scores and information matrices in terms of quantities pertaining to only one of the two (complementary) subspaces. For instance, one easily find that

$$J_F \stackrel{\text{def}}{=} [\mathbf{P}_S^T \dot{\mathbf{P}}_S]^T \text{Cov}^{\#}(\mathbf{P}_S \hat{P}_S)[\mathbf{P}_S^T \dot{\mathbf{P}}_S] \quad (33)$$

where $\hat{P}_S = I - \hat{P}_N$. The saturation condition makes it possible to compare subspace fitting to subspace matching as is done next.

3. RELATION TO SUBSPACE MATCHING

We consider optimal estimation based on $\hat{P}_N$ (subspace matching) and its relationship to subspace fitting estimates studied in previous sections. The saturation condition is assumed to hold throughout this section since no close relationships between the two approaches can be expected to be found otherwise.

Statistic matching estimates are obtained as

$$\tilde{\theta}_n^V = \arg\min_{\eta} \|\hat{P}_N - P_N(\eta)\|_V^2. \quad (34)$$

For appropriate choices of weighting matrix V, statistic matching estimates are $\mathcal{AN}(\theta, \tilde{\Sigma}_V)$ for some covariance matrix $\tilde{\Sigma}_V$. Theory of statistic matching can easily be adapted to the current context in spite of the fact that $\text{Cov}(\hat{P}_N)$ necessarily is rank deficient. In particular, one finds that the asymptotic covariance matrix $\tilde{\Sigma}_V$ is lower bounded by the inverse of matrix $\tilde{J}_M$:

$$\tilde{J}_M \stackrel{\text{def}}{=} [\dot{\mathbf{P}}_N]^T \text{Cov}^{\#}(\hat{P}_N)[\dot{\mathbf{P}}_N] \quad (35)$$

whose relation to J_F is given below. It is well known that under appropriate regularity conditions, $\tilde{J}_M^{-1}$ is an asymptotic lower bound to any estimate of θ obtained as a function of $\hat{P}_N$.

We set out to comparing subspace matching estimates based on $\hat{P}_N$ to subspace fitting estimates based on the pair $(P_S(\cdot), \hat{P}_N)$.

Theorem 9 *Under the saturation condition* (32), for any admissible weight W, $\hat{\theta}_n^W = \tilde{\theta}_n^V + o_P(n^{-1/2})$ for $V = (\mathbf{P}_N \otimes \mathbf{P}_S)^T W (\mathbf{P}_N \otimes \mathbf{P}_S)$.

Thus we see that any subspace fitting estimate using weight W is equivalent to some subspace matching estimate with a weight V which is simply related W. The converse property is more difficult to establish because not any matrix V can be put in the form mentioned in theorem 9. However, we do we have

Theorem 10 *Under the saturation condition* (32), for any admissible matrix V, $\tilde{\theta}_n^V = \hat{\theta}_n^W o_P(n^{-1/2})$ for some admissible weight W.

Unfortunately, space is lacking to describe how the weight W relates to V in th. 10. A direct consequence of the previous two theorems is

Theorem 11 *Under saturation condition* (32), $J_M = J_F$.

This theorem shows that optimally weighted subspace fitting estimates are asymptotically 'efficient' under saturation in the sense that their asymptotic performance reaches the lower bound set by optimal subspace matching estimation.

REFERENCES

[1] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramér-Rao bounds," *IEEE Tr. on Acoust. Speech and Sig. Proc.*, vol. 37, pp. 720–741, 1989.

[2] M. Viberg, *Subspace fitting concepts in sensor array processing.* PhD thesis, Department of Electrical Engineering, Linkoping Univ., 1989.

[3] M. Gurelli and C. Nikias, "EVAM: an eigenvector-based algorithm for multichannel blind deconvolution of input colored signals," *IEEE Tr. on Sig. Proc.*, vol. 43, pp. 134–149, Jan. 1995.

[4] P. Stoica, M. Cedervall, and A. Eriksson, "Combined instrumental variable and subspace fitting approach to parameter estimation of noisy input-output system," *IEEE Tr. on Sig. Proc.*, vol. 43, pp. 2386–2397, Oct. 1995.

[5] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A class of subspace methods for blind identification of certain MA systems." Accepted for publication in I.E.E.E. Trans. on Info. Theory, July 1996.

Independent Component Analysis Based on Higher-Order Statistics Only*

Lieven De Lathauwer, Bart De Moor, Joos Vandewalle
K.U.Leuven - E.E. Dept.- ESAT - SISTA
Kard. Mercierlaan 94, B-3001 Leuven (Heverlee), Belgium
tel: 32/16/321085 fax: 32/16/321986
email: Lieven.DeLathauwer@esat.kuleuven.ac.be

Abstract

Most conventional techniques for Independent Component Analysis (ICA) resort to second-order statistics to decorrelate the observed data. The prewhitening step makes these algorithms sensitive to the presence of additive Gaussian noise. In this paper a higher-order-only technique is presented. The identification problem is approached in a (linear and multilinear) algebraic framework: our derivation starts with the observation that the solution can be obtained from the Canonical Decomposition (CANDECOMP) of a higher-order cumulant tensor. Next, it is demonstrated that the CANDECOMP components follow from the simultaneous diagonalization, by congruence transformation, of a set of matrices. A reformulation in terms of orthogonal unknowns leads to a simultaneous Schur decomposition, which is solved by a Givens-type iteration. The technique can be considered as the higher-order-only equivalent of the popular JADE-algorithm.

1 Introduction

The basic statistical model for *Independent Component Analysis* (ICA), or *Blind Source Separation*, is in this paper denoted as:

$$Y = \mathbf{M}X + N \qquad (1)$$

in which the observed vector Y, the source vector X and the noise vector N are zero-mean random vectors with values in $\mathbb{R}$ or $\mathbb{C}$. The components of X are mutually statistically independent, as well as statistically independent from the noise components. The goal of ICA now consists of the estimation of the transfer matrix (or "mixture matrix") $\mathbf{M}$ and the corresponding realizations of X, given only realizations of Y.

Without a priori knowledge the ICA-problem cannot be solved using only second-order statistics. Usually the second-order statistics of the observation vector Y are used for a whitening of the data. In this way the transfer matrix can be estimated up to an orthogonal factor $\mathbf{U}$. In the second step $\mathbf{U}$ is then obtained from higher-order cumulants of the standardized data. Several algorithms have been presented in literature. Among the most well-known approaches are the one by Comon [3] (further analyzed in [7]), where $\mathbf{U}$ is computed by a Jacobi-type diagonalization of the standardized cumulant tensor, and the JADE-algorithm (Joint Approximate Diagonalization Estimation) by Cardoso and Souloumiac [2], where $\mathbf{U}$ is found as the solution of a simultaneous eigenvalue decomposition.

In our paper the problem is solved using *only* the *higher-order* cumulant. This approach has the advantage that it is conceptually blind for the noise term N, when this term is Gaussian. For simplicity of notation, the exposition in this summary is restricted to fourth-order processing of real-valued data. The technique can be applied to cumulants of any order (higher than 2), as well as to complex data.

The paper is organized as follows. In the next section the relation between the columns of $\mathbf{M}$ and the fourth order observation cumulant is explicited. This relation takes the form of a tensorial decomposition of the cumulant in a sum of symmetric rank-1 tensors, and the uniqueness of this decomposition is discussed. In Section 3 the estimation of the transfer matrix from the cumulant model is presented as a simultaneous congruence transformation. In Section 4 it is

*This research was partially supported by the Belgian Program on Interuniversity Attraction Poles (IUAP-17, IUAP-50), the European Community Research program ESPRIT, Basic Research Working Group nr. 6620 (ATHOS), the Flemish Institute for Support of Scientific-Technological Research in Industry (I.W.T.) and is part of a Concerted Action Project of the Flemish Community, entitled "Model-based Information Processing Systems". Lieven De Lathauwer is a Research Assistant supported by the I.W.T. Bart De Moor is a Research Associate of the National Fund for Scientific Research (N.F.W.O.) The scientific responsibility is assumed by the authors.

0-8186-7576-4/96 $5.00 © 1996 IEEE

explained how the problem can be reformulated in terms of unknown orthogonal matrices. This leads to the simultaneous Schur decomposition of Section 5, for which a Givens-type computation scheme is derived. Section 6 contains a concluding discussion.

2 Canonical Decomposition

2.1 Model

As the name already suggests, the CANDECOMP of higher-order tensors is a basic concept in multilinear algebra. First we define the tensorial outer product of a set of vectors:

Definition 1 *The outer product of the vectors* $U^{(1)} \in \mathbb{R}^{I_1}$, $U^{(2)} \in \mathbb{R}^{I_2}$, ..., $U^{(N)} \in \mathbb{R}^{I_N}$, *denoted as* $U^{(1)} \circ U^{(2)} \circ \ldots \circ U^{(N)}$, *is an* $(I_1 \times I_2 \times \ldots \times I_N)$*-tensor* $\mathcal{A}$, *defined by the following element-wise equation:*

$$a_{i_1 i_2 \ldots i_N} = u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(N)}_{i_N}$$

In analogy with the vector/matrix case, the outer product leads to the definition of rank-1 tensors:

Definition 2 *An* N*th-order tensor* $\mathcal{A}$ *has rank* 1 *when it equals the outer product of* N *vectors* $U^{(1)}, U^{(2)}, \ldots, U^{(N)}$.

These elementary definitions allow to define the CANDECOMP:

Definition 3 *The Canonical Decomposition (CANDECOMP) of an* N*th-order tensor* $\mathcal{A}$ *is the decomposition of* $\mathcal{A}$ *in a minimal sum of rank-1 components.*

The decomposition is also known as *Parallel Factors Model* (PARAFAC). It can be considered as the tensorial generalization of the diagonalization of matrices by equivalence transformation (unsymmetric case) or by congruence transformation (symmetric case). Despite the importance of CANDECOMP no robust general computation schemes have been proposed in the past.

2.2 Link with ICA

When the noise N is Gaussian, it does not contribute to the fourth-order cumulant of Y. This cumulant, denoted by $\mathcal{C}$, then shows the following structure:

$$\mathcal{C} = \sum_p \kappa_p M_p \circ M_p \circ M_p \circ M_p \qquad (2)$$

where κ_p denotes the fourth-order cumulant of the pth source ($1 \leqslant p \leqslant P$) and M_p symbolizes the pth "steering vector" (i.e. the pth column of $\mathbf{M}$). Eq. (2) is clearly a symmetric CANDECOMP-model. The contribution of a non-Gaussian noise component, and the effect of other estimation errors when $\mathcal{C}$ is a finite sample cumulant, is considered as a perturbation of the equation.

2.3 Uniqueness

The uniqueness properties of CANDECOMP and its matrix counterparts are thoroughly different. Here we assume that the transfer matrix is square and regular, and that all the sources have non-vanishing kurtosis. It can be proved [9] that these conditions are sufficient to guarantee that decomposition (2) is unique up to the following trivial indeterminacies:

- permutation of the terms
- scaling of the steering vectors with a factor α_p, combined with inverse scaling (factor α_p^{-4}) of the coefficients κ_p.

(The interested reader is referred to the overview paper [4] for a discussion of some other uniqueness properties.) Note that in our setting different sources can have the same probability distribution, as long as they are mutually statistically independent in fourth order. The conditions can be weakened for the identification of at most one non-kurtic source. It is also possible to handle the "more-sensors-than-sources" case.

3 Simultaneous Congruence Transformation

We associate to $\mathcal{C}$ a linear matrix transformation in the following way:

$$\mathbf{B} = \mathcal{C}(\mathbf{A}) \quad \Longleftrightarrow \quad b_{ij} = \sum_{kl} c_{ijkl} a_{kl} \qquad (3)$$

for all index values. From Eq. (2) follows that every matrix in the range space of $\mathcal{C}$ can be written as a linear combination of the "steering matrices" $M_p M_p^T$ ($1 \leqslant p \leqslant P$). In other words, the transfer matrix $\mathbf{M}$ diagonalizes every matrix in the range space of $\mathcal{C}$ by congruence transformation. Assume a basis for the range space is given by $\mathbf{T_1}, \mathbf{T_2}, \ldots, \mathbf{T_P}$, then we have the following *simultaneous congruence transformation*:

$$\begin{aligned} \mathbf{T_1} &= \mathbf{M} \cdot \mathbf{D_1} \cdot \mathbf{M}^T \\ \mathbf{T_2} &= \mathbf{M} \cdot \mathbf{D_2} \cdot \mathbf{M}^T \\ &\vdots \\ \mathbf{T_P} &= \mathbf{M} \cdot \mathbf{D_P} \cdot \mathbf{M}^T \end{aligned} \qquad (4)$$

where $\mathbf{D_1}, \mathbf{D_2}, \ldots, \mathbf{D_P}$ are diagonal. Remarkably, a similar set of equations arises in the blind separation of constant-modulus signals [11], which suggests a weird link between the constant-modulus property and non-Gaussianity. The computational technique of this paper differs from the one presented in [11], the latter being suboptimal.

Although two equations in (4) are generally sufficient to estimate the transfer matrix, we prefer to solve the complete set simultaneously, in order to exploit all the available information. This can be substantiated by numerical arguments [9].

The simultaneous solution of Eq. (4) is the higher-order-only equivalent of the *simultaneous eigenvalue decomposition* on which the ICA-algorithm by Cardoso and Souloumiac is based [2]. In the latter algorithm pre-whitening leads to a simultaneous matrix decomposition from which an orthogonal matrix has to be computed; in the current approach a general regular matrix has to be determined (up to the indeterminacies mentioned in Section 2.3), corresponding to the fact that there is no pre-whitening.

4 A new matrix representation

The fact that the unknown transfer matrix is basically an arbitrary regular matrix, makes it hard to deal with in a proper numerical way. Therefore we will represent the mixture matrix by a pair of *orthogonal* matrices, obtained from the QR-factorisation $\mathbf{M} = \mathbf{Q}^T\mathbf{R}'$ and the RQ-decomposition $\mathbf{M}^T = \mathbf{R}''\mathbf{Z}^T$. The pair $(\mathbf{Q},\mathbf{Z})$ is actually an equivalent representation of the transfer matrix, within the limits of identifiability. From the definition of $\mathbf{Q}$ and $\mathbf{Z}$ we have:

$$(\mathbf{Q}\cdot\mathbf{Z})\cdot\mathbf{R}''^T = \mathbf{R}' \tag{5}$$

The orthogonal matrix $\mathbf{Q}\cdot\mathbf{Z}$ will be denoted as $\mathbf{V}$. The lower triangular part of Eq. (5) is a system of linear equations in the unknown coefficients of $\mathbf{R}''$:

$$\begin{bmatrix} v_{P,P-1} & v_{PP} \end{bmatrix}\begin{bmatrix} r''_{P-1,P-1} \\ r''_{P-1,P} \end{bmatrix} = 0$$

$$\begin{bmatrix} v_{P-1,P-2} & v_{P-1,P-1} & v_{P-1,P} \\ v_{P,P-2} & v_{P,P-1} & v_{P,P} \end{bmatrix}\begin{bmatrix} r''_{P-2,P-2} \\ r''_{P-2,P-1} \\ r''_{P-2,P} \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}^T \tag{6}$$

$$\vdots$$

Note that a scaling of the rows of $\mathbf{R}''$ does not affect this homogeneous set of equations, which is consistent with the fact that the steering vectors can only be determined up to a scalar multiple. By substitution of $\mathbf{R}''$ in Eq. (5) $\mathbf{R}'$ can be found as well.

5 Simultaneous Schur Decomposition

5.1 Principle

The notation of the simultaneous congruence transformation in terms of $\mathbf{Q}$ and $\mathbf{Z}$ leads to a set of matrix equations that we will denote as a *simultaneous Schur decomposition*:

$$\begin{aligned} \mathbf{Q}\cdot\mathbf{T_1}\cdot\mathbf{Z} &= \mathbf{R_1} = \mathbf{R}'\cdot\mathbf{D_1}\cdot\mathbf{R}'' \\ \mathbf{Q}\cdot\mathbf{T_2}\cdot\mathbf{Z} &= \mathbf{R_2} = \mathbf{R}'\cdot\mathbf{D_2}\cdot\mathbf{R}'' \\ &\vdots \\ \mathbf{Q}\cdot\mathbf{T_P}\cdot\mathbf{Z} &= \mathbf{R_P} = \mathbf{R}'\cdot\mathbf{D_P}\cdot\mathbf{R}'' \end{aligned} \tag{7}$$

From these equations the orthogonal matrices $\mathbf{Q}$ and $\mathbf{Z}$ have to be determined such that $\mathbf{R_1}, \mathbf{R_2}, \ldots, \mathbf{R_P}$ are "as upper triangular as possible" (in least-squares sense). The criterion function f to be minimized can be written as:

$$f(\mathbf{Q},\mathbf{Z}) = \|\mathbf{Q}\cdot\mathbf{T_1}\cdot\mathbf{Z}\|^2_{LF} + \ldots + \|\mathbf{Q}\cdot\mathbf{T_P}\cdot\mathbf{Z}\|^2_{LF} \tag{8}$$

in which $\|\mathbf{A}\|_{LF}$ denotes the below-diagonal Frobenius-norm of $\mathbf{A}$, i.e.

$$\|\mathbf{A}\|_{LF} = (\sum_{j<i}\sum_{i} a_{ij}^2)^{1/2} \tag{9}$$

It can be proved that the criterion f satisfies all the conditions for a higher-order-only contrast function [3, 5] that discriminates over the set of regular transfer matrices. Contrarily to classical approaches this contrast depends on two orthogonal matrices.

5.2 Solution by Givens-iteration

The core of our method is the computation of $\mathbf{Q}$ and $\mathbf{Z}$ from Eq. (7). The criterion function $f(\mathbf{Q},\mathbf{Z})$ is optimized by an iteration technique, in which $\mathbf{Q}$ and $\mathbf{Z}$ are determined as a sequence of elementary Givens rotations. Each elementary rotation makes the set $\mathbf{R_1}, \mathbf{R_2}, \ldots, \mathbf{R_P}$ simultaneously as upper triangular as possible.

First, the estimates of $\mathbf{Q}$ and $\mathbf{Z}$ are initialized as any orthogonal matrix, e.g. as the identity matrix: $\mathbf{Q}^{(0)} = \mathbf{Z}^{(0)} = \mathbf{I}$. The estimates of $\mathbf{R_1}, \mathbf{R_2}, \ldots, \mathbf{R_P}$ are initialized accordingly: $\mathbf{R_1}^{(0)} = \mathbf{T_1}$, $\mathbf{R_2}^{(0)} = \mathbf{T_2}$, ..., $\mathbf{R_P}^{(0)} = \mathbf{T_P}$. In each iteration step k either $\mathbf{Q}^{(k)}$ or $\mathbf{Z}^{(k)}$ is updated. An update of $\mathbf{Q}^{(k)}$ takes the form of $\mathbf{Q}^{(k+1)} = \mathbf{G_{ij}}\cdot\mathbf{Q}^{(k)}$, in which $\mathbf{G_{ij}}$ denotes an elementary Givens rotation that affects rows i and j; at the same time $\mathbf{R_1}^{(k)}, \mathbf{R_2}^{(k)}, \ldots, \mathbf{R_P}^{(k)}$ are updated as $\mathbf{R_1}^{(k+1)} = \mathbf{G_{ij}}^T\cdot\mathbf{R_1}^{(k)}$, $\mathbf{R_2}^{(k+1)} = \mathbf{G_{ij}}^T\cdot\mathbf{R_2}^{(k)}$, ..., $\mathbf{R_P}^{(k+1)} = \mathbf{G_{ij}}^T\cdot\mathbf{R_P}^{(k)}$. $\mathbf{Z}^{(k)}$ is updated in a similar way, by working on the columns.

Let us focus on the updating of $\mathbf{Q}^{(k)}$ by multiplication with $\mathbf{G_{ij}}$. The Givens-rotation should be determined such that it minimizes the below-diagonal norm of all the ith and jth rows in $\mathbf{R_1}^{(k+1)}, \ldots, \mathbf{R_P}^{(k+1)}$. If we define

$$\mathbf{E_{ij}}^{(k)} = \begin{bmatrix} (\mathbf{R_1}^{(k)})_{i,\,i:j-1} & \ldots & (\mathbf{R_K}^{(k)})_{i,\,i:j-1} \\ (\mathbf{R_1}^{(k)})_{j,\,i:j-1} & \ldots & (\mathbf{R_P}^{(k)})_{j,\,i:j-1} \end{bmatrix} \tag{10}$$

using MATLAB-notation, then $\mathbf{G_{ij}}$ should be determined such that the second row of $\mathbf{E_{ij}^{(k+1)}} = \mathbf{G_{ij}^T} \cdot \mathbf{E_{ij}^{(k)}}$ has minimal Frobenius-norm. Hence $\mathbf{G_{ij}}$ can be obtained as the left singular matrix of $\mathbf{E_{ij}^{(k)}}$: in this way the norm of the second row equals the smallest singular value of $\mathbf{E_{ij}^{(k)}}$, which is the best one can do. In simulations the Givens-iteration shows a monotonous convergence to the global optimum of $f(\mathbf{Q}, \mathbf{Z})$.

6 Discussion and conclusions

We presented a new ICA-technique that resorts only to the higher-order cumulants of the observations. The transfer matrix estimate that is obtained shows exactly the same uniqueness properties as in the classical ICA-algorithms [2, 3], which also exploit second-order information in a prewhitening step. Higher-order-only Blind Source Separation has the advantage that it is asymptotically insensitive to additive Gaussian perturbations of the data. When dealing with finite sample cumulants, the accuracy of higher-order-only versus classical approaches is subject to a trade-off, caused by the fact that higher-order statistics are harder to estimate than second-order statistics [1, 10].

Our approach is based on the observation that the data cumulant can be expanded as a sum of rank-1 tensors. This implies that all the matrices in the range space of the cumulant tensor, considered as a super-symmetric matrix-to-matrix mapping, satisfy a simultaneous congruence transformation. For the numerical computation of this set of matrix equations we proposed a new representation of the transfer matrix: it turns out that any matrix, of which the columns are fixed up to multiplication with a scalar, can be represented by a pair of orthogonal matrices, obtained by QR- and RQ-factorisation. In this new format the simultaneous congruence transformation takes the form of a simultaneous Schur decomposition, that can be computed by a Givens-type iteration. The result can be considered as an approximate solution of *other* cumulant-based identification criterions, e.g. least-squares cumulant matching can be realized by means of a standard optimization routine, using the simultaneous Schur-solution as starting value [8].

The technique established in this paper is in fact the higher-order-only equivalent of the well-known combined second/higher-order ICA-algorithm by Cardoso and Souloumiac [2]. The concepts of this paper also lead to a higher-order-only equivalent ([9]) of the ICA-algorithm by Comon [3].

The technique can also be generalized for higher-order tensors without symmetry properties [6]. The unsymmetric version of the algorithm can be used for Factor Analysis of multiway datasets.

References

[1] C. Bourin and P. Bondon. Efficiency of high-order moment estimates. *Proc. IEEE SP / ATHOS Workshop on Higher-Order Statistics*: 186-190, Girona, Spain, June 12-14, 1995.

[2] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. *IEE Proceedings-F*, 140(6): 362-370, 1994.

[3] P. Comon. Independent component analysis, a new concept? *Signal Processing*, Special Issue *Higher Order Statistics*, 36(3): 287-314, April 1994.

[4] P. Comon and B. Mourrain. Decomposition of quantics in sums of powers of linear forms. Accepted for publication in *Signal Processing*.

[5] P. Comon, L. De Lathauwer and B. De Moor. A contrast-based independent component analysis without second-order moments. *Proc. SSAP-96*.

[6] L. De Lathauwer, B. De Moor and J. Vandewalle. Canonical decomposition of a fourth-order tensor. *IMA Conf. on Linear Algebra and Its Applications*, Manchester, U.K., July 10-12, 1995.

[7] L. De Lathauwer, P. Comon, B. De Moor and J. Vandewalle. Higher-order power method - Application in independent component analysis. *Proc. NOLTA'95*, 1: 91-96, Las Vegas, USA, December 10-14, 1995.

[8] L. De Lathauwer, B. De Moor and J. Vandewalle. A technique for higher-order-only blind source separation. *Proc. ICONIP'96*, Hong Kong, September 24-27, 1996 (to be published).

[9] L. De Lathauwer. *Signal Processing by Higher-Order Tensors*. Ph.D. thesis, K.U.Leuven, E.E.Dept.-ESAT, Belgium (in preparation).

[10] J. Fonollosa. Sample cumulants of stationary processes: asymptotic results. *IEEE Trans. on Signal Processing*, 43 (4): 967-977, April 1995.

[11] A.-J. van der Veen and A. Paulraj. An analytical constant modulus algorithm. Submitted to *IEEE Trans. Signal Processing*.

Cross-Correlation Based Multichannel Blind Equalization

Haralambos Pozidis and Athina P. Petropulu
Electrical and Computer Engineering Department
Drexel University, Philadelphia, PA 19104, USA
pozidis@cbis.ece.drexel.edu, athina@artemis.ece.drexel.edu

Abstract

A novel cross-correlation based framework is proposed for the problem of blind equalization in communications. We assume that we have access to two observations, corresponding to the outputs of two channels excited by the same input. We propose a new algorithm which estimates the channels using as basic tool the phase of the cross spectrum of functions of the observations. The proposed method is computationally attractive, requires small input sample sizes, and performs well in low signal-to-noise ratios.

1. Introduction

Blind equalization is the problem of reconstructing a signal from a filtered version of it, without knowledge of the signal nor the filter. Research results dealing with the case of non-white signals have been reported in the past [1], [4], [5], [2], [3], using either multiple observations of the distorted signal [4], or oversampling of the received signal, [2], [3]. Both approaches lead to a multichannel scenario where the input signal is to be estimated from multiple distorted versions of it.

The approach of [4], [5] uses higher-order statistics of the observations, while that of [2], [3] is based on the cyclic autocorrelation of the observations. The second approach has the advantage that it can be applied to any type of input signals, as opposed to the first approach which applies to non-Gaussian signals only. The cyclic approach has a lower complexity compared to the higher-order statistics based approach, however its performance and the uniqueness of the solution is critically related to the knowledge (or ability to get exact estimates) of the lengths of the unknown channels.

In this paper we present a new cross correlation based approach which estimates the channels by simultaneously minimizing two error criteria involving the phase of a combination of the two channels. This phase is estimated from the observations based on cross correlation operations. We show that the proposed method is not very sensitive to channel lengths mismatch, requires small input sample sizes, and performs well in low SNR's.

2. Problem Formulation

The two channel case will be described next, however the results can be easily extended to the multichannel case. The unknown system model is described by

$$x_i(k) = h_i(k) * s(k) + n_i(k), \qquad i = 1, 2 \qquad (1)$$

$x_i(k)$, $i = 1, 2$ denote the observations; $h_i(k), i = 1, 2$ are the unknown FIR channels; $s(k)$ is stationary, generally non-white, zero-mean random process; $n_i(k)$ are noise processes uncorrelated to each other and to $s(k)$. It is assumed that $h_1(k)$ and $h_2(k)$ have no common zeros, that there are no zero-pole cancellations between $h_i(k), i = 1, 2$ and convolutional components of $s(k)$, and that there are no common zeros between convolutional components of $s(k)$ and each of the channels. Under these conditions, the channels $h_1(k)$ and $h_2(k)$ are identifiable within a constant and a delay. In the sequel we present an algorithm that performs the identification task.

3. The Cross-Correlation Blind Equalization Algorithm

Let us model the random process $s(k)$ as

$$s(k) = e(k) * h(k) \qquad (2)$$

where $e(k)$ is a white, zero-mean process. Combining (1) and (2) we get

$$x_i(k) = e(k) * h(k) * h_i(k) + n_i(k) = e(k) * g_i(k) + n_i(k) \qquad (3)$$

0-8186-7576-4/96 $5.00

The cross correlation of $x_1(k)$ and $x_2(k)$ equals

$$r_{x_1x_2}(k) = E\{x_1(n+k)x_2^*(n)\} = \gamma_2^e \, g_1(k) * g_2^*(-k) \quad (4)$$

where γ_2^e is the variance of $e(k)$. The contribution of $n_1(k)$ and $n_2(k)$ to (4) is zero, due to the fact that the noise processes are zero-mean and uncorrelated to $e(n)$.

The minimum phase equivalent of an all-zero sequence $y(n)$, denoted by $\tilde{y}^{min}(n)$, is a minimum phase sequence, whose zeros consist of the minimum phase zeros of $y(n)$ and the maximum phase zeros of $y(n)$, reflected inside the unit circle at their conjugate reciprocal locations. One simple method to estimate $\tilde{y}^{min}(n)$ is the power cepstrum based approach,i.e.,

$$\begin{aligned} \tilde{y}^{min}(n) &= F^{-1}\{exp[C(\omega)]\} \\ c(n) &= F^{-1}\{log|Y(\omega)|^2\}u(n), \end{aligned} \quad (5)$$

where $C(\omega)$ is the Fourier transform of $c(n)$ and $u(n)$ is the unit step function. The minimum phase equivalent of a random sequence $x(k) = h(k) * e(k)$ where $e(k)$ is white and $h(k)$ deterministic, is equal to the minimum phase equivalent of the deterministic part $h(n)$.

Let $d(k)$ be defined as

$$d(k) = r_{x_1x_2}(k) * \tilde{g}_1^{min}(k) * (\tilde{g}_2^{min}(-k))^* \quad (6)$$

The z-transform of a deterministic, FIR, generally complex sequence can be decomposed as

$$H(z) = c_h z^{-r_h} I_h(z^{-1}) O_h(z) \quad (7)$$

where $I_h(z^{-1}), O_h(z)$ are the minimum and maximum phase parts of $h(n)$ respectively, c_h is a constant and r_h equals the number of zeros of $h(n)$ outside the unit circle.

Taking the Fourier transform of $d(k)$ and using (4) and (7), we get

$$D(z) = \gamma_2^e c_{g_1} c_{g_2}^* z^{r_{g_1} - r_{g_2}} [I_{g_1}(z^{-1}) I_{g_2}^*(z^*)]^2 P_1(z), \quad (8)$$

with

$$P_1(z) = O_{g_1}(z) O_{g_1}^*(1/z^*) O_{g_2}(z) O_{g_2}^*(1/z^*) \quad (9)$$

Since $P_1(\omega)$ is zero-phase, (8) leads to the following phase relation

$$\begin{aligned} arg\{D(\omega)\} &= 2arg\{I_{g_1}(\omega) I_{g_2}^*(\omega)\} + (r_{g_1} - r_{g_2})\omega \\ &= 2arg\{I_{h_1}(\omega) I_{h_2}^*(\omega)\} + (r_{h_1} - r_{h_2})\omega \end{aligned} \quad (10)$$

where we ignored the phase contribution of the terms c_{g_1} and c_{g_2}, since this is an additive constant, contributing a complex scalar to the corresponding time domain signal.

Equation (10) is of key importance, since the equalization scheme described in the sequel, is based on it.

Let us consider the filtered observations $y_i(k), i = 1, 2$ obtained as:

$$y_i(k) = x_i(k) * w_i(k) \quad (11)$$

where $w_i(k), i = 1, 2$ are FIR channels. Let

$$\begin{aligned} E_{min}(z) &= I_{y_1}(z) I_{y_2}^*(1/z^*) \\ E_{max}(z) &= O_{y_2}(1/z) O_{y_1}^*(z^*) \end{aligned} \quad (12)$$

be the cross spectrum of the minimum and maximum phase parts of the adaptive filter outputs respectively. We show that

$$\begin{aligned} &arg\{E_{min}(\omega)\} = 0 \quad \forall\, \omega \iff \\ &i_{h_1}(n) = i_{w_2}(n) \;\; and \;\; i_{h_2}(n) = i_{w_1}(n). \end{aligned} \quad (13)$$

and similarly,

$$\begin{aligned} &arg\{E_{max}(\omega)\} = 0 \quad \forall\, \omega \iff \\ &o_{h_1}(n) = o_{w_2}(n) \;\; and \;\; o_{h_2}(n) = o_{w_1}(n) \end{aligned} \quad (14)$$

The proofs can be found in [5], [6]. Combining Propositions 1 and 2, we get that

$$\begin{aligned} &arg\{E_{min}(\omega)\} = 0 \;\; and \;\; arg\{E_{max}(\omega)\} = 0 \; \forall\, \omega \iff \\ &h_1(n) = w_2(n) \quad and \quad h_2(n) = w_1(n). \end{aligned} \quad (15)$$

3.1. Channel Identification/Equalization

Let

$$W_{min}(\omega) = I_{w_1}(\omega) I_{w_2}^*(\omega), \quad H_{min}(\omega) = I_{h_1}(\omega) I_{h_2}^*(\omega), \quad (16)$$

and also let us assume temporarily that the length of $w_{min}(n)$, L_{min} is known. Setting the phase of $E_{min}(\omega)$ to zero and after some mathematical manipulations we get

$$\begin{aligned} &\textstyle\sum_{n=-N_2, n\neq 0}^{N_1} w_{min}^R(n) sin(\omega n - \psi_{min}(\omega)) - \\ &\textstyle\sum_{n=-N_2}^{N_1} w_{min}^I(n) sin(\omega n + \psi'_{min}(\omega)) = sin(\psi_{min}(\omega)) \end{aligned} \quad (17)$$

where

$$\psi_{min}(\omega) = arg\{H_{min}(\omega)\}, \quad \psi'_{min}(\omega) = \frac{\pi}{2} - \psi_{min}(\omega), \quad (18)$$

N_1 is the length of the causal part of $w_{min}(n)$, and N_2 the length of its noncausal part. Through (10), (18) becomes

$$\psi_{min}(\omega) = \frac{1}{2}(arg\{D(\omega)\} - (r_{h_1} - r_{h_2})\omega) \quad (19)$$

In (17) it was taken $w_{min}^R(0) = 1$.

Repeating (17) for different ω's in the set $\{\omega = \frac{2\pi}{L}k, k = 0, \ldots, L/2\}$, we can form the system of equations:

$$\mathbf{\Phi}_{min}\mathbf{w}_{min} = \boldsymbol{\phi}_{min}, \qquad (20)$$

and subsequently solve it for $\mathbf{w}_{min}$ via least-squares. An adaptive solution can be obtained via LMS algorithm.

Due to the structure of $w_{min}(n)$, as defined in (16), and from Proposition 1, the cepstrum of $w_{min}(n)$ equals

$$\hat{w}_{min}(n) = \hat{i}_{h_2}(n) + \hat{i}^*_{h_1}(-n) \qquad (21)$$

where $\hat{i}_{h_i}(n)$ denotes the cepstrum of the minimum phase part of $h_i(n)$. Since $\hat{i}_{h_2}(n)$ is a causal sequence and $\hat{i}^*_{h_1}(n)$ is a noncausal sequence, (21) can be used to obtain the causal cepstra of the two channels. A similar procedure can be followed to yield the noncausal cepstra $\hat{o}_{h_1}(n)$ and $\hat{o}_{h_2}(n)$, based on the minimization of the phase of $E_{max}(\omega)$. Finally, the channels can be reconstructed via inverse cepstra operations, within a scalar and a time delay. In the channel equalization case, where the inverse channels are of interest, they can be reconstructed as

$$h_i^{inv}(n) = F\{e^{F\{-\hat{h}_i(n)\}}\}, \quad i = 1, 2 \qquad (22)$$

In many cases the inverse channel obtained that way may enhance the noise at the receiver, thus raising the probability of error at the decision device. This problem can be bypassed by using the so-called constrained Wiener filter approach [7], which estimates the desired input symbols in a least-squares sense. This method was used in our case.

4. Implementation Issues

There are several issues in the reconstruction procedure that have to be addressed. As seen from (21), $w_{min}(n)$ is a two-sided sequence of unknown length L_{min}. The length of w_{min}, together with the length N_2 of its causal part, have to be taken into account in forming (17). Moreover, the phase $\psi_{min}(\omega)$, as given by (19), contains a linear phase component $(r_{h_1} - r_{h_2})\omega$ and also a constant phase $c = arg\{c_{h_1}\} - arg\{c_{h_2}\}$, both of which are unknown. Neglecting the term c will result in a sequence $w_{min}(n)$ which will differ by the true one by a complex constant. The linear phase component does not possess additional problems also, since it corresponds to a circular shift of the original sequence.

Let us suppose that L_{min} is known. Assuming an incorrect value for N_2, will effectively shift $w_{min}(n)$. This will appear as a circular shift in the reconstructed sequence $w_{min}(n)$. Therefore, if two solutions exist for two different values of shift, they should differ by a time delay. If no solution exists for some amount of shift, the algorithm will not converge. To determine the existence of a solution we look for a low value of the mean square error corresponding to the shortest length $L_{min}(L_{max})$.

Now suppose L_{min} is unknown. Assuming a value $L > L_{min}$ for the length of $w_{min}(n)$ in (17), we will result in a sequence $w'_{min}(n)$ which will have the same phase as $w_{min}(n)$, but greater length. Therefore, $w'_{min}(n)$ will be related to $w_{min}(n)$ by a zero-phase sequence, which has zeros in conjugate reciprocal pairs. This, together with (21), imply that the minimum phase parts of the reconstructed channels $h'_1(n)$ and $h'_2(n)$ will have common zeros. The number of common zeros is the difference between L and L_{min}, and can be found as the number of zero eigenvalues of the Sylvester matrix formed based on $i'_{h_1}(n)$ and $i'_{h_2}(n)$, the minimum-phase equivalents of $h'_1(n)$ and $h'_2(n)$ respectively.

5. Simulations

A channel was generated from two delayed raised cosine pulses, as an approximation to a two-ray multipath environment. The channel is given by

$$h(t) = 0.2c(t, 0.11) + 0.4c(t - 2.5, 0.11), \qquad (23)$$

where $c(t, a)$ denotes a raised-cosine pulse and a is the roll-off factor. The length of the channel was taken to be 6 symbols. Two virtual channels $h_1(n)$ and $h_2(n)$ were generated by oversampling $h(t)$ by a factor of two. The source symbols were drawn from a 16 QAM signal constellation with uniform distribution. The noise processes were white, zero-mean and Gaussian distributed.

In the implementation of the algorithm we used 100 input symbols, solving (20) in a least-squares sense. The average least-squares errors over 100 simulations were used to estimate the lengths L_{min}, L_{max} and the parameter N_2. The first substantial drops of the errors ocurred at the correct values, however a length mismatch was found not to cause errors to the estimation procedure. Fig. 1 shows the actual channels and the sample mean of 100 estimates of the channels for SNR 10 dB. Fig. 2 shows a plot of 1000 output symbols of the unequalized channel $h_1(n)$ corrupted by additive noise at SNR 20 dB. The channels were then reconstructed and equalized. 1000 symbols were then transmitted and the equalized channel output is shown in Fig. 3, which indicates that the eye is well opened.

References

[1] M. Gurelli and C.L. Nikias, "EVAM: An Eigenvector-Based Algorithm for Multichannel Blind Deconvolution of Input Colored Signals", *IEEE Trans. on Signal Processing*, vol. 43(1), pp. 134-149, Jan. 1995.

[2] H. Liu, G. Xu, and L. Tong, "A Deterministic Approach to Blind Equalization", *Asilomar Conference on Signals, Systems and Computers*, pp.751-755, Nov. 1993, Pacific Grove, California.

[3] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace Methods for the Blind Identification of Multichannel FIR Filters", *IEEE Transactions on Signal Processing*, vol. 43, pp.516-525, February 1995.

[4] A.P. Petropulu, C.L. Nikias, "Blind deconvolution using reconstruction from partial Higher Order cepstral information", *IEEE Trans. on Signal Processing*, vol. 41(6), pp. 2088-2095, June 1993.

[5] A.P. Petropulu, H. Pozidis, "Multichannel Blind Equalization of Colored Signals", *IEEE Signal Processing, Athos Workshop on Higher-Order Statistics*, pp. 35-38, Girona, Spain, June 1995.

[6] H. Pozidis and A.P. Petropulu, "Cross-correlation Based Blind Equalization," *IEEE Trans. on Signal Processing*, submitted in 1996.

[7] S. Treitel and L.R. Lines, "Linear Inverse Theory and Deconvolution," *Geophysics*, vol. 47(8), pp. 1153-1159, August 1982.

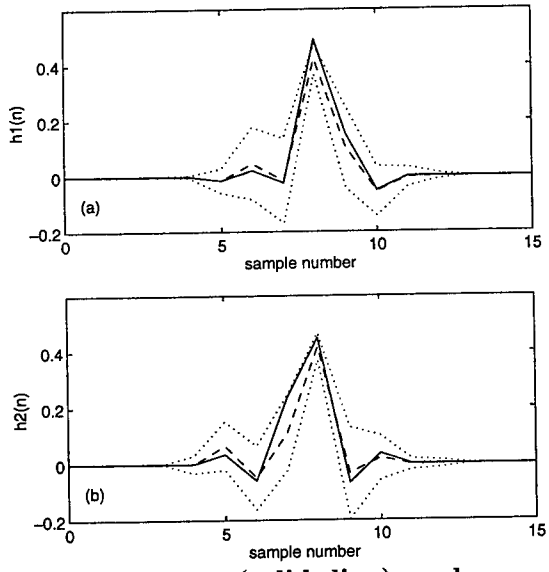


Figure 1. True (solid line) and reconstructed channels at SNR=10 dB. Dashed line indicates sample mean of 100 Monte Carlo runs of the reconstructed channels. Dotted lines indicate standard deviation. 100 output symbols were used in the estimation procedure.

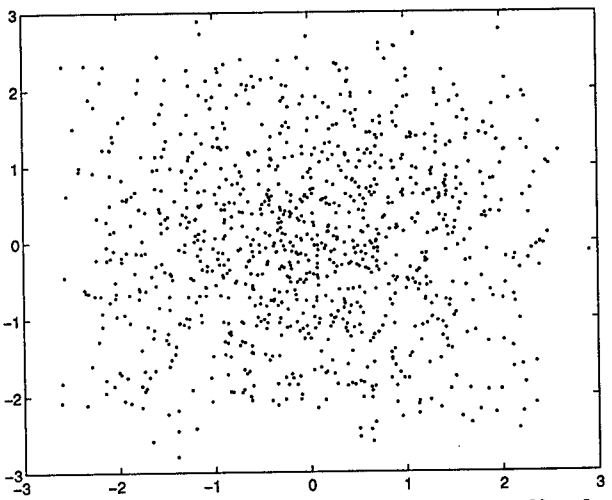


Figure 2. The output of the unequalized channel $h_1(n)$, for SNR=20 dB.

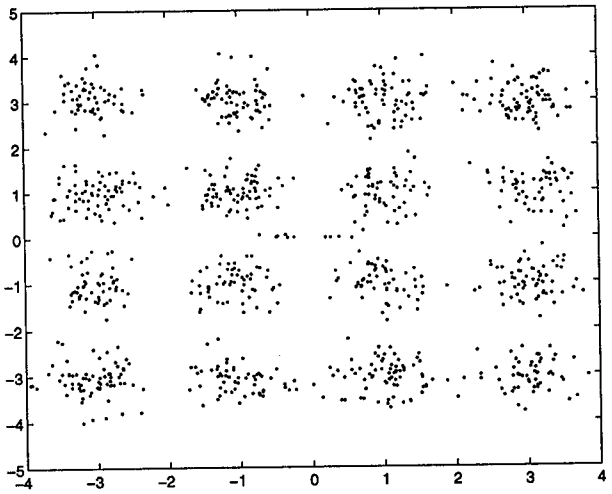


Figure 3. The output of the equalized channel $h_1(n)$. 1000 symbols were plotted. 100 output symbols were used in the estimation, at SNR=20 dB.

Plenary

Nonlinear Systems and Nonstationary Signals: Theory and Applications

T. Subba-Rao, J. Hammond

TP-1: Nonlinear Signal Processing

Nonlinear Autoregressive Exogenous Time Series: Structural Identification Via Projection Estimates

Elias Masry
Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92093 U.S.A.
masry@ece.ucsd.edu

Dag Tjøstheim
Department of Mathematics
University of Bergen
5007 Bergen, NORWAY
dagt@mi.uib.no

Abstract

We consider additive nonlinear Autoregressive Exogenous (ARX) time series and we propose projections as means of identifying and estimating its endogenous and exogenous components. The estimates are nonparametric in nature and involve averaging of kernel type estimates. Such estimates have very recently been treated informally in a univariate time series situation. Here we extend the scope to nonlinear ARX models and present a rigorous theory, including the establishment of consistency and asymptotic normality for the projection estimates.

1. Introduction

Nonlinear time series analysis has received much attention in recent years due primarily to the fact that linear models, such as ARMA, fail to capture many nonlinear features present in commonly encountered time series, as in econometric data. See Tjøstheim [10] for a recent review. Both parametric (Lewis and Stevens [4] and Granger and Terasvitra [3]) and nonparametric models (Chen and Tsay [2]) were introduced in the literature. While a rigorous theory is available in the parametric case (Potscher and Prucha (1991)), much less has so far been achieved for nonparametric models; the work of Chen and Tsay is algorithmic/computational in nature lacking analytical convergence results.

In this paper we consider nonlinear bivariate autoregressive exogenous (ARX) time series modeled by:

$$\begin{aligned} &Y_{t+q+1} = \\ &\quad g_1(Y_{t+1},\ldots,Y_{t+q}) + g_2(X_{t+1},\ldots,X_{t+p}) + e_{t+q+1} \quad (1) \\ &X_{t+p} = g_3(X_{t+1},\ldots,X_{t+p-1}) + \varepsilon_{t+p}. \end{aligned}$$

where $p \le q+1$ and $\{e_t\}$ and $\{\varepsilon_t\}$ are independent series each consisting of zero-mean independent identically distributed variables (iid) with variance σ_e^2 and σ_ε^2, respectively. Under certain regularity conditions, the bivariate process (1) is jointly stationary. The variables $\{X_t\}$ and $\{Y_t\}$ are exogenous and endogenous respectively. The nonlinear ARX model is of fundamental importance in modelling econometric time series. A popular subclass is when both g_1 and g_2 are themselves additive so that, for example,

$$g_1(x_1,\ldots,x_q) = \sum_{i=1}^{q} g_{1,i}(x_i)$$

(see for example Chen and Tsay [2]). We do not limit ourselves to this special case. Our goal is to identify/estimate the functional structure of the time series from the observations $\{Y_i, X_i\}_{i=0}^{n-1}$. We note that

$$\begin{aligned} &E[X_{t+p}|X_{t+1}=x_1,\ldots,X_{t+p-1}=x_{p-1}] = \\ &\quad g_3(x_1,\ldots,x_{p-1}) \end{aligned} \quad (2)$$

and

$$\begin{aligned} &m(x_1,\ldots,x_p;y_1,\ldots,y_q) \triangleq \\ &E[Y_{t+q+1}|X_{t+1}=x_1,\ldots,X_{t+p}=x_p; \\ &\qquad\qquad Y_{t+1}=y_1,\ldots,Y_{t+q}=y_q] \\ &\quad = g_1(y_1,\ldots,y_q) + g_2(x_1,\ldots,x_p). \end{aligned} \quad (3)$$

The function g_3 can be estimated in a straightforward manner by kernel methods and was treated in Masry and Tjøstheim [5]. From (3) it is seen that regression estimation methods could consistently estimate only the *sum* of the functions g_1 and g_2. Our approach to additive modeling in general and to the additive ARX model (1) in particular is through projections. Projections were introduced by Auestad and Tjøstheim in [1][11] for a univariate additive model with the purpose of identifying the functional structure of the components. In this paper we establish the consistency and asymptotic normality of projection estimates for the more general ARX framework. In this way we extend rigorous

0-8186-7576-4/96 $5.00

analysis of estimates in additive models from the independent component case of Stone [9] to the present ARX case, where dependence is an integral part of the system.

We remark that projection estimation theory draws on traditional kernel regression results (cf. Robinson [8]) and on corresponding results for the ARCH (autoregressive conditionally hetroskedastic) model treated in Masry and Tjøstheim [5].

It is possible to generalize the system (1) to a full nonlinear ARX-ARCH system by multiplying e_{t+q+1} and ϵ_{t+p} by nonlinear functions $g_4(Y_{t+1},\ldots,Y_{t+q})$ and $g_5(X_{t+1},\ldots,X_{t+p-1})$, respectively. A univariate ARCH system was treated at length in Masry and Tjøstheim [5]. By combining the results of that paper with the theory of the present one it is possible to construct an ARX-ARCH estimation theory.

2. Projection Estimates

Projection estimates are defined as follows. Let

$$\underline{Y}_t = (Y_{t+1},\ldots,Y_{t+q}) \quad \text{and} \quad \underline{X}_t = (X_{t+1},\ldots,X_{t+p}).$$

The vectors $\underline{x}$ and $\underline{y}$ are defined analogously, and a more general version of (3) can be written as

$$m(\underline{x},\underline{y};p,q) = E\{\phi(Y_{t+q+1}) \mid \underline{X}_t = \underline{x}, \underline{Y}_t = \underline{y}\} \quad (4)$$

where ϕ is measurable on the real line and $E\{|\phi(Y_t)|\} < \infty$. The introduction of ϕ allows us to estimate conditional moments, $\phi(Y) = Y^r$, and conditional distributions, $\phi(Y) = I\{Y \le u\}$. Set

$$h(\underline{x},\underline{y};p,q) = m(\underline{x},\underline{y};p,q) f(\underline{x},\underline{y};p,q) \quad (5)$$

where $f(\underline{x},\underline{y};p,q)$ is the joint density of $(\underline{X}_t,\underline{Y}_t)$, assumed to exist. Let b_n be the bandwidth parameter and set

$$K_n(\underline{u}) = b_n^{-(p+q)} K(\underline{u}/b_n)$$

where $K(\underline{u})$ is a kernel function. Given the observations $\{X_i, Y_i\}_{i=0}^{n-1}$ we estimate f by

$$\hat{f}_n(\underline{x},\underline{y};p,q) = \frac{1}{n_1+1}\sum_{t=0}^{n_1} K_n(\underline{x}-\underline{X}_t, \underline{y}-\underline{Y}_t) \quad (6)$$

and

$$\hat{h}_n(\underline{x},\underline{y};p,q) = \frac{1}{n_1+1}\sum_{t=0}^{n_1} \phi(Y_{t+q+1}) K_n(\underline{x}-\underline{X}_t, \underline{y}-\underline{Y}_t) \quad (7)$$

where $n_1 = n-q-2$ is assumed to be positive. We now estimate $m(\underline{x},\underline{y},p,q)$ by

$$\hat{m}_n(\underline{x},\underline{y};p,q) = \frac{\hat{h}_n(\underline{x},\underline{y};p,q)}{\hat{f}_n(\underline{x},\underline{y};p,q)}. \quad (8)$$

For the ARX model (1) we estimate the sum of the functions g_1 and g_2 as follows:

$$\{g_1(\underline{y}) + g_2(\underline{x})\}\hat{}_n = \hat{m}_n(\underline{x},\underline{y};p,q) \quad (9)$$

where $\hat{m}_n$ is given in (8) with $\phi(x) = x$. We then employ the projection technique to estimate $g_1(\underline{y})$ essentially up to scale and location. Let S_1 be a compact subset of R^p and S_2 be a compact subset of R^q and let $D = S_1 \times S_2$. Put

$$w(\underline{x},\underline{y}) = \begin{cases} 1 & \text{for } (\underline{x},\underline{y}) \in D \\ 0 & \text{for } (\underline{x},\underline{y}) \notin D \end{cases} \quad (10)$$

Then define the projection

$$P_Y(\underline{y}) = E[m(\underline{X}_t,\underline{y};p,q)w(\underline{X}_t,\underline{y})], \quad (11)$$

and, for the ARX system (1), we then have

$$P_Y(\underline{y}) = 1_{S_2}(\underline{y})\left[g_1(\underline{y})Pr\{\underline{X}_0 \in S_1\} + E\{g_2(\underline{X}_0)1_{S_1}(\underline{X}_0)\}\right]$$

Thus the projection identifies $g_1(\underline{y})$ for $\underline{y} \in S_2$ up to a multiplicative and additive constant. Moreover, it is seen that the multiplicative constant will approach 1 when the support D of w is taken to be large enough.

In view of (11), we estimate $P_Y(\underline{y})$ by

$$\hat{P}_Y(\underline{y}) = \frac{1}{n_2+1}\sum_{t=0}^{n_2} w(\underline{X}_t,\underline{y})\hat{m}_n(\underline{X}_t,\underline{y};p,q) \quad (12)$$

where $n_2 = n-p-1$ and $\hat{m}_n(\underline{x},\underline{y};p,q)$ is given by (8). One can similarly estimate the function g_2, related to the projection,

$$P_X(\underline{x}) = 1_{S_1}(\underline{x})\left[g_2(\underline{x})Pr\{\underline{Y}_0 \in S_2\} + E\{g_1(\underline{Y}_0)1_{S_2}(\underline{Y}_0)\}\right]$$

by

$$\hat{P}_X(\underline{x}) = \frac{1}{n_3+1}\sum_{t=0}^{n_3} w(\underline{x},\underline{Y}_t)\hat{m}_n(\underline{x},\underline{Y}_t;p,q) \quad (13)$$

where $n_3 = n-q-1$.

3. Results

We present in this section the main results of the paper without proofs. The full proofs, along with the precise regularity conditions needed, can be found in Masry and Tjøstheim [6]. We first show that the projection estimate $\hat{P}_Y(\underline{y})$ is consistent. We have

Theorem 1. Assume that the functions f and h are Lip_γ for some $0 < \gamma \le 1$ and that the bandwidth parameter

b_n satisfies $nb_n^q \to \infty$ as $n \to \infty$. Then, under certain regularity conditions we have

$$\widehat{P}_Y(\underline{y}) - P_Y(\underline{y}) = O_p\{(nb_n^q)^{-1/2}\} + O_p(b_n^\gamma).$$

The first term on the right-side is the contribution of the variance and the second term is the contribution of the bias. Note that the above rates for the projection estimates coincide with the classical regression rates.

Now define

$$H(\underline{x}, \underline{y}) = w(\underline{x}, \underline{y}) f(\underline{x}\,; p) / f(\underline{x}, \underline{y}\,; p, q)$$

and

$$V^2(\underline{u}, \underline{v}; p, q) = \mathrm{var}\{\phi(Y_{t+q+1} \mid \underline{X}_t = \underline{u}, \underline{Y}_t = \underline{v}\} \,.$$

Also define

$$a^2(\underline{v}, \underline{y}) = \int_{R^p} \left[V^2(\underline{u}, \underline{v}; p, q) + \{m(\underline{u}, \underline{v}\,; p, q) - m(\underline{u}, \underline{y}\,; p, q)\}^2\right] \times H^2(\underline{u}, \underline{y}) f(\underline{u}, \underline{v}\,; p, q) d\underline{u}. \quad (14)$$

Assume that the kernel $K(\underline{u})$ on R^{p+q} is factorable: With $\underline{u} = (\underline{x}, \underline{y})$, let $K(\underline{u}) = K^{(1)}(\underline{x}) K^{(2)}(\underline{y})$. We then have the following result on the asymptotic normality of the projection estimate $\widehat{P}_Y(\underline{y})$.

Theorem 2. Assume that the bandwidth parameter b_n satisfies $nb_n^q \to \infty$ as $n \to \infty$. Then, under certain regularity conditions we have

$$(nb_n^q)^{1/2}\{\widehat{P}_Y(\underline{y}) - P_Y(\underline{y}) - B_n(\underline{x}, \underline{y})\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, a^2(\underline{y}, \underline{y}) \|K^{(2)}\|_2^2) \quad (15)$$

at continuity points of $a^2(\underline{v}, \underline{y})$ as a function of $\underline{v}$.

Remark. The term $B_n(\underline{x}, \underline{y})$ in Theorem 2 represents the "bias" of the projection estimates. When the functions f and h are Lip_γ, as in Theorem 1, then $B_n(\underline{x}, \underline{y}) = O_p(b_n^\gamma)$. It is seen from Theorem 2 that the projection estimate $\widehat{P}_Y(\underline{y})$ is asymptotically normal and a precise expression for the asymptotic variance is given by $a^2(\underline{y}, \underline{y}) \|K^{(2)}\|_2^2$.

4. Example

We carried out a small simulation experiment for the first order system

$$Y_{t+1} = 0.5 Y_t + 0.5 X_{t+1}^2 + e_{t+1} \quad (16)$$

$$X_{t+1} = 0.5 X_t + \varepsilon_{t+1} \quad (17)$$

where $\{e_t\}$ and $\{\varepsilon_t\}$ are generated as independent processes consisting of Gaussian iid random variables with zero mean and variance. The $\{X_t\}$ and $\{Y_t\}$ processes were subsequently adjusted so that they have zero mean (already the case for the $\{X_t\}$-process) and unit variance. The projection estimates (12) and (13) were computed for the scaled processes taking the set $[-3,3] \times [-3,3]$ as the compact set D and using a bandwidth $b_n = n^{-1/5}$. The results clearly reveal the linear dependence on Y_t and the quadratic x-dependence on X_{t+1} in (16).

References

[1] B. Auestad and D. Tjøstheim. Functional identification in nonlinear time series. In *Nonparametric Functional Estimation and Related Topics*. Ed. G.G. Roussas, Amsterdam: Kluwer Academic Publishers, 493-507, 1991.

[2] R. Chen and R. Tsay. Nonlinear additive ARX models. *Journal of the American Statistical Association*, 88: 955-967, 1993.

[3] C.W.J. Granger and T. Teräsvirta. *Modelling Nonlinear Dynamic Relationships* . Oxford: Oxford University Press, 1993.

[4] P.A. Lewis and J.G. Stevens. Nonlinear modeling of time series using multivariate adaptive regression splines (MARS). *Journal of the American Statistical Association*, 86: 864-877, 1991.

[5] E. Masry and D. Tjøstheim. Nonparametric estimation and identification of ARCH nonlinear time series: Strong convergence and asymptotic normality. *Econometric Theory*, 11: 258-289, 1995.

[6] E. Masry and D. Tjøstheim. Additive nonlinear ARX time series and projection estimates. *Econometric Theory*, 1996, to appear.

[7] B.M. Pötscher and I.R. Prucha. Basic structure of the asymptotic theory in dynamic nonlinear econometric models, Part II: Asymptotic normality. *Econometric Reviews*, 10: 253-325, 1991.

[8] P.M. Robinson. Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4: 185-207, 1983.

[9] C.J. Stone. Additive regression and other nonparametric models. *Annals of Statistics*, 13: 689-705, 1985.

[10] D. Tjøstheim. Nonlinear time series: A selected review. *Scandinavian J. of Statistics*, 21: 97-130, 1994.

[11] D. Tjøstheim and B. Auestad. Nonparametric identification of nonlinear time series: Projections. *Journal of the American Statistical Association*, 89: 1398-1409, 1994.

BLIND EQUALIZERS OF MULTICHANNEL LINEAR-QUADRATIC FIR VOLTERRA CHANNELS

Georgios B. Giannakis and Erchin Serpedin
Dept. of Electrical Engr., Univ. of Virginia, Charlottesville, VA 22903-2442, USA

ABSTRACT

Blind equalization of general Volterra models has not been addressed, despite its practical value in communications, acoustics, and physiological modeling. Relying upon diversity (sufficient number of multiple outputs), we establish existence and uniqueness conditions which guarantee that single-input, FIR *nonlinear* Volterra channels can be perfectly but blindly equalized using *linear* FIR equalizers. Apart from a minimal order persistence-of-excitation condition (also present with input-output approaches), the inaccessible input is allowed to be deterministic or random and of unknown color or distribution. With the kernels also satisfying a certain co-primeness condition, we develop direct blind equalizers which by-pass the channel estimation step. Preliminary simulations corroborate our analytical results.

1. INTRODUCTION

Identification of nonlinear systems is of paramount importance in acoustics, physiological modeling [5], magnetic recording [1], satellite and microwave communication links [4]. Using input-output data, methods for identifying FIR Volterra models have been proposed (see e.g., [9]). But apart from special cases, dealing with memoryless nonlinearities and imposing extra conditions on the input [8], [7], the blind scenario has not been addressed. Its practical significance is evident with high-speed (over 5 kb/s) communication channels, especially when no training inputs are available or when new receivers are added in the link and transmission can not be interrupted to initiate a new training session.

In this paper we address the blind equalization and identification of FIR nonlinear Volterra channels by exploiting the temporal and/or spatial diversity offered in the form of multichannel output time series. The latter are collected by oversampling the continuous-time data at a rate faster than the symbol rate and/or by multiple antennas. Diversity is also exploited in [10], [6], [11], [2] for blind identification and equalization of *linear* time-invariant FIR channels, and the present work extends these ideas to the challenging set-up of *nonlinear* Volterra models.

We present our results in the linear-quadratic case (proofs and generalizations to nonlinearities of arbitrary order are reported in [3]). To link temporal with spatial diversity we start with the (baseband) continuous-time Volterra model $x_c(t) = \sum_l s(l)h_{1c}(t-lT)+\sum_{l_1}\sum_{l_2} s(l_1)s(l_2)h_{2c}(t-l_1T, t-l_2T)$, where T is the symbol period. As with the linear case, oversampling by a rate of M/T yields $x(n) := x_c(t)|_{t=nT/M} = \sum_l s(l)h_1(n-lM) + \sum_{l_1}\sum_{l_2} s(l_1)s(l_2)h_2(n-l_1M, n-l_2M)$, where $h_1(n) := h_{1c}(nT/M)$, and similarly for $h_2(n_1,n_2)$. Time series $x(n)$ is cyclostationary with period M. But upon defining the sub-processes $x^{(m)}(n) := x(nM+m-1)$, $m = 1,\ldots,M$, the M-channel process $\mathbf{x}'(n) := [x^{(1)}(n)\ldots x^{(M)}(n)]$, becomes stationary, and for $n = 0,1,\ldots,N-1$ is given by:

$$\mathbf{x}'(n) = \sum_{l=0}^{L_1} \mathbf{h}_1'(l)s(n-l) + \mathbf{v}(n) + \sum_{l_1=0}^{L_2}\sum_{l_2=0}^{l_1} \mathbf{h}_2'(l_1,l_2)s(n-l_1)s(n-l_2)\,, \quad (1)$$

where: (i)prime denotes transpose and lower (upper) bold is used for vectors (matrices); (ii)$M \times 1$ vector $\mathbf{h}_1(\mathbf{h}_2)$ corresponding to the linear (quadratic) kernel is defined similar to $\mathbf{x}$; (iii) the inaccessible scalar input $s(n)$ is allowed to be either deterministic or a realization of a random process (white or colored); (iv)the range of l_2 is chosen so that $\mathbf{h}_2(l_1,l_2)$ is defined over its non-redundant region $0 \le l_2 \le l_1$; (v)$\mathbf{v}(n)$ is AWGN (see also Fig. 1).

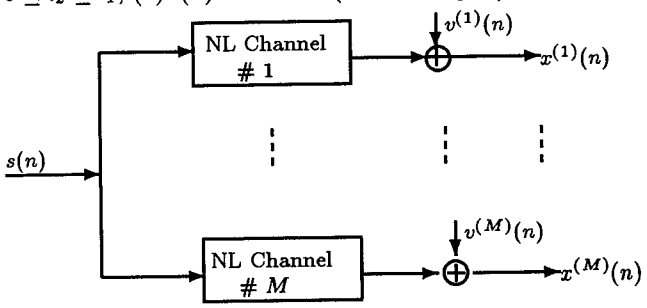


Figure 1. SIMO linear-quadratic model

Given $\{\mathbf{x}(n)\}_{n=0}^{N-1}$ obeying (1), we first transform it to a specific multi-input multi-output (MIMO) linear model, determine orders L_1, L_2, and establish conditions for FIR vector equalizers to exist (Section 2). Blind linear FIR equalizers $\{\mathbf{g}_{1,i}(k)\}_{k=0}^{K}$ of order K satisfying (in the absence of noise) the zero-forcing condition

$$\sum_{k=0}^{K} \mathbf{x}'(n-k)\,\mathbf{g}_{1,i}(k) = s(n-i)\,, \quad (2)$$

are derived in Section 3 and simulated (with noise present) in Section 4. Their uniqueness is established within a shift $i \in [0, L_1+K]$ which is non-identifiable in the blind case.

2. MULTICHANNEL APPROACH

With new variables $i = l_1 - l_2$, $l = l_2$, we view the 2-d kernel $\mathbf{h}_2(l_1,l_2)$ as a collection of L_2+1 linear (1-d) kernels

0-8186-7576-4/96 $5.00

defined as: $\mathbf{h}_{2i}(l) := \mathbf{h}_2(l+i,l)$, $i = 0,1,\ldots,L_2$ and $l = 0,1,\ldots,L_2-i$. Such an interpretation reduces (1) to:

$$\mathbf{x}'(n) = \sum_{l=0}^{L_1} \mathbf{h}_1'(l)s(n-l) + \mathbf{v}(n) + \sum_{i=0}^{L_2}\left[\sum_{l=0}^{L_2-i} \mathbf{h}_{2i}'(l)s(n-i-l)s(n-l)\right], \quad (3)$$

and casts the SIMO problem into a MIMO one, but with special L_2+2 inputs $s_1(n) = s(n)$, $s_{20}(n) = s^2(n)$, $s_{21}(n) = s(n-1)s(n)$, $\ldots$, $s_{2L_2}(n) = s(n-L_2)s(n)$.

Let us now define the $(L_1+K+1)\times M(K+1)$ block Toeplitz matrix associated with $\mathbf{h}_1'(l)$

$$\mathbf{H}_1 := \begin{bmatrix} \mathbf{h}_1'(0) & \ldots & \mathbf{0}' \\ \vdots & \ddots & \vdots \\ \mathbf{h}_1'(L_1) & \ldots & \mathbf{h}_1'(L_1-K) \\ \vdots & \ddots & \vdots \\ \mathbf{0}' & \ldots & \mathbf{h}_1(L_1) \end{bmatrix}.$$

Similarly, for each $i \in [0, L_2]$, we denote the $(L_2+K+1-i)\times M(K+1)$ matrix corresponding to $\mathbf{h}_{2i}'(l)$ as $\mathbf{H}_{2i}$. We also define the $1\times(L_1+K+1)$ input vector $\mathbf{s}_{10}'(n) := [s(n); s(n-1); \ldots; s(n-K-L_1)]$ and respectively for each $i \in [0, L_2]$ the $1\times(L_2+K+1-i)$ vector $\mathbf{s}_{2i}'(n) := [s(n-i)s(n); \ldots; s(n-K-L_2)s(n-K-L_2+i)]$. With these definitions, we obtain the matrix version of (3), $\mathbf{X} = \mathbf{SH}$, where the $(N-K)\times M(K+1)$ block Hankel data matrix $\mathbf{X}$ is formed as

$$\mathbf{X} := \begin{bmatrix} \mathbf{x}'(N-1) & \ldots & \mathbf{x}'(N-1-K) \\ \vdots & \vdots & \vdots \\ \mathbf{x}'(K) & \ldots & \mathbf{x}'(0) \end{bmatrix}, \quad (4)$$

and $(N-K)\times d(L_1,L_2,K)$ input matrix $\mathbf{S}$ and $d(L_1,L_2,K)\times M(K+1)$ system matrix $\mathbf{H}$ are given by

$$\begin{bmatrix} \mathbf{s}_{10}'(N-1) & \ldots & \mathbf{s}_{2L_2}'(N-1-K) \\ \mathbf{s}_{10}'(N-2) & \ldots & \mathbf{s}_{2L_2}'(N-2-K) \\ \vdots & \vdots & \vdots \\ \mathbf{s}_{10}'(K) & \ldots & \mathbf{s}_{2L_2}'(0) \end{bmatrix}, \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_{20} \\ \vdots \\ \mathbf{H}_{2L_2} \end{bmatrix}.$$

The common dimension of $\mathbf{S}$ and $\mathbf{H}$ is $d(L_1,L_2,K) = L_1+K+1+\sum_{i=0}^{L_2}(L_2+K+1-i)$, or,

$$d(L_1,L_2,K) = (L_2+2)(K+1) + (L_2^2+L_2+2L_1)/2\,. \quad (5)$$

We adopt the following assumptions:
(a1) $N-K \geq \max(L_1,L_2)+K+1$, which is easily met in practice by collecting sufficient data;
(a2)input $s(n)$ is persistently exciting (p.e.) of order $\rho_s = d(L_1,L_2,K)$; i.e., $\mathrm{rank}(\mathbf{S}) = d(L_1,L_2,K)$. White noise is p.e. of any order, but $d(L_1,L_2,K)$ modes in the spectrum of $s(n)$ may not guarantee p.e. as in the linear case; $s(n)$ must also have sufficient amplitude levels (note that if e.g., $s(n) = 0,1$, $\mathbf{S}$ is rank deficient because $s(n) = s^2(n)$);
(a3)quadruplet (M,K,L_1,L_2) obeys

$$M(K+1) \geq d(L_1,L_2,K)\,, \quad (6)$$

which for a given M and (L_1,L_2) is satisfied by choosing $K \geq \lceil (L_2^2+L_2+2L_1)/2(M-L_2-2)\rceil - 1$, where $\lceil\alpha\rceil$ denotes the smallest integer $\geq |\alpha|$. The minimum number of channels required is thus $M_{min} = L_2+3$ which depends on the memory of the nonlinearity; e.g., memoryless nonlinearities (often encountered with satellite links) correspond to $i=0$ in (3) and require $M_{min} = 3$ channels and a minimum equalizer order $K_{min} = L_1+L_2-1$ (recall that for linear channels $M_{min} = 2$ and $K_{min} = L_1-1$) [10], [11];
(a4)matrix $\mathbf{H}$ has full row rank; i.e., $\mathrm{rank}(\mathbf{H}) = d(L_1,L_2,K)$, which implies that there are no common zeros among the 1-d kernel transfer functions $\{H_1^{(m)}(z), H_{20}^{(m)}(z), H_{21}^{(m)}(z), \ldots, H_{2L_2}^{(m)}(z)\}$ across all M channels.

We stress that apart from p.e. (also needed for input-output methods) no extra assumptions are imposed on $s(n)$. Matrix $\mathbf{H}$ must be at least fat (square if equality holds in (6)), which along with (a4) expresses the need for diversity (sufficient number of "sufficiently different" channels).

2.1. Order determination

Assume noise-free data and equality in (6) to obtain, [3]:
Lemma 1: *Under (a1)-(a4), matrix* $\mathbf{X}$ *in (4) has* $\mathrm{rank}(\mathbf{X}) = d(L_1,L_2,K)$. □
Let $(\bar{L}_1,\bar{L}_2)$ be known upper bounds on (L_1,L_2). With M given, choose two distinct orders $(\bar{K}_1,\bar{K}_2)$ both $\geq \bar{K}_{min} := \lceil(\bar{L}_2^2+\bar{L}_2+2\bar{L}_1)/2(M-\bar{L}_2-2)\rceil - 1$, and form matrices $\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2$ as in (4). It follows from Lemma 1 that $\rho_i := \mathrm{rank}(\bar{\mathbf{X}}_i) = d(L_1,L_2,\bar{K}_i)$, $i=1,2$. Knowing $(\bar{K}_1,\bar{K}_2)$, evaluating ρ_1,ρ_2 (using SVD), and using (5), we establish:
Corrolary 1: *For the model in (1), the orders* L_1, L_2 *can be found as* $L_2 = (\rho_1-\rho_2)/(\bar{K}_1-\bar{K}_2) - 2$, *and* $L_1 = \rho_1 - (L_2+2)(\bar{K}_1+1) - (L_2^2+L_2)/2$. □
From now on we assume L_1, L_2 known and choose K to satisfy (6) for a given M.

2.2. Existence and uniqueness

Consider (2) with $n = N-1,\ldots,N-K$, and define $\mathbf{g}_{1,i}' := [\mathbf{g}_{1,i}'(0)\ldots\mathbf{g}_{1,i}'(K)]$ to obtain the matrix-forms

$$\mathbf{X}\,\mathbf{g}_{1,i} = \mathbf{s}_i \;\Leftrightarrow\; \mathbf{S}\,\mathbf{H}\,\mathbf{g}_{1,i} = \mathbf{s}_i\,, \quad (7)$$

where $\mathbf{s}_i$ denotes the $(i+1)$st column of $\mathbf{S}$. But (7) holds iff: $\mathbf{H}\mathbf{g}_{1,i} = \mathbf{e}_i$ where $\mathbf{e}_i$ is a $d(L_1,L_2,K)\times 1$ vector with unity in its $(i+1)$st entry and zero elsewhere. Given $\mathbf{H}$ and a fixed shift $i \in [0, L_1+K]$, we prove that [3]:
Theorem 1: *Under (a3) and (a4), a linear FIR equalizer* $\mathbf{g}_{1,i} = \mathbf{H}^\dagger\mathbf{e}_i$, $i \in [0, L_1+K]$, *exists. It is unique if (6) holds as equality, or, if the pseudoinverse* $\mathbf{H}^\dagger$ *is adopted when (6) holds as strict inequality (minimum-norm solution).* □
The vital role of diversity offered by multichannel data is transparent if we note that (6) is not satisfied with $M=1$; i.e., single channel linear FIR equalizers of FIR Volterra channels are impossible. With $M_{min} = L_2+3$, (6) is satisfied as an equality if we choose $K_{min} = (L_2^2+L_2+2L_1-2)/2$. On the other hand, $\mathbf{g}_{1,i}^{min-norm}$ may be preferable for suppressing AWN as in the linear channel case [2].

Although the number of antennas (and thus complexity) increases in the nonlinear case, our ability to equalize nonlinear channels with linear FIR equalizers is very appealing, especially because stability of inverse Volterra systems is difficult to define and even more difficult to check.

3. DIRECT BLIND EQUALIZERS

To solve (7) in the blind case we must eliminate the input dependence. Consider (2) with $i=0$ to find $\sum_{k=0}^{K}\mathbf{x}'(n-k)\mathbf{g}_{1,0}(k) = s(n)$, and also substitute $n+i \leftarrow n$ in (2) to obtain $\sum_{k=0}^{K}\mathbf{x}'(n+i-k)\mathbf{g}_{1,i}(k) = s(n)$. Eliminating $s(n)$

from these equations, we arrive at the cross-relation:

$$\sum_{k=0}^{K} \mathbf{x}'(n-k)\,\mathbf{g}_{1,0}(k) \;=\; \sum_{k=0}^{K} \mathbf{x}'(n+i-k)\,\mathbf{g}_{1,i}(k)\,,$$

which forms the basis of our blind approach. In matrix form we use Matlab's notation $\mathbf{X}(i_1 : i_2, :)$ to denote a submatrix of $\mathbf{X}$ formed by the i_1 through i_2 rows and all columns of $\mathbf{X}$. For the $\mathbf{s}_i$ vector in (7), it holds that $\mathbf{s}_0(i+1 : N-K, :) = \mathbf{s}_i(1 : N-K-i, :)$, which implies that the left-hand sides of (7) for different i's are related. Upon defining

$$\mathbf{X}_i := \mathbf{X}(i : N-K, :)\;,\;\; \mathbf{X}_{0,i} = \mathbf{X}(1 : N-K-i, :)\,,$$

we infer that $\mathbf{X}_i\mathbf{g}_{1,0} = \mathbf{X}_{0,i}\mathbf{g}_{1,i}$; i.e., for $i = 1, \ldots, L_1 + K$,

$$\mathcal{X}_{0,i}\mathbf{g}_{1,0i} := [\mathbf{X}_{0,i} - \mathbf{X}_i] \begin{bmatrix} \mathbf{g}_{1,0} \\ \mathbf{g}_{1,i} \end{bmatrix} = \mathbf{0}\,. \qquad (8)$$

The pair of equalizers $(\mathbf{g}_{1,0}, \mathbf{g}_{1,i})$ will be identifiable (up to a scale) as the eigenvector corresponding to the minimum eigenvalue of $\mathcal{X}_{0,i}$ iff the nullity $\nu(\mathcal{X}_{0,i}) = 1$. It is also possible to collect all pairs of shifts $(0,1), (0,2), \ldots, (0, L_1 + K)$ and recover simultaneously equalizers corresponding to all shifts by solving $\mathcal{X}_{0:L_1+K}\boldsymbol{g}_1 = \mathbf{0}$; i.e.,

$$\begin{bmatrix} \mathbf{X}_1 & -\mathbf{X}_{0,1} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_{L_1+K} & \mathbf{0} & \ldots & -\mathbf{X}_{0,L_1+K} \end{bmatrix} \begin{bmatrix} \mathbf{g}_{1,0} \\ \vdots \\ \mathbf{g}_{1,L_1+K} \end{bmatrix} = \mathbf{0} \qquad (9)$$

where $\mathcal{X}_{0:L_1+K}$ has dimensions $\sum_{i=1}^{L_1+K+1}(N-K-i)\times(L_1+K+1)M(K+1)$. With regards to the ranks of $\mathcal{X}_{0,L_1+K}$ and $\mathcal{X}_{0:L_1+K}$ in (8) and (9) we prove [3]:

Theorem 2: *Suppose* $\mathbf{x}(n)$ *comes from (1) with* $\mathbf{v}(n) \equiv \mathbf{0}$, *(a1), (a3), (a4) are satisfied, and (6) holds as equality. If* $L_1 > L_2$ *and p.e. order* $\rho_s \geq 2d(L_1, L_2, K) + 1$, *then* $\nu(\mathcal{X}_{0,L_1+K}) = 1$ *and the minimum and maximum delay equalizers are identifiable from (8). If* $L_1 > L_2$ *and (a2) holds, then* $\nu(\mathcal{X}_{0:L_1+K}) = 1$ *and all equalizers in* $\boldsymbol{g}_1$ *are identifiable from (9).* □

Note that (8) involves a smaller matrix than (9), but also requires stronger p.e. conditions. Among all $(0, i)$ pairs only the $(0, L_1 + K)$ pair of equalizers can be identified alone. Two questions arise at this point: when does $L_1 > L_2$ hold in practice? and what if $L_1 = L_2$? Condition $L_1 > L_2$ requires memory domination of the linear part which is expected in most practical cases. Also, in magnetic recording applications we have $L_1 = L_2 = L$ =2 or 3, but $s(n) = 0, 1$; hence, $s(n) = s^2(n)$ which allows us to combine the quadratic kernel $\mathbf{h}_{20}(l)$ (of order L) with the linear one $\mathbf{h}_1(l)$, leaving the remaining kernels $\mathbf{h}_{2i}(l)$ with orders $L-i$, $i \in [1, L]$. In this case too, Theorem 2 applies because there is a single kernel attaining the maximum order L.

If $L_1 = L_2$, then it turns out that $\nu(\mathcal{X}_{0:L_1+K}) = 2$, in which case supervector $\boldsymbol{g}_1$ in (9) is a linear combination of the null eigenvectors $\mathbf{u}_1, \mathbf{u}_2$: $\boldsymbol{g}_1 = \lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2$. To determine the λ_1, λ_2 constants we take advantage of "quadratic" equalizers such as $\mathbf{g}_{20,i}$ whose output satisfies:

$$\sum_{k=0}^{K} \mathbf{x}'(n-k)\,\mathbf{g}_{20,i}(k) = s^2(n-i)\,. \qquad (10)$$

It turns out that similar to $\boldsymbol{g}_1$, supervector $\boldsymbol{g}_{20}$ also satisfies (9) and hence, $\boldsymbol{g}_{20} = \mu_1\mathbf{u}_1 + \mu_2\mathbf{u}_2$. Eliminating $s(n-i)$ from (2) and (10), a cross-relation between $\boldsymbol{g}_1$ and $\boldsymbol{g}_{20}$ results which allows determination of the $(\lambda_1, \lambda_2, \mu_1, \mu_2)$ coefficients [3].

If the linear kernel is absent (homogeneous model) we again find $\nu(\mathcal{X}_{0:L_1+K}) = 1$, and the $\boldsymbol{g}_{20}$ equalizer is uniquely identifiable although of limited value since its output $s^2(n)$ is only sufficient when sign ambiguity is not a problem (e.g., transmission with non-negative signals).

Our conclusions on the nullity of $\mathcal{X}_{0:L_1+K}$ carry over to nonlinearities of arbitrary order although some cases are easier than others (e.g., odd order nonlinearities only) [3]. Equalizing nonlinear channels with linear deconvolvers is neat and can be justified intuitively if one views the vector equalizer as a beamformer which, thanks to its diversity, is capable of nulling the nonlinearities and equalizing the linear part. With equalizers corresponding to all possible shifts, we can align their outputs and average in order to estimate the input via (c.f.(2)):

$$s(n) = \frac{1}{L_1+K+1} \sum_{i=0}^{L_1+K} \left[\sum_{k=0}^{K} \mathbf{x}'(n+i-k)\mathbf{g}_{1,i}(k) \right].$$

If blind equalization is the goal, direct equalizers offer advantages over indirect approaches which estimate first the channel $\mathbf{H}$ and then invert it to obtain the equalizer. When the noise spectrum is known, the Wiener inverse trades off perfect (or zero-forcing) equalization with SNR improvement. Similarly, SNR gain is obtained if (9) is solved using (weighted or total) least-squares depending on whether $\mathbf{v}(n)$ is colored or white.

If blind channel identification is the objective, the estimated equalizers can be used to recover $s(n)$ from which $\mathbf{H}$ can be obtained by solving (3) using (batch or recursive) linear regression methods. The linear forms of (8), and (9) should lend themselves to adaptive schemes which are currently under investigation together with linear prediction formulations and methods needed to select the optimum shift.

4. SIMULATIONS

Example 1: We generated 2-level PAM i.i.d. data ($s(n) = 0, 1$) and passed them through $M = 3$ FIR channels ($m = 1, 2, 3$) to obtain the data:

$$x^{(m)}(n) = \sum_{l=0}^{2} h_1^{(m)}(l)s(n-l) + \sum_{l=0}^{1} h_{21}^{(m)}(l)s(n-1-l)s(n-l).$$

The impulse response vectors were $\mathbf{h}_1(0) = [1, 0.5, 2]'$, $\mathbf{h}_1(1) = [-2.5, 3, 0]'$, $\mathbf{h}_1(2) = [1, 5, 2]'$, $\mathbf{h}_{21}(0) = [2, 0.3, -0.7]'$, $\mathbf{h}_{21}(1) = [0.7, 1.2, 3]$. Such a channel has form similar to that used in magnetic recording models [1]. Theorem 2 applies to this channel ($L_1 = 2 > L_{21} = 1$) and using one SVD we computed the 3×1 vector equalizer of order $K = 2$ by solving (8) with $i = L_1 + K$. Fig. 2a depicts root mean-square error (rmse) between the true and estimated equalizer coefficients for lengths $N = 100, \ldots, 900$ at SNR 20dB and 40dB; rmse vs. SNR is shown in Fig. 2b for $N = 50, 100$ (averages were computed based on 100 Monte Carlo runs). Interestingly, with as little as $N = 200$ symbols, it is possible to equalize linear-quadratic channels with rmse=$O(10^{-2})$ at SNR=20dB. A typical eye-diagram of one channel's output is plotted in Fig. 3a along with its equalized version in Fig. 3b.

Example 2: A similar simulation was carried with 4-level PAM data ($s(n) = \pm 3, \pm 1$) and $M = 4$ channel outputs were generated according to the model ($m = 1, 2, 3, 4$):

$$\begin{aligned} x^{(m)}(n) \;&=\; \sum_{l=0}^{1} h_1^{(m)}(l)s(n-l) + \sum_{l=0}^{1} h_{20}^{(m)}(l)s^2(n-l) \\ &+\; h_{21}^{(m)}(0)s(n-1)s(n) + v^{(m)}(n)\,, \end{aligned}$$

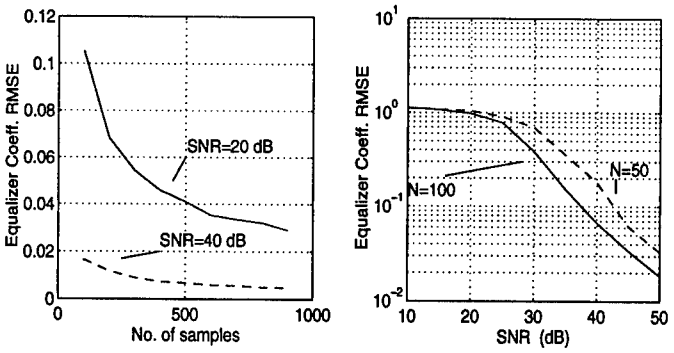


Figure 2. RMSE curves for Example 1

with $\mathbf{h}_1(0) = [1, 0.5, 2, 0.1]'$, $\mathbf{h}_1(1) = [-2.5, 3, 0, -1.1]'$, $\mathbf{h}_{20}(0) = [0.01, 0.5, 0.2, 0.03]'$, $\mathbf{h}_{20}(1) = [0.2, 0.3, -0.7, -0.001]$, $\mathbf{h}_{21}(0) = [0.007, 0.001, 0.3, -0.15]$. Figs. 4 and 5 show that about an order of magnitude more data are required to achieve performance similar to that in Figs. 2 and 3, a consequence of the fact that two SVDs are required for this model [3] (note that here $L_1 = L_{20} = 1$, $L_{21} = 0$, $K = 1$). To illustrate the importance of incorporating nonlinearities over adopting linear approximations, we supposed that the data come from a linear channel of order $L = 3$, and using $M = 2$ outputs we designed an order $K = 2$ linear equalizer by inverting the channel estimate of [11]. The equalized eye-patterns for the 2- and 4-level PAM data are shown in Fig. 6. The importance of adopting the correct model is evident if one compares Figs. 3 and 5 with Fig. 6.

REFERENCES

[1] E. Biglieri, E. Chiaberto, G.P. Maccone, E. Viterbo, "Compensation of nonlinearities in high-density magnetic recording channels," *IEEE Tr. on Magnetics*, p. 5079, 1994.

[2] G. B. Giannakis and S. Halford, "Blind fractionally-spaced equalization of noisy FIR channels: adaptive and optimal solutions," *Proc. of Intl. Conf. on ASSP*, vol. 3, pp. 1972-1975, Detroit, MI, May 8-12, 1995.

[3] G. B. Giannakis and E. Serpedin, "Linear multichannel blind equalizers of nonlinear FIR Volterra channels," *IEEE Trans. on Signal Proc.*, 1996 (submitted).

[4] G. Lazzarin, S. Pupolin, and A. Sarti, "Nonlinearity compensation in digital radio systems," *IEEE Trans. on Communications*, pp. 988-998, 1994.

[5] P. Z. Marmarelis and V. Z. Marmarelis, *Analysis of Physiological Systems*, Plenum, New York, 1978.

[6] E. Moulines, P. Duhamel, J. Cardoso, S. Mayrargue, "Subspace Methods for the Blind Identification of Multichannel FIR Filters", *IEEE Trans. on SP*, p. 516, 1995.

[7] S. Prakriya and D. Hatzinakos, "Blind identification of nonlinear channels with higher-order cyclic spectral analysis," *Proc. of Intl. Conf. on HOS*, p. 366, Spain, 1995.

[8] N. Rozario and A. Papoulis, "The identification of certain nonlinear systems by only observing the output," *Proc. of HOSA Wrksp*, p.78, Vail, CO, 1989.

[9] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, New York: Wiley, 1980.

[10] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE-IT*, p. 340, 1994.

[11] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. on Signal Proc.*, pp. 2982-2993, 1995.

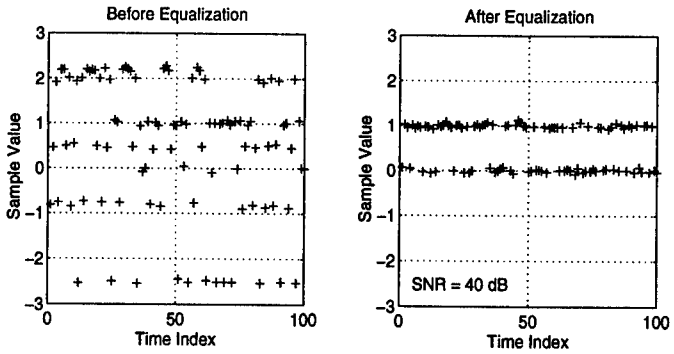


Figure 3. Eye-patterns for Example 1

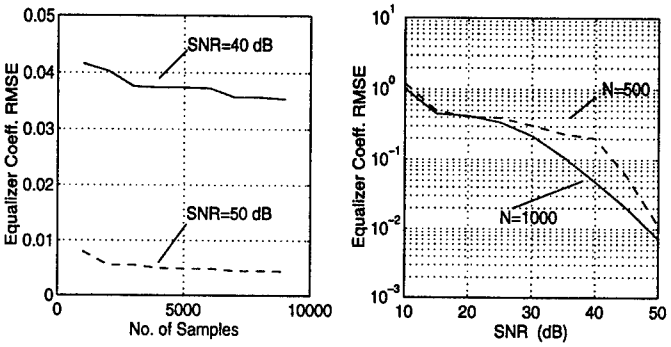


Figure 4. RMSE curves for Example 2

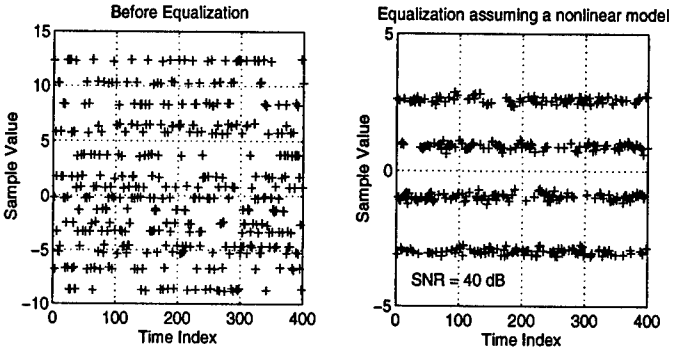


Figure 5. Eye-patterns for Example 2

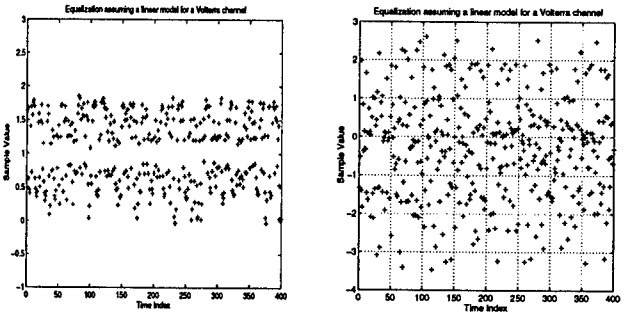


Figure 6. Eye-patterns – linear approximations

Memoryless Predistortion of Nonlinear Amplifiers Based on Fourier Series Based Models

A. Pagès-Zamora, Miguel A. Lagunas, Tomás Jiménez

Dpt.Teoria Senyal i Comunicacions, ETSE Telecom., Universitat Politècnica de Catalunya (UPC)
Campus Nord UPC, Ed. D-5. c/ Gran Capità, s/n. 08034 Barcelona, Spain.
e-mail: alba@gps.tsc.upc.es

Abstract*

In order to maximise the efficiency of the RF amplifier located in a transmitter, for instance in both analog and digital terrestrial TV links, it is forced to work near saturation introducing thus an undesirable nonlinear effect. A common solution includes a predistortion system before the modulation that compensates as much as possible the posterior nonlinear distortion, in such a way that the overall performance of the transmitter results in a linear and efficient amplifier. Polynomial models usually implement the predistortion, but in this paper we propose an alternative model based on the Fourier-exponential series that shows better performance in the design stage without a significant increase of the complexity.

1. Introduction

The distortion introduced by a High Power Amplifier (HPA) located in the transmitter of a terrestrial link is usually equalised by the so-called feed-forward method, the negative feedback method or the predistortion method. The first one has a cost limitation since it needs two HPAs, which are quite expensive elements of the RF link. The so-called negative feedback method is another RF or intermediate frequency solution, which has an inherent instability problem. On the contrary, the predistortion method has no loop (avoiding thus any instability) and it results in a cheap solution since it can be implemented at the baseband level by a DSP. Nevertheless, in order to be able to apply the predistortion at the symbol level, several aspects should be taken into account. First of all, the fact that the HPA is located in the transmitter ensures the introduced nonlinear distortion is memoryless. This property, along with the bandpass behaviour of the HPA [1], allows a lowpass equivalent formulation where the HPA is completely characterised by the so-called AM/AM and AM/PM curves which relates the input amplitude to the output amplitude and output phase, respectively. These curves are supposed to be independent of the frequency for narrow bandpass signals and they can be obtained by measuring the output of the HPA when it is driven with a pure tone of the carrier frequency. The AM/AM and AM/PM relations are needed for the design stage, where the parameters of the predistortion are set. In general, a memoryless Volterra system is chosen to model the predistortion, being its coefficients fitted by an adaptive learning that is applied periodically before the data transmission due to the slow-time variation in the HPA characteristics.

In this paper, the authors propose an alternative system to model the predistortion that shows better performance in adaptive designs than the Volterra model does. Section 2 is thus devoted to present this model, which is based on a Fourier series development. In Section 3, the particular HPA predistortion problem is focused emerging the role and design of the memoryless nonlinear models. Finally, the simulation results are included in Section 4 where the performance of the Fourier versus the Volterra model in this particular topic is compared.

2. The Fourier model

The Fourier model arises from the Fourier series development of the input/output relation of the actual nonlinear system (NLS). If g[x] denotes the relation of a given memoryless NLS, and x is the input, the approximation of a N-order Fourier is the following one.

$$g[x] \approx \hat{g}[x] = \sum_{n=-N}^{N} c_n \cdot e^{jn\omega_o x} \qquad (1)$$

It is important to remark that, in order to avoid aliasing in the approximation provided by the Fourier model, the input signal x should be bounded, i.e. $x \in [-X_{max}, X_{max}]$, being also the principal frequency upper-bounded.

$$\omega_0 = \frac{2\pi}{2X_0} \leq \frac{2\pi}{2X_{max}} \qquad (2)$$

Some previous works about the Fourier model have been already done, even with nonlinear problems with memory [2,3]. Without going into details, an important feature of this model is the fact that, once the order N and the principal frequency ω_0 are chosen, the model is linear with the rest of coefficients, $\{c_n\}$. This property allows a MMSE criterion for designing the coefficients $\{c_n\}$ and,

*This work was supported by PRONTIC/CICYT TIC-95-1022-C05-01 and CIRIT/"Generalitat de Catalunya" GRQ93-3021

0-8186-7576-4/96 $5.00 © 1996 IEEE

moreover, the use of classical adaptive methods to lead the model to this solution can be also applied [4].

In the HPA predistortion problem, the simplified model versions that consider an even or odd symmetry in the NLS input/output relation are specially interesting. Thus, the Fourier model allows a simplification when g[x] has an odd (eq.3.a) or an even (eq.3.b) symmetry. The complexity is considerably reduced in comparison with the general Fourier model (eq.1) due to the real character of the coefficients and the functions.

$$If\ g[x]=-g[-x] \ \Rightarrow\ \hat{g}(x)=\sum_{n=1}^{N} a_n \cdot \sin(n\,\omega_0\,x) \qquad (3.a)$$

$$If\ g[x]=g[-x] \ \Rightarrow\ \hat{g}(x)=\sum_{n=0}^{N} b_n \cdot \cos(n\,\omega_0\,x) \qquad (3.b)$$

Before dealing with the HPA predistortion problem, it is interesting to compare the Fourier model and the Volterra model in terms of complexity. Thus, an N-order Volterra model needs of the order of O(N) multiplications to provide the successive powers of the input signal, whereas an N-order Fourier model needs O(4N) real multiplications to compute the successive powers of the first order complex exponential, exp(jω0x). In fact, the memoryless Fourier model can be basically viewed as an N-order Volterra model preceded by an exponential transformation (fig.1) and, in consequence, both models involve the same order of operations to generate the respective input data space (apart from the cost of computing the first complex exponential function).

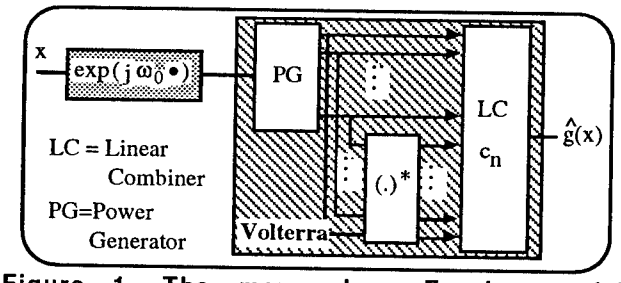


Figure 1. The memoryless Fourier model implementation

In case of dealing with the symmetric models, the Volterra model uses the half of the operations, whereas the Fourier model (eq.3.a,3.b) needs the same because the cosine/sine functions are obtained as the real/imaginary parts of the respective exponential functions. Concerning the number of coefficients that determine the computation load in the adaptive design, the Volterra model has N coefficients and the Fourier model has (2N+1) complex ones. Nevertheless, this number for the symmetric Volterra model is N/2, and for the Fourier Cosine or Sine models becomes N real coefficients.

3. The HPA predistortion problem

Let's consider a digital link with a transmitter as the one shown in figure (2). In a practical situation, the predistortion is located before the modulation, being usually designed by means of an adaptive method applied previously to the data transmission.

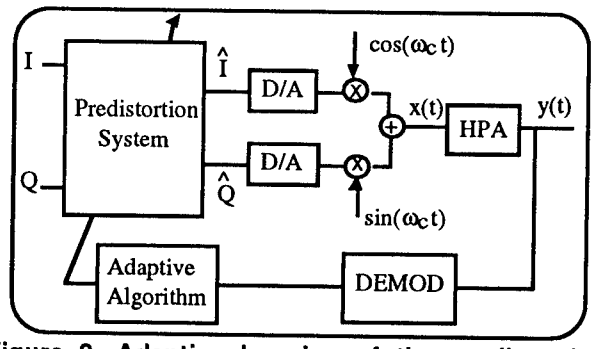


Figure 2. Adaptive learning of the predistortion system.

The input to the HPA is denoted by x(t) and results in a narrow bandpass signal, centred round the frequency ω_c with an instantaneous amplitude and phase represented by, R_t and θ_t, respectively. Thus, the output of the HPA can be approximated by the following bandpass signal,

$$y(t)=HPA[x(t)]=F[R_t]\cdot\cos(\omega_c t+\theta_t+\phi[R_t]) \qquad (4)$$

which involves the functions F[R] and φ[R] that represent the so-called AM/AM and AM/PM relations of the HPA. In the simulations, these functions approximate the actual AM/AM and AM/PM curves proposed in [5]. Whereas the amplitude distortion (fig.5.a) follows the relation,

$$F[R]=sign(R)*0.62*\left(1-\exp\left(-R^2/0.25\right)\right) \qquad (5)$$

the phase predistortion φ[R] (fig.5.b) is implemented by an even polynomial with 9 coefficients.

The actual HPA output (eq.4) makes evident the fact that these curves completely characterise the HPA and, moreover, that they can be seen as a lowpass equivalent transformations. Thus, as it is shown in figure (3) for a discrete system, the predistortion design allows a lowpass formulation useful not only to find the design equations, but also for the simulations since we have not available real data.

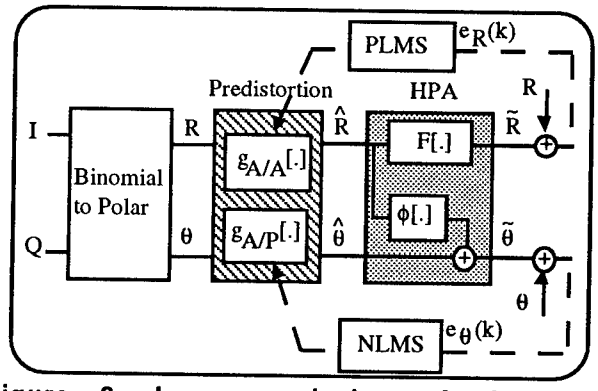


Figure 3. Lowpass design of the HPA predistortion.

Concerning the amplitude predistortion denoted by $g_{A/A}[.]$, it can be designed in order to minimise the amplitude error $e_R(k)$, defined as the difference between the HPA AM/AM output and the desired magnitude R_k (note that the HPA amplification is an scale factor, not included in the predistortion design).

$$e_R(k) = R_k - F[\hat{R}_k] = R_k - F\left[g_{A/A}\left[R_k\right]\right] \approx 0 \quad (6)$$

Thus, ideally, the relation $g_{A/A}[R]$ should be the inverse transformation of F[R]. With respect to the phase predistortion denoted by the function $g_{A/P}[.]$, note that the minimisation of the phase error,

$$e_\theta(k) = \theta_k - \hat{\theta}_k + \phi\left[\hat{R}_k\right] \approx 0 \Rightarrow g_{A/P}\left[\hat{R}_k\right] \approx -\phi\left[\hat{R}_k\right] \quad (7)$$

conveys to a basic identification problem since the phase predistortion is applied after the amplitude predistortion.

It is important to remark that, although some noise will be present at the output of the actual HPA, the predistortion design equations derived from expressions (6) and (7) are useful since the main goal of the predistortion is to compensate nonlinearities, without taking care of the noise. In the simulations, an additive Gaussian noise at the output of the HPA curves will be considered with a high SNR (as it usually happens in terrestrial RF transmission) that allows a good performance of the proposed predistortion design.

3.1. The Predistortion Models

The odd symmetry of the AM/AM curve determines the model that implements the amplitude predistortion. In fact, the function $g_{A/A}[.]$ should also follow an odd relation and two possible models arise from the respective Volterra or Fourier series.

$$g^V_{A/A}\left[R_k\right] = \sum_{n=1}^{N} a^V_k(n) \cdot R_k^{(2k-1)} = \left(a^V_k\right)' \cdot u^V_k \quad (8.a)$$

$$g^{FS}_{A/A}\left[R_k\right] = \sum_{n=1}^{N} a^{FS}_k(n) \cdot \sin\left(n\omega_0 R_k\right) = \left(a^{FS}_k\right)' \cdot u^{FS}_k \quad (8.b)$$

The linear dependence of both models allows a vector notation in terms of the coefficient vector $\mathbf{a}_k$ and the data vector $\mathbf{u}_k$, which gathers the power functions for the Volterra model and the sine functions for the Fourier-Sine model. The coefficients in both models are time dependent since they are modified during the learning stage in order to minimise the power of the amplitude error. In [5], a kind of gradient adaptive algorithm is considered to update the value of the coefficients in the opposite direction of the instantaneous gradient. The resulting adaptive algorithm is called predistortion LMS (PLMS) algorithm and it involves the gradient of the AM/AM characteristic of the HPA with respect to the input value.

$$a(k+1) = a(k) + \mu\, e_R(k) \left.\frac{\partial F(x)}{\partial x}\right|_{x=\hat{R}_k} u(k) \quad (9)$$

In the simulations, the exact gradient of the proposed function F[.] has been used although in a real situation it can be also estimated in sections and stored in a table.

Similarly, the AM/PM curve follows an even relation and thus, the phase predistortion system should be also even. An even memoryless Volterra model (eq.10.a) and also a Fourier-Cosine model (eq.10.b) both of order N are proposed to implement the phase predistortion denoted by $g_{A/P}[.]$.

$$g^V_{A/P}\left[\hat{R}_k\right] = \sum_{n=0}^{N-1} b^V_k(n) \cdot \hat{R}_k^{2k} = \left(b^V_k\right)' \cdot v^V_k \quad (10.a)$$

$$g^{FC}_{A/P}\left[\hat{R}_k\right] = \sum_{n=0}^{N-1} b^{FC}_k(n) \cdot \cos\left(n\omega_0 \hat{R}_k\right) = b_k^{FC'} \cdot v^{FC}_k \quad (10.b)$$

The update equations for the coefficients $\mathbf{b}_k$ involved in the phase predistortion models can be performed by classical adaptive methods (the NLMS algorithm is proposed), since the design problem consists in a basic identification problem with a model that is linear in terms of its coefficients.

$$\boldsymbol{b}(k+1) = \boldsymbol{b}(k) + \frac{\mu}{p(k)} e_\theta(k) \boldsymbol{v}(k) \quad (11)$$

The term p(k) denotes the estimated input power approximated by a lowpass filtering of the input data vector $\mathbf{v}_k$ with a memory factor named β. It is interesting to note that, since the output of the amplitude predistortion drives the phase predistortion, the learning of the coefficients of the $g_{A/P}[.]$ model will be conditioned to the learning of the amplitude predistortion system.

4. Simulation Results

In this section, the results obtained in the simulation of the adaptive learning of the predistortion (fig.3) are presented. The input signal is a 64-QAM modulation generated from two 8-PAM signals, for the in-phase I_k and the in-quadrature Q_k components. The resulting magnitude should be less than 0.62, which is the range capable of being compensated since the output of the normalised F[.] function of the BJT transistor is bounded to this range (eq.5). A Gaussian noise (SNR=60dB) is also added to the output of the HPA distortion.

Concerning the amplitude distortion, $g_{A/A}[.]$, two different models are considered: an odd Volterra (oV) model (eq.8.a) and a Fourier-Sine (FS) model (eq.8.b), both with N=5 coefficients. The principal frequency for the Fourier model is chosen equal to $\omega_0=\pi/(2*0.62)$ since the input value R is bounded to 0.62. The coefficients are updated by the PLMS algorithm (eq.9), where the step size parameters are normalised to the power estimate of the respective data vector $\mathbf{u}$ ($\mu_{oV}(k)=0.003/(p_{oV}(k))$ with $p_{oV}(-1)\approx 0$, $\mu_{FS}(k)=2/N$). At this point it is worth valued to remark that the diversity managed by the Fourier model, i.e. the sine functions of successive harmonics, has not so scattered power values as the Volterra model.

This feature, together with other aspects concerning the correlation between the components of the vector $\mathbf{u}_k$ [3], usually provides the Fourier model with a better performance than the Volterra model in adaptive solutions. A clear example is shown in figure (4.a) that represents the squared amplitude error achieved by both models averaged over 25 independent realisations (the better result is achieved by the FS model). The step-sizes have been fitted after various tests to achieve a similar convergence rate for both models. Additionally, the amplitude predistortion system implemented by the Fourier model and Volterra model after the learning stage are included in figure (5.c).

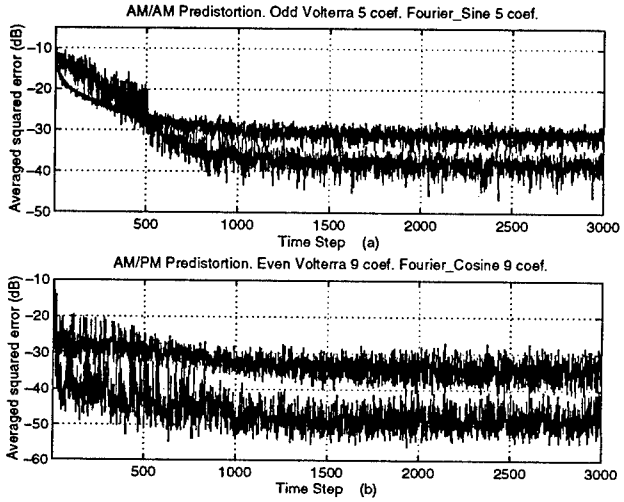


Figure 4. Squared envelope error (a) and squared phase error (b) averaged over 25 realisations.

Similarly, the averaged squared phase error achieved by the AM/PM predistortion is included in figure (4.b) (the Fourier model also shows the better results). For this nonlinear system, an even Volterra (eV) model and a Fourier-Cosine (FC) model with 9 coefficients are used. The respective coefficients are updated by the NLMS algorithm (eq.11) with the following step-sizes (μ_{Ve}=0.05 $p_{Ve}(-1)\approx 0$, μ_{FC}=0.3 $p_{FC}(-1)$=1). In this case, the principal frequency of the Fourier model is chosen equal to $\omega_0=\pi/(2)$ because the input to our model, denoted by $\hat{R}_k$ is bounded to one. At the beginning of the learning, this assumption does not hold and the phase predistortion learning could be in troubles. Thus, in the simulations, the output of the amplitude predistortion system is forced to be less than one in order to avoid this problem. From the error performance, it can be seen how the phase predistortion is conditioned to the convergence of the amplitude predistortion as it was expected. Finally, the AM/PM predistortion implemented by both models after the learning stage are also shown in figure (5.d), where the superior performance of the Fourier model becomes evident.

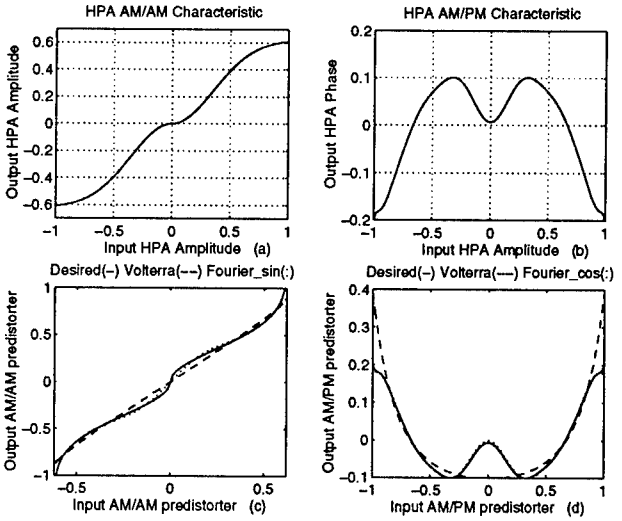


Figure 5. (a) HPA AM/AM relation. (b) HPA AM/PM relation. (c) Ideal amplitude predistortion (solid) and final amplitude predistortion of the Fourier-Sine model (dotted) and odd Volterra model (dashed). (d) Ideal phase predistortion (solid) and final phase predistortion of the Fourier-Cosine model (dotted) and odd-Volterra model (dashed).

Remarks

In this paper the HPA predistortion implemented by a Fourier model is compared with the performance achieved by the classical solution of using polynomial models. Although the Fourier model requires more computational load than the Volterra model, the existing fast DSP processors as well as the considerably superior performance achieved in this particular problem by the Fourier model seem to justify the use of this last one.

References

[1] Benedetto S., Biglieri E., Castellani V., "Digital Transmission Theory", Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.

[2] Pagès-Zamora A.,Lagunas M.A., Nájar M., Pérez-Neira A., "The K-filter: A new architecture to model and design non-linear systems from Kolmogorov's theorem", EURASIP Signal Processing, Vol.44, No.3, pp.249-267, July 1995.

[3] Pagès-Zamora A., Lagunas M.A., "Joint Probability Density Function Estimation by Spectral Estimate Methods", Proc. IEEE-ICASSP'96, Atlanta, Georgia, USA, May 1996.

[4] Mathews V.J., "Adaptive Polynomial Filters", IEEE Signal Processing Magazine, V.8, N°3, pp.10-26, July 1991.

[5] Stonick J.T. *et al.*, "Memoryless Polynomial Adaptive Predistortion", Proc. IEEE-ICASSP'95, pp.981-984, Detroit, Michigan, USA, May 1995.

Identification of a Class of Multiple Input–Output Nonlinear Systems Driven by Stationary Non-Gaussian Processes

Jonathon C. Ralston and Abdelhak M. Zoubir
Signal Processing Research Centre,
GPO Box 2434, Brisbane, Qld. 4001, Australia

Abstract

We consider the identification of a class of multiple input–output nonlinear systems when the inputs are stationary non-Gaussian processes. Currently, there are very few identification techniques which exist to solve this complicated problem. In an attempt to provide a solution, we extend the single input–output Hammerstein series to a multiple input–output version. Our solution for the multiple input–output problem in the non-Gaussian case is mathematically tractable and computationally attractive. Real data experiments are shown to indicate the usefulness of the method.

1. Introduction

The identification and analysis of multiple input–output systems is a problem of practical importance, which finds special application in seismic and array processing, physiological modelling, and vibration analysis (e.g., [1, 2, 4]). Most multiple input–output system identification procedures are based on the assumptions of linearity and Gaussianity, which inevitably exhibit limitations and weaknesses in practice. This is problematic as there are *many* multiple input–output nonlinear systems where the inputs are non-Gaussian processes (e.g., [4, 1, 9]). However, very few computationally efficient techniques exist to solve this identification problem.

A multiple input–output Volterra model has been proposed [1], but this can lead to an unwieldy multiple input–output system description. In addition, the Volterra approach necessitates large computational requirements in estimation. In an attempt to overcome these problems, we consider extending the single input–output *Hammerstein series* model (see [8]) to the multiple input–output case. The Hammerstein series has been shown to exhibit distinct practical advantages over the Volterra series in the non-Gaussian case. We derive solutions for the multiple input–output problem using the Hammerstein series and also derive multiple nonlinear coherence functions. The approach represents a mathematically tractable and computationally attractive solution to the multiple input–output nonlinear system identification problem in the non-Gaussian case.

2. The Identification Procedure

2.1 The Multiple Input–Output Model

We consider the multiple-input multiple-output (MIMO) nonlinear system, as it represents the most generalised configuration of the four multiple input–output scenarios (i.e., the single/multiple input–output configurations). We define the m-input, k-output, nth order discrete-time time-invariant Hammerstein series as

$$Y_r(t) = \sum_{q=1}^{n} \sum_{p=1}^{m} \sum_{\tau=-\infty}^{\infty} g_{prq}(\tau) X_p(t-\tau)^q + N_r(t)\,, \quad (1)$$

with inputs $X_p(t)$, $p = 1, 2, \ldots, m$ and outputs $Y_r(t)$, $r = 1, 2, \ldots, k$ and where the Hammerstein kernel, $g_{prq}(\tau)$, relates to the pth input, rth output, and qth nonlinearity respectively for $q = 1, 2, \ldots, n$. Note also that we have allowed for an additive, zero-mean disturbance signal $N_r(t)$, where we assume that $X_p(t)$ and $N_r(t)$ are independent for all p, r and that $N_r(t)$ and $N_s(t)$ are independent for all $s, r = 1, 2, \ldots, k$, $s \neq r$, and all t. We also assume that the inputs are real, zero-mean, and stationary processes with bounded cumulants up to $2n$th order. Equation (1) represents an extension of the Hammerstein series which has been successfully applied in the single-input single-output (SISO) case [9, 8].

In order to simplify the formulation of the problem, consider a *vectorial* version of the MIMO nonlinear system in (1), i.e.,

$$\mathbf{Y}(t) = \sum_{q=1}^{n} \sum_{\tau=-\infty}^{\infty} \mathbf{g}_q(\tau)' \mathbf{X}^{\odot q}(t-\tau) + \mathbf{N}(t)\,,$$

where $\mathbf{X}(t)$, $\mathbf{Y}(t)$, and $\mathbf{N}(t)$ respectively represent the m, k, and k vector-valued input, output and noise processes, and where the qth order $[m \times k]$ MIMO Hammerstein kernel matrix $\mathbf{g}_q(t)$ represents the *collection* of all Hammerstein kernels of order q, i.e., the p, rth element of $\mathbf{g}_q(t)$ is $g_{prq}(t)$. The notation $\mathbf{X}^{\odot q}(t)$ represents the q-fold Hadamard self-product of $\mathbf{X}(t)$ with itself, i.e., the pth element of $\mathbf{X}^{\odot q}(t)$ is $X_p(t)^q$. The vectorial notation conveniently enables us to express the MIMO model in similar manner to the SISO nonlinear case [8]. Note however that unlike most other nonlinear system identification procedures in the non-Gaussian case, we do not linearise the model in finding a solution (e.g., cf. [5, 6]).

0-8186-7576-4/96 $5.00

2.2 Solving for the MIMO Hammerstein Kernels

Since we are considering the identification of a nonlinear system in the non-Gaussian case, our solution naturally involves the use of higher order cumulant sequences [3]. We are able to greatly simplify the formulation of the solution by only requiring a particular *slice* of the cumulant sequence instead of the usual cumulant sequences [8]. The required $(p+q)$th multivariate cumulant sequence slice, $\mathbf{c}_{\underline{X}^p\underline{X}^q}(\tau)$, is defined as

$$\mathbf{c}_{\underline{X}^p\underline{X}^q}(\tau) \triangleq \text{cum}\left\{\mathbf{X}^{\odot p}(t), \mathbf{X}^{\odot q}(t-\tau)'\right\},$$

for $p, q = 1, 2, \ldots, n$, where cum represents the cumulant operator and $'$ is the matrix transpose operator. Note that the above cumulant sequence slice involves products of multivariate random processes, which can be expressed in terms of $(p+q)$th and lower order cumulant sequences (see [3]).

A minimum mean-square error criterion for the MIMO nth order Hammerstein kernels leads to the set of $[n \times n]$ linear (block) matrix equations

$$\mathbf{c}_{\underline{Y}\underline{X}^u}(v) = \sum_{q=1}^{n}\sum_{\tau=-\infty}^{\infty} \mathbf{g}_q(\tau)'\mathbf{c}_{\underline{X}^q\underline{X}^u}(v-\tau), \qquad (2)$$

for $u = 1, 2, \ldots, n$. Taking the Fourier transform of (2) with respect to v gives

$$\mathbf{C}_{\underline{Y}\underline{X}^u}(\omega) = \sum_{q=1}^{n} \mathbf{G}_q(\omega)'\mathbf{C}_{\underline{X}^q\underline{X}^u}(\omega), \qquad (3)$$

for $u = 1, 2, \ldots, n$ where $\mathbf{C}_{\underline{X}^q\underline{X}^u}(\omega)$ is the *one-dimensional* (integrated) polyspectral representation [8] corresponding to the Fourier transform of $\mathbf{c}_{\underline{X}^q\underline{X}^u}(\tau)$, and $\mathbf{G}_q(\omega)$ is the Fourier transform of $\mathbf{g}_q(t)$ with respect to t which we call the Hammerstein transfer functions. The key result to note is that the MIMO Hammerstein transfer functions have *separated* from the multivariate integrated polyspectra (cf. the Volterra series in the non-Gaussian case). Using the Hammerstein series greatly reduces the computational requirements in estimation. Simultaneously solving (3) leads to optimal mean-square solutions for $\mathbf{G}_q(\omega)$, $q = 1, 2, \ldots, n$

$$\begin{bmatrix}\mathbf{G}_1(\omega)\\ \mathbf{G}_2(\omega)\\ \vdots\\ \mathbf{G}_n(\omega)\end{bmatrix} = \begin{bmatrix}\mathbf{C}_{\underline{X}\underline{X}}(\omega)' & \cdots & \mathbf{C}_{\underline{X}^n\underline{X}}(\omega)'\\ \mathbf{C}_{\underline{X}\underline{X}^2}(\omega)' & \cdots & \mathbf{C}_{\underline{X}^n\underline{X}^2}(\omega)'\\ \vdots & \ddots & \vdots\\ \mathbf{C}_{\underline{X}\underline{X}^n}(\omega)' & \cdots & \mathbf{C}_{\underline{X}^n\underline{X}^n}(\omega)'\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{C}_{\underline{Y}\underline{X}}(\omega)'\\ \mathbf{C}_{\underline{Y}\underline{X}^2}(\omega)'\\ \vdots\\ \mathbf{C}_{\underline{Y}\underline{X}^n}(\omega)'\end{bmatrix}. \qquad (4)$$

Thus we have derived a general solution for a multiple input–output nth order nonlinear system identification in the general non-Gaussian case. Note that we do not perform a discrete frequency regression as in [5], but solve (4) with respect to the Hammerstein transfer functions as a continuous function of ω.

2.3 MIMO Nonlinear Coherence Function

The coherence function is very important in system identification as it provides a practical mechanism for model validation and system analysis. We use the notion of system coherency [5, 8] to derive a MIMO nonlinear coherence function. A block matrix version of the MIMO model is introduced so that the derivation of the MIMO nonlinear coherence function is not obscured by notation. Let the system of block matrix equations in (2) be expressed as

$$\mathcal{C}_{YX}(\omega) = \mathcal{G}(\omega)\mathcal{C}_{XX}(\omega), \qquad (5)$$

where $\mathcal{C}_{YX}(\omega)$ and $\mathcal{C}_{XX}(\omega)$ are $[k \times mn]$ and $[mn \times mn]$ block matrices, respectively. The above equation leads to the general solution for $\mathcal{G}(\omega)$ as in (4),

$$\mathcal{G}(\omega) = \mathcal{C}_{YX}(\omega)\,\mathcal{C}_{XX}(\omega)^{-1}. \qquad (6)$$

Let the $[k \times k]$ output spectral density matrix of the MIMO nonlinear system, $C_{\underline{Y}\underline{Y}}(\omega)$ be denoted by

$$C_{\underline{Y}\underline{Y}}(\omega) = \mathcal{G}(\omega)\mathcal{C}_{XX}(\omega)\mathcal{G}(\omega)^H + \mathbf{C}_{\underline{N}\underline{N}}(\omega), \qquad (7)$$

where the dimensions of (7) are $[k \times k] = [k \times mn]\,[mn \times mn]\,[mn \times k]$. An expression for the $[k \times k]$ MIMO nonlinear coherence function is found by substituting the expressions for $\mathcal{G}(\omega)$ in (6) into (7). Thus the MIMO nonlinear coherence matrix, $\mathcal{R}(\omega)$, is given by

$$\mathcal{R}(\omega) = \mathbf{C}_{\underline{Y}\underline{Y}}(\omega)^{-1}\mathcal{C}_{YX}(\omega)\mathcal{C}_{XX}(\omega)^{-1}\mathcal{C}_{YX}(\omega)^H.$$

where the above matrix dimensions are $[k \times k] = [k \times mn]\,[mn \times mn]\,[mn \times k]$. This yields an expression for the MIMO nonlinear coherence function. Since the MIMO nth order Hammerstein series has k-outputs, it follows that $\mathcal{R}(\omega)$ is a $[k \times k]$ matrix with elements

$$\mathcal{R}(\omega) \equiv \begin{bmatrix}R_{11}(\omega) & R_{12}(\omega) & \cdots & R_{1k}(\omega)\\ R_{21}(\omega) & R_{22}(\omega) & \cdots & R_{2k}(\omega)\\ \vdots & \vdots & \ddots & \vdots\\ R_{k1}(\omega) & R_{k1}(\omega) & \cdots & R_{kk}(\omega)\end{bmatrix}, \qquad (8)$$

where $R_{uv}(\omega), u, v = 1, 2, \ldots, k$ represents the nth order coherence function of the MIMO model. Thus a solution for the MIMO nth order coherence function has been derived which does not depend on any unknown MIMO Hammerstein transfer functions.

2.4 Discussion

Special Cases. It is straightforward to show how the MIMO model in (4) reduces to the single-input multiple-output, multiple-input single-output, and single-input single-output configurations, i.e., $m = 1, r = 1$, and $m = r = 1$, respectively.

Parameterisation Considerations. Although the Hammerstein series is not as general a mathematical model as the Volterra series, its use leads to significant reductions in the number of coefficients required in system modelling. As a simple example of the parameterisation advantages achieved in using the Hammerstein series, consider a two-input, two-output cubically nonlinear system with a system memory of $M = 10$ lags. To model this system, the Hammerstein series requires $Mnkm = 120$ coefficients, whereas the Volterra series requires a total of $\sum_{q=1}^{n}(M+q-1)!/((M-1)!q!)km = 1140$ distinct coefficients.

2.5 Estimation

We make estimates of the multivariate integrated polyspectra using an averaged periodogram based approach [3]. A similar approach is used for the univariate case in [8, 10].

We assume that the input-output sequences $X(t)$ and $Y(t)$ are available for $t = 0, \ldots, T-1$. Given that the data is stationary, we segment the sequences into M stretches each of length N, denoted by $X(t,m)$ and $Y(t,m)$, respectively, for $m = 1, \ldots, M$, such that $T = MN$. The finite Fourier transform of $X(t,m)^n$ is given by [7]

$$d_{X^n}^{(N)}(\omega, m) = \sum_{t=0}^{N-1} \left(X(t,m) - \hat{c}_{X_m}\right)^n e^{-j\omega t}, \quad n = 1, 2,$$

where $\hat{c}_{X_m}$ is the sample mean of $X(t,m)$. For the third-order case, the cumulant-moment relationship is given by

$$\begin{aligned} c_{X^2X}(\tau) &\triangleq \text{cum}\{X(t)X(t), X(t-\tau)\} \\ &\equiv \mathsf{E}\{X(t)^2 X(t-\tau)\} \end{aligned}$$

for $\mathsf{E}\{X(t)\} = 0$, which represents the sliced third order cumulant sequence of $X(t)$. This expression suggests the estimator

$$I_{X^2X}^{(N)}(\omega, m) = \frac{1}{(2\pi)N} d_{X^2}^{(N)}(\omega, m) d_X^{(N)}(\omega, m)^*, \qquad (9)$$

which is a form of cross-periodogram. An estimate for $C_{X^2X}(\omega)$ is found by averaging over the M terms in (9), and smoothing with an appropriate weighting function in the frequency domain [7]. In using the weighting functions we assume that the spectra have some smoothness properties. The large sample properties of this class of estimate are discussed in [3].

Estimates of $C_{XX}(\omega)$, $C_{YY}(\omega)$, $C_{XX^2}(\omega)$, $C_{X^2X^2}(\omega)$, $C_{YX}(\omega)$ and $C_{YX^2}(\omega)$ are found in a similar manner to the above. The estimates of the Hammerstein transfer functions and the nonlinear coherence function are subsequently found by substituting the estimated polyspectra into (4) and (8), respectively for a given n, k, and m.

3. Engine Transmission Modelling

We first verified the multiple input–output system identification technique using a simulated nonlinear system where we obtained good results. We then applied the method to a practical identification problem relating to engine knock (see [8, 12]).

An effective means for lowering fuel consumption and improving the efficiency of a combustion engine is to increase the compression ratio. However, this also increases the occurrence of an abnormal combustion phenomenon called *knock*. Knock needs to be avoided as it results in an excessively noisy, over-heated and inefficient engine. If the knocking condition can be detected, then it can be minimised without adversely effecting overall engine efficiency. The knocking condition can be detected by placing relatively inexpensive vibration sensors on the engine housing. Previously, a SISO quadratic Volterra series has been used to model the engine transmission characteristics between the cylinder pressure signal and a structural vibration signal [11].

We propose the use of a single-input two-output quadratically nonlinear Hammerstein series model (i.e., $m = 1, r = 2, n = 2$ in (1)) to model the transmission characteristics of a combustion engine operating in a knocking condition. Note that we do not attempt to solve the engine knock problem here, but rather focus on the application of the MIMO Hammerstein series as a nonlinear model. The cylinder pressure and engine vibration cycles measured from the engine were used as the system's input and output signals respectively[1]. A 1.8 l, 4 cylinder engine operating under strong knocking conditions at full load was used in the experiment.

A schematic of the assumed quadratically nonlinear system is shown in Figure 1, and has an input–output relationship given by

$$Y_r(t) = \sum_{\tau=-\infty}^{\infty} g_{r1}(\tau) X(t-\tau) + \sum_{\tau=-\infty}^{\infty} g_{r2}(\tau) X(t-\tau)^2 + N_r(t)$$

with input $X(t)$ and outputs $Y_r(t)$ for $r = 1, 2$. We assume that the additive noise terms $N_r(t)$ are zero-mean and stationary, and that $X(t)$ and $N_r(t)$ are independent. The cylinder pressure and structural vibration signals were used form estimates of the quadratically nonlinear Hammerstein transfer functions and the quadratic coherence function.

Using the results from Section 2, explicit solutions for the Hammerstein transfer functions are given by

$$G_{11}(\omega) = \frac{C_{X^2X^2}(\omega) C_{Y_1X}(\omega) - C_{X^2X}(\omega) C_{Y_1X^2}(\omega)}{C_{XX}(\omega) C_{X^2X^2}(\omega) - C_{XX^2}(\omega) C_{X^2X}(\omega)},$$

$$G_{12}(\omega) = \frac{C_{XX}(\omega) C_{Y_1X^2}(\omega) - C_{XX^2}(\omega) C_{Y_1X}(\omega)}{C_{XX}(\omega) C_{X^2X^2}(\omega) - C_{XX^2}(\omega) C_{X^2X}(\omega)},$$

$$G_{21}(\omega) = \frac{C_{X^2X^2}(\omega) C_{Y_2X}(\omega) - C_{X^2X}(\omega) C_{Y_2X^2}(\omega)}{C_{XX}(\omega) C_{X^2X^2}(\omega) - C_{XX^2}(\omega) C_{X^2X}(\omega)},$$

$$G_{22}(\omega) = \frac{C_{XX}(\omega) C_{Y_2X^2}(\omega) - C_{XX^2}(\omega) C_{Y_2X}(\omega)}{C_{XX}(\omega) C_{X^2X^2}(\omega) - C_{XX^2}(\omega) C_{X^2X}(\omega)}.$$

The above results represent an extremely efficient method for computing the MIMO Hammerstein transfer functions as the solution is in closed form (cf. matrix inversion techniques [5]). The quadratic coherence function associated with the quadratic model is found in a similar manner. Figure 2 shows estimates of the linear and quadratic Hammerstein series kernels for the two paths of the model, and the linear and quadratic coherences for the model. The closeness of the coherence function to unity indicates the general goodness of fit of the model. We found that the quadratic model provided a *better* characterisation of the engine block than a single-input multiple-output linear model.

4. Summary

We have formulated a procedure for identifying a class of multiple input–output nth order nonlinear systems when the inputs

[1] **Acknowledgement:** We would like to thank Professor J. F. Böhme from the Signal Theory Division of Ruhr University Bochum and Volkswagen AG, Wolfsburg, Germany, for kindly providing the knock data used in the paper.

are stationary non-Gaussian processes. We have validated the model by deriving a multiple input–output nonlinear coherence function. The solution has been formulated using special forms of polyspectra, which significantly simplifies the estimation and implementation of the model. In addition, we have avoided the parameterisation issues associated with the Volterra series by using the Hammerstein series as the system model. The solution represents a simple and practical approach to a difficult system identification problem. The technique has been validated with an application to an automotive engineering problem.

References

[1] C. K. An, E. J. Powers, and C. P. Ritz. Frequency domain modeling of dual-input/multiple-output quadratic systems with general random inputs. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, volume 3, pages 2209–2212, Espoo, Finland, 7-9 June 1988.

[2] J. S. Bendat and A. G. Piersol. *Random data: Analysis and measurement procedures*. John Wiley & sons, New York, 2nd edition, 1986.

[3] D. R. Brillinger. *Time series : data analysis and theory*. Holden-Day, San Francisco, 1981.

[4] H. W. Chen, L. D. Jacobson, and J. P. Gaska. Structural classification of multi-input nonlinear systems. *Biological Cybernetics*, 63:341–357, 1990.

[5] K. I. Kim and E. J. Powers. A digital method of modelling quadratically nonlinear systems with a general random input. *IEEE Transactions on ASSP*, 36(11):1758–1769, 1988.

[6] M. Korenburg, S. Bruder, and P. McIlroy. Exact orthogonal kernel estimation from finite data records: Extending Wiener's identification of nonlinear systems. *Annals of Biomedical Engineering*, 16:201–214, 1988.

[7] M. B. Priestley. *Non-Linear and Non-Stationary Time Series Analysis*. Academic Press, London, 1988.

[8] J. C. Ralston and A. M. Zoubir. Identification of quadratically nonlinear systems under stationary non-Gaussian excitation. In *Proceedings of the IEEE Signal Processing ATHOS Workshop on Higher-Order Statistics*, pages 419–423, Girona, Spain, June 1995.

[9] J. C. Ralston, A. M. Zoubir, and B. Boashash. Identification of a class of time-invariant and time-varying nonlinear systems under non-Gaussian processes. In T. Luk, editor, *Proceedings of the SPIE Conference on Advanced Signal Processing Algorithms, Architectures and Implementations*, pages 144–155, San Diego, USA, July 1995. Proc. SPIE.

[10] J. Tugnait. Detection of non-Gaussian signals using the integrated polyspectrum. *IEEE Transactions on Signal Processing*, 42(11):3137–3149, November 1994.

[11] A. M. Zoubir. Identification of second-order Volterra systems driven by non-Gaussian stationary processes. In T. Luk, editor, *Proceedings of the SPIE Conference on Advanced Signal Processing Algorithms, Architectures and Implementations*, pages 327–338, San Diego, USA, July 1992. Proc. SPIE.

[12] A. M. Zoubir and J. F. Böhme. Application of higher order spectra to the analysis and detection of knock in combustion engines. In B. Boashash, E. J. Powers, and A. M. Zoubir, editors, *Higher Order Statistical Signal Processing*, pp. 269–290, Longman Cheshire, Melbourne, 1995.

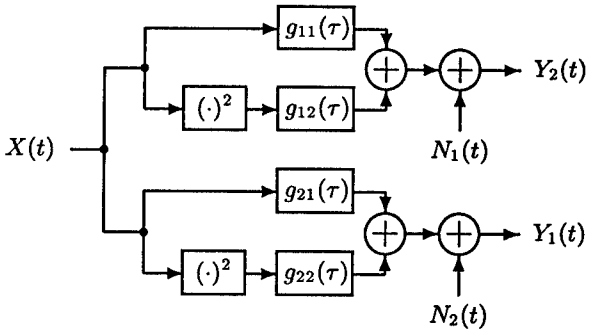
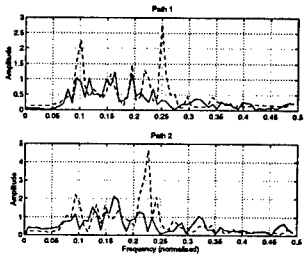
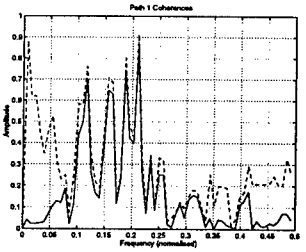


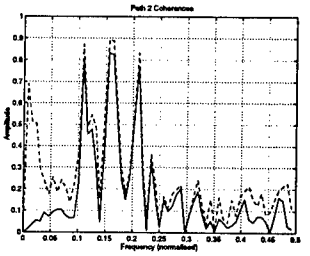
Figure 1. The single-input dual-output quadratically nonlinear Hammerstein series used as the system model.



(a) Hammerstein transfer functions



(b) Path 1 coherences



(c) Path 2 coherences

Figure 2. (a) Hammerstein transfer functions: Path 1 (top) and path 2 (bottom) showing linear (solid line) and quadratic (dashed line) transfer functions; (b) Ordinary linear (solid line) and quadratic (dashed line) coherences for path 1; (c) Ordinary linear (solid line) and quadratic (dashed line) coherences for path 2.

Second Order Volterra System Identification

Panos Koukoulas
koukoula@di.uoa.gr

Nicholas Kalouptsidis
University of Athens
Department of Informatics
Division of Communications and Signal Processing
kalou@di.uoa.gr

Abstract

In this paper closed form expressions for the identification of second order Volterra systems are developed. Two main cases are considered. The first case imposes natural constraints on the Volterra kernels leaving the input signal quite general. The second case leaves the kernels in general form but puts constraints on the input. In particular signals obtained as outputs of linear systems driven by higher order white noise are considered.

1. Introduction

We will be concerned with second order Volterra systems of the form

$$y(n) = \sum_{k_1=0}^{\infty} h_1(k_1)u(n-k_1) + \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} h_2(k_1,k_2)u(n-k_1)u(n-k_2) + \eta(n) \qquad (1)$$

The disturbance and input signals are independent zero mean processes. The Volterra kernels $h_1(k_1), h_2(k_1,k_2)$ are causal, absolutely summable, symmetric sequences.

Closed form expressions for the Volterrra kernels, for a second order system, have been determined when the input is a zero mean stationary Gaussian process [1]. The general $p-th$ order system with the same assumptions for the input is treated in [2]. These expressions utilize cumulant information in the time or frequency domain. Recall that if $x(k)$ is a stationary discrete time random process then the $p-th$ order cumulant of $x(k)$, denoted $c_x^p(k_1,k_2,\cdots,k_{p-1})$, is defined as the joint $p-th$ order cumulant of the random variables $x(k), x(k+k_1),\cdots,x(k+k_{p-1})$, i.e.,

$$c_x^p(k_1,k_2,\cdots,k_{p-1}) = \text{cum}(x(k), x(k+k_1),\cdots,x(k+k_{p-1}))$$

The $p-th$ order polyspectrum is the $(p-1)$-dimensional discrete time Fourier transform of the $p-th$ order cumulant $c_x^p(k_1,k_2,\cdots,k_{p-1})$. Cross-cumulant and cross-polyspectra of two jointly stationary stochastic processes are similarly defined.

The Gaussian asumption is not always realistic. Attempts to handle the non Gaussian case for kernels of compact support are presented in [3]. This paper derives closed form expressions for the identification of the Volterra kernels of (1) in two important cases : a) banded Volterra kernels and general inputs b) general Volterra kernels and inputs of special type.

2 Formulation as a fredholm integral equation

We first compute the cross-cumulant of y with one and two copies of the input, respectively. Using the properties of cumulants and Leonov - Shiryaev theorem [4], we obtain

$$\text{cum}[y(n), u(n-s_1)] = \sum_{k_1} h_1(k_1)\text{cum}[u(n-k_1), u(n-s_1)] + \sum_{k_1}\sum_{k_2} h_2(k_1,k_2)\text{cum}[u(n-k_1), u(n-k_2), u(n-s_1)]$$

$$\text{cum}[y(n), u(n-s_1), u(n-s_2)] = \sum_{k_1} h_1(k_1) \cdot \text{cum}[u(n-k_1), u(n-s_1), u(n-s_2)] + \sum_{k_1}\sum_{k_2} h_2(k_1,k_2) \cdot \text{cum}[u(n-k_1), u(n-k_2), u(n-s_1), u(n-s_2)] + 2\sum_{k_1}\sum_{k_2} h_2(k_1,k_2)\text{cum}[u(n-k_1), u(n-s_1)] \cdot \text{cum}[u(n-k_2), u(n-s_2)]$$

Passing to frequency domain we obtain

0-8186-7576-4/96 $5.00

$$C_{yu}(-w) = H_1(w)C_u^2(w) +$$
$$+\frac{1}{2\pi}\int_{-\pi}^{\pi} H_2(w-w_3,w_3)C_u^3(w-w_3,w_3)dw_3 \qquad (2)$$

$$C_{yuu}(-w_1,-w_2) = H_1(w_1+w_2)C_u^3(-w_1,-w_2)+$$
$$+2H_2(w_1,w_2)C_u^2(w_1)C_u^2(w_2)+$$
$$+\frac{1}{2\pi}\int_{-\pi}^{\pi} H_2(w_1+w_2-w_3,w_3)$$
$$\cdot C_u^4(-w_2,w_1+w_2-w_3,w_3)dw_3 \qquad (3)$$

Let $w \in [-\pi,\pi]$. Note that as long as we move on the line $w_1+w_2=w$, eq. (3) takes the form

$$C_{yuu}(-(w-w_2),-w_2) = H_1(w)C_u^3(-(w-w_2),-w_2)+$$
$$+2H_2(w-w_2,w_2)C_u^2(w-w_2)C_u^2(w_2)+$$
$$+\frac{1}{2\pi}\int_{-\pi}^{\pi} H_2(w-w_3,w_3)C_u^4(-w_2,w-w_3,w_3)dw_3 \qquad (4)$$

Solving eq. (2) with respect to $H_1(w)$ and substituting into eq. (4) we obtain

$$H_1(w) = \frac{C_{yu}(-w)}{C_u^2(w)} - \frac{1}{C_u^2(w)}\frac{1}{2\pi}$$
$$\cdot\int_{-\pi}^{\pi} H_2(w-w_3,w_3)C_u^3(w-w_3,w_3)dw_3 \qquad (5)$$

$$H_2(w-w_2,w_2) - (-\frac{1}{2\pi})$$
$$\cdot\int_{-\pi}^{\pi}(\frac{C_u^4(-w_2,w-w_3,w_3)C_u^2(w)}{2C_u^2(w)C_u^2(w-w_2)C_u^2(w_2)} -$$
$$-\frac{C_u^3(-(w-w_2),-w_2)C_u^3(w-w_3,w_3)}{2C_u^2(w)C_u^2(w-w_2)C_u^2(w_2)})H_2(w-w_3,w_3)dw_3$$
$$= \frac{C_{yuu}(-(w-w_2),-w_2)}{2C_u^2(w-w_2)C_u^2(w_2)} -$$
$$-\frac{C_{yu}(-w)C_u^3(-(w-w_2),-w_2)}{2C_u^2(w)C_u^2(w-w_2)C_u^2(w_2)} \qquad (6)$$

We observe that eq. (6) is a Fredholm integral equation of the second kind of the form

$$x(t) - \lambda\int_a^b K(t,\xi)x(\xi)d\xi = f(t)$$

The $2-D$ function $K(t,\xi)$ is the so called, kernel of the integral equation. In our case

$$x(t) = H_2(w-w_2,w_2) \qquad \lambda = -\frac{1}{2\pi}$$

$$K(t,\xi) = \frac{C_u^4(-w_2,w-w_3,w_3)C_u^2(w)}{2C_u^2(w)C_u^2(w-w_2)C_u^2(w_2)} -$$
$$-\frac{C_u^3(-(w-w_2),-w_2)C_u^3(w-w_3,w_3)}{2C_u^2(w)C_u^2(w-w_2)C_u^2(w_2)}$$

$$f(t) = \frac{C_{yuu}(-(w-w_2),-w_2)}{2C_u^2(w-w_2)C_u^2(w_2)} -$$
$$-\frac{C_{yu}(-w)C_u^3(-(w-w_2),-w_2)}{2C_u^2(w)C_u^2(w-w_2)C_u^2(w_2)}$$

The solution of these equations can be studied by various methods, including iterated kernels, successive approximation, the determinants method and the eigenvalues method. Approximate expressions are obtained if the integral is replaced by a finite sum. Then finite dimensional linear systems of equations result. Here we confine ourselves to the determinants method. It can be proved [5] that a necessary and sufficient condition for the existence of a unique continuous solution is that the Fredholm's determinant

$$\Delta(\lambda) = 1 + \sum_{\nu=1}^{\infty}(-1)^\nu\frac{\lambda^\nu}{\nu!}\int_a^b\cdots\int_a^b K\begin{pmatrix}\xi_1 & \xi_2 & \cdots & \xi_\nu\\ \xi_1 & \xi_2 & \cdots & \xi_\nu\end{pmatrix}$$
$$d\xi_1 d\xi_2\cdots d\xi_\nu$$

is not zero. The solution is given by

$$x(t) = f(t) - \lambda\int_a^b \Gamma(t,\xi;\lambda)f(\xi)d\xi$$

where the kernel $\Gamma(t,\xi;\lambda)$ is

$$\Gamma(t,\xi;\lambda) = -\frac{\Delta(t,\xi;\lambda)}{\Delta(\lambda)}$$

and

$$\Delta(t,\xi;\lambda) = K(t,\xi) + \sum_{\nu=1}^{\infty}(-1)^\nu\frac{\lambda^\nu}{\nu!}$$
$$\cdot\int_a^b\cdots\int_a^b K\begin{pmatrix}t & \xi_1 & \xi_2 & \cdots & \xi_\nu\\ \xi & \xi_1 & \xi_2 & \cdots & \xi_\nu\end{pmatrix}d\xi_1 d\xi_2\cdots d\xi_\nu$$

The notation

$$K\begin{pmatrix}\xi_1 & \xi_2 & \cdots & \xi_\nu\\ \xi_1 & \xi_2 & \cdots & \xi_\nu\end{pmatrix} \equiv$$

$$\begin{vmatrix} K(\xi_1,\xi_1) & K(\xi_1,\xi_2) & \cdots & K(\xi_1,\xi_\nu)\\ K(\xi_2,\xi_1) & K(\xi_2,\xi_2) & \cdots & K(\xi_2,\xi_\nu)\\ \vdots & \vdots & \ddots & \vdots\\ K(\xi_\nu,\xi_1) & K(\xi_\nu,\xi_2) & \cdots & K(\xi_\nu,\xi_\nu)\end{vmatrix}$$

is employed. The above expressions are invoked in section 4 for the computation of the Volterra kernels when the input is of special type.

An alternative expression for the system of eqs. (2) and (4) is next derived having the advantage that the resulting format is pointwise linear in the kernels. More specifically let us introduce the LTI filters with impulse responses

$$h_{2,k}(l) = h_2(k+l,l) \quad , \qquad k,l \in Z$$

Then

$$H_2(w_1,w_2) = \sum_m \sum_n h_2(m,n) e^{-jw_1 m} e^{-jw_2 n} =$$

$$= \sum_k \sum_l h_{2,k}(l) e^{-jw_1(k+l)} e^{-jw_2 l} =$$

$$= H_{2,0}(w_1+w_2) + \sum_{k=1}^{\infty} H_{2,k}(w_1+w_2)(e^{-jw_1 k} + e^{-jw_2 k})$$

Hence eqs. (2) and (4) become

$$C_{yu}(-w) = H_1(w) C_u^2(w) + H_{2,0}(w) \frac{1}{2\pi}$$

$$\cdot \int_{\pi}^{\pi} C_u^3(w-w_3,w_3) dw_3 + \sum_{l=1}^{\infty} H_{2,l}(w) \frac{1}{2\pi}$$

$$\cdot \int_{-\pi}^{\pi} C_u^3(w-w_3,w_3)(e^{-jl(w-w_3)} + e^{-jlw_3}) dw_3 \qquad (7)$$

$$C_{yuu}(-(w-w_2),-w_2) = H_1(w) C_u^3(-(w-w_2),-w_2) +$$

$$+ H_{2,0}(w)(2C_u^2(w-w_2) C_u^2(w_2) + \frac{1}{2\pi}$$

$$\cdot \int_{-\pi}^{\pi} C_u^4(-w_2, w-w_3, w_3,) dw_3) + \sum_{l=1}^{\infty} H_{2,l}(w)$$

$$[2C_u^2(w-w_2) C_u^2(w_2)(e^{-jl(w-w_2)} + e^{-jlw_2}) + \frac{1}{2\pi}$$

$$\int_{-\pi}^{\pi} C_u^4(-w_2, w-w_3, w_3)(e^{-jl(w-w_3)} + e^{-jlw_3}) dw_3] \qquad (8)$$

3 Banded second order volterra forms

We say that (1) is banded Volterra if the matrix H_2 associated with the second order kernel is a banded matrix, i.e. there exists an integer M such that $h_2(k_1,k_2) = 0$ for $|k_1 - k_2| > M$. The meaning of these systems is that input products with sufficiently wide time separation do not contribute to the output. In the banded case eqs. (7), (8) reduce to the linear system $Ax = b$, where the vector of the unknown parameters is

$$x = [H_1(w) \quad H_{2,0}(w) \quad H_{2,1}(w) \quad \cdots \quad H_{2,M}(w)]^T$$

the first row of the matrix A has the form

$$(C_u^2(w) \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} C_u^3(w-w_3,w_3) dw_3$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} C_u^3(w-w_3,w_3)(e^{-j1(w-w_3)} + e^{-j1w_3}) dw_3 \cdots$$

$$\cdots \frac{1}{2\pi} \int_{-\pi}^{\pi} C_u^3(w-w_3,w_3)(e^{-jM(w-w_3)} + e^{-jMw_3}) dw_3)$$

The first column of A has the form

$$(C_u^2(w) \quad C_u^3(-(w-w_{2,0}),-w_{2,0}) \quad \cdots$$

$$\cdots \quad C_u^3(-(w-w_{2,M}),-w_{2,M}))^T$$

Likewise the second column of A is

$$(\frac{1}{2\pi} \int_{-\pi}^{\pi} C_u^3(w-w_3,w_3) dw_3 \quad 2C_u^2(w-w_{2,0}) C_u^2(w_{2,0}) +$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} C_u^4(-w_{2,0}, w-w_3, w_3,) dw_3 \cdots 2C_u^2(w-w_{2,M})$$

$$\cdot C_u^2(w_{2,M}) + \frac{1}{2\pi} \int_{-\pi}^{\pi} C_u^4(-w_{2,M}, w-w_3, w_3,) dw_3)^T$$

The remaining elements of A are

$$a_{m,k} = 2C_u^2(w-w_{2,m-2}) C_u^2(w_{2,m-2})(e^{-j(k-2)(w-w_{2,m-2})}$$

$$+ e^{-j(k-2)w_{2,m-2}}) + \frac{1}{2\pi} \int_{-\pi}^{\pi} C_u^4(-w_{2,m-2}, w-w_3, w_3)$$

$$\cdot (e^{-j(k-2)(w-w_3)} + e^{-j(k-2)w_3}) dw_3$$

Finally the vector b is given by

$$b = (C_{yu}(-w) \quad C_{yuu}(-(w-w_{2,0}),-w_{2,0}) \cdots$$

$$\cdots \quad C_{yuu}(-(w-w_{2,M}),-w_{2,M}))^T$$

If the banded Volterra model is restricted to compact support Volterra kernels the above set up is analogous with the frequency domain approach of [3].

4 Volterra systems with special non-gaussian inputs

In this section identification of second order Volterra systems using inputs of special type is considered.

Let us first assume that the input signal is an IID zero mean random process with

$$c_u^k(i_1,\cdots,i_{k-1}) = \gamma_k \delta(i_1,\cdots,i_{k-1})$$

$\delta(i_1,\cdots,i_{k-1})$ is the $(k-1)$ dimensional unit sample, i.e,

$$\delta(i_1,\cdots,i_{k-1}) = \begin{cases} 1 & \text{if } i_1 = \cdots = i_{k-1} = 0 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$K(w_2,w_3) = \frac{\gamma_4\gamma_2 - \gamma_3^2}{2\gamma_2^3}$$

$$\Delta(\lambda) = 1 + \frac{\gamma_4\gamma_2 - \gamma_3^2}{2\gamma_2^3} = \frac{2\gamma_2^3 + \gamma_4\gamma_2 - \gamma_3^2}{2\gamma_2^3}$$

$$\Delta(w_2,w_3;\lambda) = \frac{\gamma_4\gamma_2 - \gamma_3^2}{2\gamma_2^3}$$

and

$$\Gamma(w_2,w_3;\lambda) = -\frac{\gamma_4\gamma_2 - \gamma_3^2}{2\gamma_2^3 + \gamma_4\gamma_2 - \gamma_3^2}$$

Therefore the first and second degree kernels are given by

$$H_2(w-w_2,w_2) = \frac{C_{yuu}(-(w-w_2),-w_2)}{2\gamma_2^2} -$$

$$\frac{\gamma_4\gamma_2 - \gamma_3^2}{2\gamma_2^2(2\gamma_2^3 + \gamma_4\gamma_2 - \gamma_3^2)} \frac{1}{2\pi}\int_{-\pi}^{\pi} C_{yuu}(-(w-w_3),-w_3)dw_3$$

$$-\frac{\gamma_3}{2\gamma_2^3 + \gamma_4\gamma_2 - \gamma_3^2} C_{yu}(-w)$$

$$H_1(w) = \frac{2\gamma_2^2 + \gamma_4}{2\gamma_2^3 + \gamma_4\gamma_2 - \gamma_3^2} C_{yu}(-w) -$$

$$\frac{\gamma_3}{2\gamma_2^3 + \gamma_4\gamma_2 - \gamma_3^2} \frac{1}{2\pi}\int_{-\pi}^{\pi} C_{yuu}(-(w-w_3),-w_3)dw_3$$

Next we assume that the input signal is obtained as the output of a linear time invariant system with transfer function $G(w)$ driven by a higher-order white noise. Then the spectra of the input signal have the form

$$C_u^2(w_1) = \gamma_2 G(w_1)G^*(w_1)$$

$$C_u^3(w_1,w_2) = \gamma_3 G(w_1)G(w_2)G^*(w_1+w_2)$$

$$C_u^4(w_1,w_2,w_3) = \gamma_4 G(w_1)G(w_2)G(w_3)G^*(w_1+w_2+w_3)$$

It can be proved in this case that $\Delta(\lambda)$ and $\Delta(w_2,w_3;\lambda)$ are given by

$$\Delta(\lambda) = 1 + \frac{1}{2\pi}\int_{-\pi}^{\pi} \Big(\frac{C_u^4(-\xi_1, w-\xi_1, \xi_1)C_u^2(w)}{2C_u^2(w)C_u^2(w-\xi_1)C_u^2(\xi_1)} -$$

$$-\frac{C_u^3(-(w-\xi_1),-\xi_1)C_u^3(w-\xi_1,\xi_1)}{2C_u^2(w)C_u^2(w-\xi_1)C_u^2(\xi_1)}\Big)d\xi_1$$

$$\Delta(w_2,w_3;\lambda) = \frac{C_u^4(-w_2, w-w_3, w_3)C_u^2(w)}{2C_u^2(w)C_u^2(w-w_2)C_u^2(w_2)} -$$

$$-\frac{-C_u^3(-(w-w_2),-w_2)C_u^3(w-w_3,w_3)}{2C_u^2(w)C_u^2(w-w_2)C_u^2(w_2)}$$

Hence

$$\Gamma(w_2,w_3;\lambda) = -\frac{\Delta(w_2,w_3;\lambda)}{\Delta(\lambda)}$$

Therefore the Volterra kernels are

$$H_2(w-w_2,w_2) = \frac{C_{yuu}(-(w-w_2),-w_2)}{2C_u^2(w-w_2)C_u^2(w_2)} -$$

$$-\frac{C_{yu}(-w)C_u^3(-(w-w_2),-w_2)}{2C_u^2(w)C_u^2(w-w_2)C_u^2(w_2)} +$$

$$+\frac{1}{2\pi}\int_{-\pi}^{\pi}\Gamma(w_2,w_3;\lambda)\frac{C_{yuu}(-(w-w_3),-w_3)}{2C_u^2(w-w_3)C_u^2(w_3)}dw_3 -$$

$$-\frac{1}{2\pi}\int_{-\pi}^{\pi}\Gamma(w_2,w_3;\lambda)\frac{C_{yu}(-w)C_u^3(-(w-w_3),-w_3)}{2C_u^2(w)C_u^2(w-w_3)C_u^2(w_3)}dw_3$$

References

[1] L.J. Tick, "The Estimation of Transfer Functions of Quadratic Systems", *Technometrics*, 3(4), pp. 562-567, November 1961.

[2] P. Koukoulas and N. Kalouptsidis, "Nonlinear System Identification Using Gaussian Inputs", *IEEE Trans. on Signal Processing*, vol. SP-38, pp. 1831-1841, August 1995.

[3] K.I. Kim and E.J. Powers, "A Digital Method of Modeling Quadratically Nonlinear Systems with a General Random Input", *IEEE Trans. ASSP*, vol. 36, pp. 1758-1769, November 1988.

[4] V.P.Leonov and A.N.Shiryaev, "On a Method of Calculation of Semi-Invariants", *Theory Prob. Appl.*, vol.4, pp 319-329, 1959.

[5] J. Kondo, *Integral Equations*, Oxford University Press, 1991.

The Computational Complexity of Nonlinear Compensators Based on the Volterra Inverse

John Tsimbinos[†,‡] and Kenneth V. Lever [‡]

[†] Defence Science and Technology Organisation, Australia*
[‡] Institute for Telecommunications Research, University of South Australia

Abstract

The performance of many analogue and digital signal processing systems is limited by nonlinear distortion mechanisms which can be modelled with a Volterra series. The nonlinear distortion can be compensated by the application of post (or pre)–distortion based on a Volterra inverse. The computational complexity associated with this type of compensation can be very high, particularly for systems with high nonlinearity order and long memory. In this paper we determine the 3rd and 5th order analytical Volterra inverses, and examine their associated computational complexity. We show how the analytical Volterra inverse can be used to determine the memory span of the kernels of an adaptive Volterra inverse, leading to computational complexity expressions. We then compare the computational complexity of the analytical and adaptive Volterra inverse. The results show that the analytical inverse has a much lower complexity than the adaptive inverse.

1. Introduction

The Volterra representation uses a set of functionals and kernels to model a wide class of nonlinear systems with memory [1]. The continuous time Volterra model is given by (1):

$$y(t) = H_0 + H_1[x(t)] + \ldots H_n[x(t)] + \ldots H_N[x(t)] \quad (1)$$

where $H_0 = h_0$ is the DC term, $H_n[.]$ is the nth order Volterra operator given by (2), and $h_n(\tau_1,\tau_2, \ldots,\tau_n)$ is the nth order Volterra kernel.

$$H_n[x(t)] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} h_n(\tau_1,\tau_2 \ldots \tau_n)x(t-\tau_1)x(t-\tau_2)\ldots x(t-\tau_n)d\tau_1 d\tau_2 \ldots d\tau_n \quad (2)$$

A Volterra inverse can be used to compensate for nonlinear distortion. For example, in a previous paper [2] we demonstrated how a Volterra inverse may be used to compensate for nonlinearities in a sample and hold with input dependent timing jitter. In this paper we give analytical expressions for the 3rd and 5th order Volterra inverses, and examine their computational complexity. We show how the analytical expressions can be used to determine the memory span of the kernels of an adaptive Volterra inverse. We then compare the computational complexity of analytical and adaptive Volterra inverses, illustrating that the analytical inverse has a much lower computational complexity than the adaptive inverse.

*Contact information: Defence Science and Technology Organisation, Communications Division, PO BOX 1500, Salisbury, South Australia, 5108, Australia, Tel: +61–8–82596403, Fax: +61–8–82596328, Email: john.tsimbinos@dsto.defence.gov.au

2. Computational complexity of the Volterra model

The discrete–time Nth order Volterra model with memory length truncated to M for all orders, and symmetric Volterra kernels (to avoid redundancy), can be written as:

$$y(k) = H_{(N)}[x(k)]$$

$$= H_0 + H_1[x(k)] + \ldots + H_n[x(k)] + \ldots + H_N[x(k)] \quad (3)$$

where $H_n[x(k)] =$

$$\sum_{m_1=0}^{M-1} \sum_{m_2=m_1}^{M-1} \ldots \sum_{m_n=m_{n-1}}^{M-1} h_n(m_1, m_2 \ldots m_n)x(k-m_1)x(k-m_2)\ldots x(k-m_n) \quad (4)$$

For the Volterra model given by (3), the total number of multiplications required gives a measure of the complexity, $\mathcal{C}(N, M)$:

$$\mathcal{C}(N, M) = \sum_{n=1}^{N} \frac{(M-1+n)!}{(n-1)!(M-1)!} \quad (5)$$

Fig. 1 shows how the complexity varies with Volterra model nonlinearity order N, and memory M. The high computational complexity places much emphasis on the development of efficient implementations and fast kernel estimation algorithms [3]–[6].

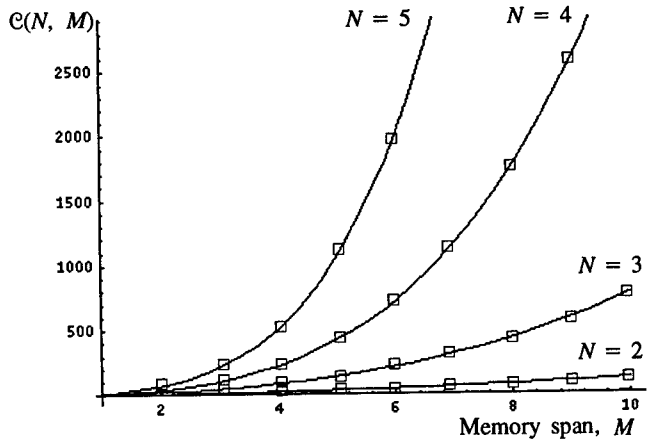


Fig. 1: Computational complexity as a function of Volterra model order N, and memory span M

0-8186-7576-4/96 $5.00 © 1996 IEEE

3. The Volterra inverse

Consider Fig. 2, where an Nth order Volterra model represented by $H_{(N)}[.]$ as in (3), is followed by a pth order Volterra inverse compensator $G_{(p)}[.]$ as in (6). In this paper we consider two types of inverses: an analytical inverse which is derived to eliminate all terms up to pth order, and an adaptive inverse which is obtained by minimising the mean square error between $x[k]$ and $\hat{x}[k]$.

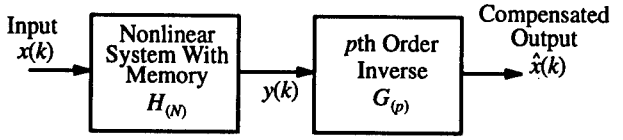


Fig. 2: Nonlinear system followed by a Volterra inverse

$$G_{(p)}[y(k)] = G_0 + G_1[y(k)] + \ldots + G_m[y(k)] + \ldots + G_p[y(k)] \quad (6)$$

3.1 Analytical Volterra inverse

The analytical pth order Volterra inverse $G_{(p)}$ is defined as one which, when cascaded with the Nth Volterra model $H_{(N)}$, results in a system $Q[.]$ with Volterra operators $Q_j[.]$, in which the 1st order Volterra kernel is a unit impulse and the higher jth order Volterra kernels are zero, for $j = 2, \ldots p$, [1], [7], as in (7).

$$\hat{x}(k) = G_{(p)}[H_{(N)}[x(k)]] = Q[x(k)]$$

$$= x + Q_{p+1}[x(k)] + Q_{p+2}[x(k)] + \ldots + Q_{pN}[y(k)]$$

$$= x + \sum_{j=p+1}^{pN} Q_j[x(k)] \quad (7)$$

We first consider a 3rd order Volterra inverse. Using (7), it is possible to derive the expressions for the Volterra inverse operators:

$$G_1 = H_1^{-1}\ ,\quad G_2 = -G_1H_2G_1\ , \quad (8)$$

$$G_3 = G_1[-H_2 + H_2[1 + G_1H_2]-H_2G_1H_2-H_3]G_1 \quad (9)$$

Fig. 3 gives a block diagram of the 3rd order Volterra inverse, and G_3 is shown in Fig. 4.

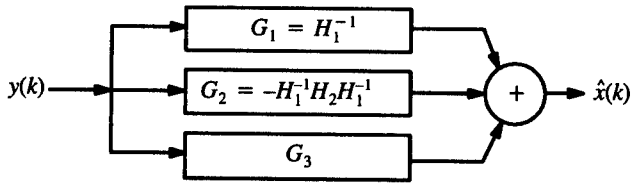


Fig. 3: 3rd order Volterra inverse

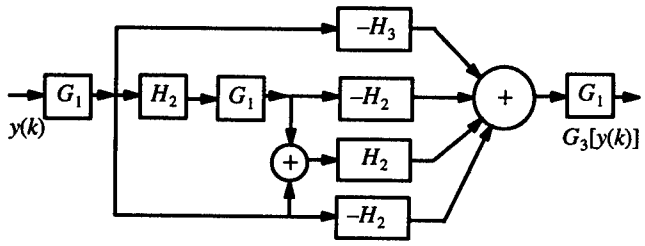


Fig. 4: G_3 of the 3rd order Volterra inverse

We now consider a 5th order Volterra inverse. It is possible to obtain the 1st, 2nd, 3rd, 4th and 5th order Volterra inverse operators [8]. However, for the purpose of this paper we will consider the case of a 5th order Volterra model with only odd order terms, such as the one we considered in [2]. Note that $G_2 = G_4 = 0$, and the remaining Volterra inverse operators are:

$$G_1 = H_1^{-1}\ ,\quad G_3 = -G_1H_3G_1\ , \quad (10)$$

$$G_5 = G_1[-H_5 - H_3[1 + G_1H_3] - 3H_3 + 0.5H_3G_1H_3 + 0.5H_3[2 + G_1H_3]]G_1 \quad (11)$$

Fig. 5 gives a block diagram of the 5th order Volterra inverse with only odd order terms, and the 5th order operator G_5 is shown in Fig. 6.

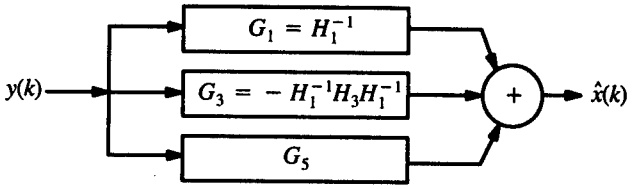


Fig. 5: 5th order inverse of a Volterra system with only odd order kernels

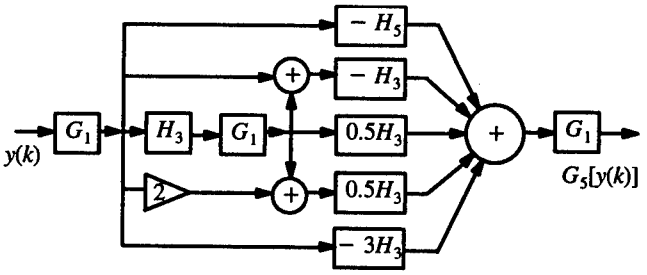


Fig. 6: G_5 of the 5th order inverse of a Volterra system with only odd order kernels

3.2 Adaptive Volterra inverse

It is also possible to obtain a Volterra inverse compensator by using an adaptive approach illustrated in Fig. 7. The inverse is obtained by minimising the mean square error between the ideal output $x[k]$ and the compensated output $\hat{x}[k]$, as in (12).

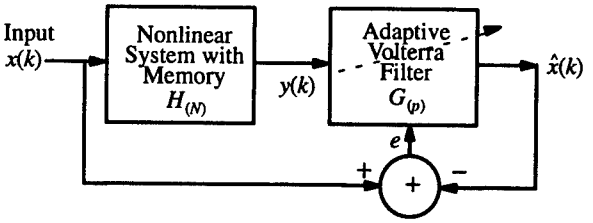


Fig. 7: Volterra inverse by adaptive method

$$E[e^2[k]] = E[(x[k] - \hat{x}[k])^2] \quad (12)$$

The adaptive method of obtaining a Volterra inverse compensator may appear to be more straightforward than deriving an analytical pth order Volterra inverse. After all, the kernels for the Volterra *inverse* are estimated without the need for obtaining a Volterra *model* of the original system. However, this method would be more difficult to apply in practice. Setting the nonlinearity order and memory of the

adaptive Volterra inverse requires some prior information about the original system. The significant nonlinearity orders of the system would determine the nonlinearity order of the adaptive inverse. However, determining the memory requirement of the inverse Volterra kernels is not trivial, even if the Volterra kernel memories of the original system are known. The memory of the inverse Volterra kernels will usually be higher than that of the Volterra kernels of the original system. The memory lengths of the kernels of the derived Volterra inverse given in Section 3.1 provide a method of determining the required memory of the kernels for the adaptive Volterra inverse. This will be discused in Section 4.2.

4. Computational complexity of analytical and adaptive Volterra inverses

We will now determine the computational complexity of the Volterra inverses. We restrict our discussion to two cases: one involving a 3rd order Volterra system, the other involving a 5th order Volterra system with only odd order terms. For both cases, all Volterra *system* kernels are assumed to have the same memory span, M.

4.1 Complexity of analytical Volterra inverse

First we consider the complexity of a 3rd order analytical Volterra inverse. We assume that an IIR filter is used to implement the first order inverse operator $G_1 = H_1^{-1}$, resulting in a memory span of M, the same as that of H_1. In any case, G_1 does not contribute significantly to the overall computational complexity of a 3rd order Volterra inverse. From (8), we have $G_2 = -G_1H_2G_1$, and the computational complexity contributions of G_2 are summarised in Table 1.

components of inverse operator G_2	memory span	nonlinearity order	complexity	number of components
G_1	M	1	M	2
H_2	M	2	$\frac{(M+1)!}{(M-1)!}$	1
Scaling coefficients	–	–	1	1

Table 1: Computational complexity contributions of G_2

From (9) and Fig. 4, we can summarise the computational complexity contributions of G_3 in Table 2.

components of inverse operator G_3	memory span	nonlinearity order	complexity	number of components
G_1	M	1	M	3
H_2	M	2	$\frac{(M+1)!}{(M-1)!}$	4
H_3	M	3	$\frac{(M+2)!}{2(M-1)!}$	1
Scaling coefficients	–	–	1	2

Table 2: Computational complexity contributions of G_3

By summing all contributions of Table 1 and Table 2, we can obtain the total complexity of the 3rd order analytical inverse, $\mathcal{C}_{an}(3, M)$:

$$\mathcal{C}_{an}(3, M) = 3 + 6M + 5\frac{(M+1)!}{(M-1)!} + \frac{(M+2)!}{2(M-1)!} \qquad (13)$$

Next we consider the complexity of the 5th order analytical Volterra inverse with only odd order terms. By considering all computational complexity contributions from each of the components shown in Fig. 5 , it can be shown that the complexity of the 5th order analytical inverse with only odd order terms is given by $\mathcal{C}_{an}(5, M)$:

$$\mathcal{C}_{an}(5, M) = 7 + 6M + 6\frac{(M+2)!}{2(M-1)!} + \frac{(M+4)!}{4!(M-1)!} \qquad (14)$$

4.2 Complexity of adaptive Volterra inverse

Now consider an adaptive Volterra inverse. Sufficient memory has to be set for the measurement of all inverse Volterra kernels. In order to determine the required memory span of each adaptive Volterra inverse kernel, it is necessary to make use of analytical Volterra inverse operators. In general, the adaptive Volterra inverse (which minimises the mean square error at the output), would not be directly equivalent to, or give the same compensation performance as the pth order analytical Volterra inverse (which is designed to remove the nonlinear distortion terms up to pth order). However, the analytical Volterra inverse operators provide a method of determining the memory requirements of the adaptive Volterra inverse operators.

First we consider the 3rd order Volterra system with all orders 1st, 2nd and 3rd, having a memory span M. Again we assume that the memory span for G_1 is M. To determine the memory requirement of the 2nd and 3rd order Volterra kernels, we make use of the analytical expressions for of G_2 and G_3 given in (8) and (9). Since the memory span of G_1 and H_2 is M, the memory requirement of G_2 is $(2(M-1) + 1)$. Since the memory span of G_1, H_2, and H_3 is M, the memory requirement of G_3 is $(4(M-1) + 1)$. Using these memory spans as a guide to the memory requirement of the adaptive Volterra inverse operators, it can be shown that the computational complexity of the 3rd order adaptive inverse $\mathcal{C}_{ad}(3, M)$, is given by:

$$\mathcal{C}_{ad}(3, M) = M + \frac{(2M)!}{(2M-2)!} + \frac{(4M-1)!}{2(4M-4)!} \qquad (15)$$

Now we consider an adaptive 5th order Volterra inverse with only odd order terms. Using the kernel memory length of the Volterra model, and the analytical Volterra inverse operators G_3 and G_5, given in (10) and (11), we can determine the memory requirement of each of the adaptive Volterra inverse kernels. We have $G_3 = -G_1H_3G_1$. Since the memory span of G_1 and H_3 is M, the overall memory requirement of G_3 is $(2(M-1) + 1)$. Now consider the expression for G_5. Since the memory span of G_1, H_3, and H_5 is M, the memory requirement of G_5 is $(4(M-1) + 1)$. Again, using these memory spans as a guide to the memory requirement of the adaptive Volterra inverse operators, it can be shown that the computational complexity of the 5th order adaptive inverse $\mathcal{C}_{ad}(5, M)$, is given by:

$$\mathcal{C}_{ad}(5, M) = M + \frac{(2M+1)!}{2(2M-2)!} + \frac{(4M+1)!}{4!(4M-4)!} \qquad (16)$$

The computational complexities of the adaptive inverses are compared with those of the analytical (derived) inverses for varying values of M. The 3rd order case is shown in Fig. 8, and the 5th order case is shown in Fig. 9. As can be seen, the analytical inverse has a much lower computational complexity than the corresponding adaptive inverse. This may not be surprising since the analytical inverse is derived using the actual Volterra model of the system which would be obtained by estimating a set of Volterra kernels. The adaptive inverse on the other hand, is implemented using less prior information about the system, resulting in a more general Volterra inverse, and a correspondingly higher computational complexity. We can also compare the complexity of the Volterra *inverses* with that of the corresponding Volterra *models* shown in Fig. 1. It can be shown that the computational complexity of the analytical Volterra inverse is of the same order of magnitude as the corresponding Volterra model, while the adaptive Volterra inverse has much higher complexity.

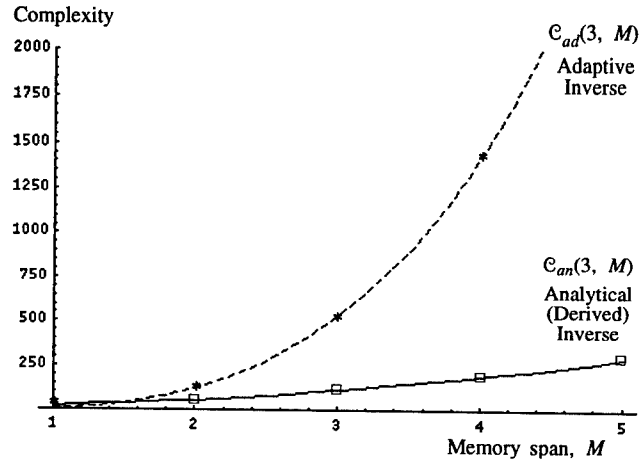


Fig. 8: Computational complexity comparison between the analytical inverse and adaptive inverse (3rd order case)

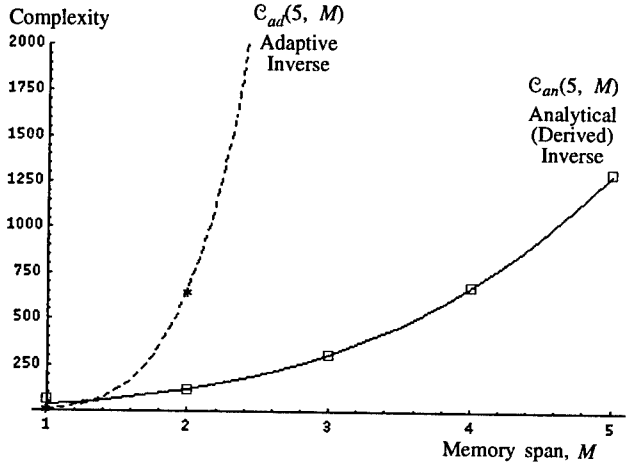


Fig. 9: Computational complexity comparison between the analytical inverse and adaptive inverse (5th order case with only odd kernels)

5. Conclusion

In this paper we have presented the architectures for 3rd and 5th order analytical Volterra inverses. We showed how a Volterra inverse can also be obtained by an adaptive approach. We explained how the analytical Volterra inverse may be used to determine the memory requirements of the adaptive Volterra inverse kernels. The computational complexity of the two types of Volterra inverse (analytical and adaptive) was examined. It was shown that the analytical Volterra inverse gives much lower computational complexity than its adaptive counterpart. The computational complexity of the analytical Volterra inverse is of the same order of magnitude as the Volterra model, while the more general, adaptive Volterra inverse has much higher complexity than the Volterra model being compensated. Processor technology governs the operation rate, so for high sampling rates, real time compensation using a Volterra inverse is limited to low order, short memory cases. Using a Volterra model of the system to derive an analytical pth order Volterra inverse would give implementation advantages over the use of the more general, higher–complexity, adaptive Volterra inverse. This paper carried out comparisons for cases involving the compensation of Volterra systems for which all kernels are assumed to have the same memory span. An extension to this work would consider a Volterra system with different memory spans for each order. Future work will investigate the use of more efficient implementation and computing structures to reduce the computational burden of the Volterra inverse based compensators.

References

[1] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, Reprint Edition, Krieger, Malabar, Florida, 1989.

[2] J. Tsimbinos and K. V. Lever, 'Applications of Higher–order Statistics to Modelling, Identification and Cancellation of Nonlinear Distortion in High–speed Samplers and Analogue–to–digital Converters using the Volterra and Wiener Models,' *Proc. IEEE Signal Processing Workshop on Higher–Order Statistics*, South Lake Tahoe, California, U.S.A., pp. 379–383, June 1993.

[3] W. A. Frank, 'MMD – An Efficient Approximation to the 2nd Order Volterra Filter,' *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, (ICASSP'94)*, Adelaide, Australia., pp. III–517–520, April 1994.

[4] B. G. Mertzios, 'Parallel Modeling and Structure of Nonlinear Volterra Discrete Systems,' *IEEE Trans. on Circuits and Systems–I: Fundamental Theory and Applications*, Vol. 41, No. 5, pp. 359–371, May, 1994.

[5] R. D. Nowak and B. D. Van Veen, 'Reduced Parameter Volterra Filters,' *Proceedings of ICASSP95*, pp. 1569–1572, May 1995.

[6] R. D. Nowak and B. D. Van Veen, 'Efficient Methods for Identification of Volterra Filter Models,' *Signal Processing*, pp. 417–428, Vol 38, No. 3, Aug. 1995.

[7] M. Schetzen, 'Theory of pth–Order Inverses of Nonlinear Systems,' *IEEE Trans. on Circuits and Systems*, pp. 285–291, Vol. 23, No. 5, May 1976.

[8] J. Tsimbinos, *Identification And Compensation of Nonlinear Distortion*, PhD Dissertation, School of Electronic Engineering, University of South Australia, Adelaide, February 1995.

Frequency Rate Estimation Based on Fourth Order Sample Moments*

Peter Händel[†‡]
Systems and Control Group
Uppsala University
Box 27, S–751 03 Uppsala, Sweden

Petr Tichavský[§]
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Box 18, 182 08 Prague, Czech Republic

Abstract

Estimation of frequency rate of linear frequency modulated signals based on phase angles of fourth order sample moments is considered. Three low-complexity estimators are proposed whose performance is close to optimal, that is their error variance is close to the Cramér-Rao lower bound.

1. Introduction

An important signal processing problem is the estimation of the parameters of complex-valued linear frequency modulated signals from noisy discrete time observations. Often, the frequency rate is the only parameter of interest. In this paper, a novel set of methods is proposed for this estimation problem, that is methods based on normalized phase angles of fourth order sample moments.

An extensive review of different algorithms for this estimations problem is given in [2]. Methods that have been suggested include maximum likelihood estimates, [1], estimates utilizing the polynomial phase transform, [4], and a Markov based estimator derived from the phase angles of the sequence $\{z_k\}_{k=3}^N$ where $z_k = x_k x_{k-1}^* x_{x-1}^* x_{k-2}$ and where $\{x_k\}_{k=1}^N$ denotes the observations, [3]. For the estimator proposed in [4] the frequency rate is estimated by the spectral position of the highest peak of the magnitude squared discrete time ambiguity function. This method is easily implemented by a grid search of the periodogram of the sequence $\{y_k\}_{k=1+\tau}^N$ where $y_k = x_k x_{k-\tau}^*$. For the choice $\tau = N/2$ the quotient of the error variance divided by the Cramér-Rao lower bound (CRB) tends for SNR$\rightarrow \infty$ to $16/15 \approx 1.07$, [4]. The estimators proposed here require no numerical search, thus they directly provide an estimate of the frequency rate.

Consider,

$$\begin{aligned} x_k &= s_k + v_k \qquad k = 1, \ldots, N \\ s_k &= Ae^{i\phi_k} \end{aligned} \tag{1}$$

where A is a complex-valued amplitude, and the noise v_k is zero mean complex-valued white Gaussian with variance σ^2. The real and imaginary parts of v_k are independent with variances $\sigma^2/2$, respectively. Further,

$$\phi_k = 2\pi\left(fk + \frac{\alpha}{2}k^2\right) \tag{2}$$

where $f \in (-1,\ 1)$ is the normalized frequency, and $\alpha \in (-0.5,\ 0.5)$ is the frequency rate. The parameters (A, f, α, σ^2) are all unknown, but often the frequency rate is the only parameter of interest. For this estimation problem, the CRB is given by, [3]

$$\mathrm{CRB}[\hat{\alpha}] = \frac{90}{\mathrm{SNR}\pi^2 N(N^2-1)(N^2-4)} \tag{3}$$

where $\hat{\alpha}$ denotes the estimated frequency rate and where the SNR is defined by $\mathrm{SNR} = |A|^2/\sigma^2$.

A set of frequency rate estimators is proposed based on normalized phase angles of fourth order sample moments, that is

$$\hat{C}(m) = \sum_{k=2m+1}^{N} x_k x_{k-m}^* x_{k-m}^* x_{k-2m} \quad m = 1, \ldots, \frac{N-1}{2} \tag{4}$$

where, for simplicity, N is assumed to be odd. Further,

$$\hat{\Phi}(m) = \frac{1}{2\pi m^2}\angle\left[\hat{C}(m)\right]. \tag{5}$$

In (4)-(5), $*$ denotes complex conjugate, and $\angle[\cdot]$ denotes the phase angle (in $(-\pi,\ \pi)$) of the expression

*In part supported by the ISS'90 Foundation.
†Now with Ericsson Radio Systems AB, Stockholm, Sweden.
‡ph@syscon.uu.se, Fax: +46-18-503611.
§tichavsk@utia.cas.cz, Fax: +42-2-66414903.

0-8186-7576-4/96 $5.00

between the brackets. If $|\alpha| > 1/(2m^2)$, equation (5) has to be replaced by another one that takes into the account the phase unwrapping, for example

$$\hat{\Phi}(m) = \frac{1}{2\pi m^2}\begin{cases}\angle\left[\hat{C}(1)\right] & m=1\\ \angle\left[\hat{C}(m)\right]+2\pi\psi(m) & m=2,\ldots,J\end{cases} \quad (6)$$

$$\psi(m) = \left\{m^2\hat{\Phi}(1) - \frac{\angle\left[\hat{C}(m)\right]}{2\pi}\right\}_Z \quad (7)$$

where $\{\cdot\}_Z$ denotes the round-off operation to the nearest integer applied to the quantity between the brackets.

The estimator is motivated by the fact that for a noiseless signal s_k it gives the correct (true) value of the frequency rate, because

$$\begin{aligned} C(m) &= \sum_{k=2m+1}^{N} |A|^4 e^{i(\phi_k - 2\phi_{k-m} + \phi_{k-2m})} \\ &= |A|^4(N-2m)e^{i2\pi m^2\alpha}. \end{aligned} \quad (8)$$

Here, $C(m)$ is the sample moment calculated for the noiseless signal, ($v_k \equiv 0$).

In this paper, frequency rate estimators are considered that are based on the set of sample moments $\{\hat{C}(m)\}_{m=1}^{J}$ where $J = 1,\ldots,(N-1)/2$. For the unwrapped sequence $\{\hat{\Phi}(m)\}_{m=1}^{J}$ it holds that $\hat{\Phi}(m) = \alpha + e(m)$ where $e(m)$ is a zero mean *colored* noise. Therefore, with $\mathbf{1} = (1\ \cdots\ 1)^T$ and $\Psi = (\hat{\Phi}(1)\ \cdots\ \hat{\Phi}(J))^T$, the Markov estimator of α is given by, [5]

$$\hat{\alpha} = \frac{\mathbf{1}^T\mathbf{R}^{-1}\Psi}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}} \quad (9)$$

where $\mathbf{R}$ is the $(J|J)$-covariance matrix with elements $\mathbf{R}_{m,n} = \text{cov}(\hat{\Phi}(m), \hat{\Phi}(n))$.

The matrix $\mathbf{R}$ depends on SNR as well as on N and J. A full expression for $\mathbf{R}$ is in principle possible to derive. In this paper, however, two approximate expressions of $\mathbf{R}$ are derived and used instead of $\mathbf{R}$. One expression valid for high SNR (SNR/$N \gg 1$), and one valid for low SNR (N/SNR$\gg 1$). The motivation to use approximate expressions for $\mathbf{R}$ is as follows. The SNR is, in general, unknown and has to be estimated leading to a multi step procedure where in the final step the frequency rate is estimated using (9) with $\mathbf{R}$ replaced by an estimate $\hat{\mathbf{R}}$. The use of approximate expressions, however, give estimators independent of SNR, leading to a direct method for which closed form expressions can be derived. As shown in the sequel, the performance of the proposed methods is (very) close to the CRB.

2. Covariance elements of $\hat{\Phi}(m)$

First, the covariance elements of $\hat{\Phi}(m)$ is expressed in terms of covariance elements of $\hat{C}(m)$. Let $\delta C(m) = \hat{C}(m) - C(m)$ and

$$\begin{aligned} \delta\Phi(m) &= \hat{\Phi}(m) - \Phi(m) \\ &= \frac{1}{2\pi m^2}\text{Im}\left\{\log\left(1+\frac{\delta C(m)}{C(m)}\right)\right\} \quad (10) \\ &\simeq \frac{1}{2\pi m^2}\text{Im}\left\{\frac{\delta C(m)}{C(m)}\right\}. \end{aligned}$$

The last approximation in (10) is valid for $|\delta C(m)| \ll 1$. Using the assumptions that the noise is circular white Gaussian one can find that $\text{E}[\hat{C}(m)] = C(m)$ for $m > 0$. Next, let

$$\mathbf{G}_{m,n} = \text{cov}(\hat{C}(m), \hat{C}(n)) = \text{E}[\delta C^*(m)\delta C(n)] \quad (11)$$

$$\mathbf{H}_{m,n} = \text{cov}(\hat{C}^*(m), \hat{C}(n)) = \text{E}[\delta C(m)\delta C(n)] \quad (12)$$

$$\mathbf{R}_{m,n} = \text{cov}(\hat{\Phi}(m), \hat{\Phi}(n)) \approx \text{E}[\delta\Phi(m)\delta\Phi(n)]. \quad (13)$$

Using (10) it follows that

$$\begin{aligned} \mathbf{R}_{m,n} &\approx \frac{1}{4\pi^2m^2n^2}\text{E}\left[\frac{1}{2i}\left(\frac{\delta C(m)}{C(m)} - \frac{\delta C^*(m)}{C^*(m)}\right)\right. \\ &\quad \left.\times\frac{1}{2i}\left(\frac{\delta C(n)}{C(n)} - \frac{\delta C^*(n)}{C^*(n)}\right)\right] \quad (14) \\ &= \frac{1}{8\pi^2m^2n^2}\text{Re}\left\{\frac{\mathbf{G}_{m,n}}{C^*(m)C(n)} - \frac{\mathbf{H}_{m,n}}{C(m)C(n)}\right\} \end{aligned}$$

where the identity $\text{Im}\{z\} = (z - z^*)/(2i)$ was used. Next, note the dependence of $\mathbf{G}_{m,n}$, $\mathbf{H}_{m,n}$ and $\mathbf{R}_{m,n}$ on the SNR. Following the reasoning of [6], it can be seen that

$$\mathbf{G}_{m,n} = \sum_{k=1}^{4} g_{mnk}\text{SNR}^{-k} \quad (15)$$

$$\mathbf{H}_{m,n} = \sum_{k=1}^{4} h_{mnk}\text{SNR}^{-k} \quad (16)$$

$$\mathbf{R}_{m,n} = \sum_{k=1}^{4} r_{mnk}\text{SNR}^{-k}. \quad (17)$$

For simplicity, we restrict ourselves to calculating the terms proportional to SNR^{-1} and SNR^{-4}. Thus, we obtain two approximations to the true covariance elements: the former approximation is valid for high SNR scenarios, and the latter one is good for low SNR scenarios. Indeed, in the latter case the sample size has to be considerably large in order to achieve reasonable estimates.

The decomposition of the covariance elements in (15)–(17) corresponds to the decomposition

$$\delta C(m) = \sum_{k=1}^{4} \delta^{(k)} C(m) \tag{18}$$

where $\delta^{(k)}C(m)$ consists of the terms where the noise appears in the k-th power, namely

$$\delta^{(1)}C(m) = \sum_{k=2m+1}^{N} v_k s_{k-m}^* s_{k-m}^* s_{k-2m} \tag{19}$$
$$+2s_k v_{k-m}^* s_{k-m}^* s_{k-2m} + s_k s_{k-m}^* s_{k-m}^* v_{k-2m}$$
$$\delta^{(4)}C(m) = \sum_{k=2m+1}^{N} v_k v_k^* v_k^* v_{k-2m}. \tag{20}$$

A straightforward calculation gives

$$\begin{aligned} g_{mn4}\mathrm{SNR}^{-4} &= \mathrm{E}[\delta^{(4)}C^*(m)\delta^{(4)}C(n)] \\ &= \sum_{k=2m+1}^{N}\sum_{\ell=2n+1}^{N} \mathrm{E}[(v_k v_{k-m}^* v_{k-m}^* v_{k-2m})^* \\ &\quad \times (v_\ell v_{\ell-n}^* v_{\ell-n}^* v_{\ell-2n})] \\ &= 2(N-2m)\sigma^8\delta_{mn} \end{aligned} \tag{21}$$

$$\begin{aligned} h_{mn4}\mathrm{SNR}^{-4} &= \mathrm{E}[\delta^{(4)}C(m)\delta^{(4)}C(n)] \\ &= \sum_{k=2m+1}^{N}\sum_{\ell=2n+1}^{N} \mathrm{E}[v_k v_{k-m}^* v_{k-m}^* v_{k-2m} \\ &\quad \times v_\ell v_{\ell-n}^* v_{\ell-n}^* v_{\ell-2n}] = 0 \end{aligned} \tag{22}$$

$$\begin{aligned} r_{mn4}\mathrm{SNR}^{-4} &= \frac{1}{8\pi^2 m^4}\frac{g_{mn4}\mathrm{SNR}^{-4}}{|C(m)|^2} \\ &= \frac{\delta_{mn}\mathrm{SNR}^{-4}}{4\pi^2 m^4(N-2m)} \quad m,n = 1,\ldots,J. \end{aligned} \tag{23}$$

In order to evaluate coefficient r_{mn1} introduce a normalized noise $\varepsilon_k = v_k/s_k$. Properties of ε_k read $\mathrm{E}[\varepsilon_m^*\varepsilon_n] = \sigma^2/|A|^2\,\delta_{mn}$, $\mathrm{E}[\varepsilon_m\varepsilon_n] \equiv 0$. Using this notation, a combination of (8) and (19) gives

$$\delta^{(1)}C(m) = \frac{C(m)}{N-2m}\sum_{k=2m+1}^{N} \varepsilon_k + 2\varepsilon_{k-m}^* + \varepsilon_{k-2m} \tag{24}$$

$$\begin{aligned} g_{mn1} &= \mathrm{E}[\delta^{(1)}C^*(m)\delta^{(1)}C(n)]\,\mathrm{SNR} \\ &= \frac{C^*(m)}{N-2m}\frac{C(n)}{N-2n}\sum_{k=2m+1}^{N}\sum_{\ell=2n+1}^{N}[\delta_{k\ell} + \delta_{k,\ell-2n} \\ &\quad +\delta_{k-2m,\ell} + \delta_{k-2m,\ell-2n} + 4\delta_{k-m,\ell-n}] \end{aligned} \tag{25}$$

$$\begin{aligned} h_{mn1} &= \mathrm{E}[\delta^{(1)}C(m)\delta^{(1)}C(n)]\mathrm{SNR} \\ &= 2\frac{C(m)}{N-2m}\frac{C(n)}{N-2n}\sum_{k=2m+1}^{N}\sum_{\ell=2n+1}^{N}[\delta_{k,\ell-n} \\ &\quad +\delta_{k-2m,\ell-n} + \delta_{k-m,\ell} + \delta_{k-m,\ell-2n}] \end{aligned} \tag{26}$$

$$r_{mn1} = \frac{1}{8\pi^2 m^2 n^2}\left[\frac{g_{mn1}}{C^*(m)C(n)} - \frac{h_{mn1}}{C(m)C(n)}\right]. \tag{27}$$

For $m, n < N/4$ it holds that

$$r_{mn1} = \frac{\max(0, \min(m,n) - |m-n|)}{2\pi^2 m^2 n^2 (N-2m)(N-2n)}. \tag{28}$$

3. Three Estimators

The low SNR estimator

For large N and low SNR, it follows from (23) that $\mathbf{R}$ is approximately diagonal, given by

$$\mathbf{R} \simeq \frac{1}{4\pi^2\mathrm{SNR}^4 N}\mathrm{diag}\left(1, \frac{1}{2^4}, \cdots, \frac{1}{J^4}\right). \tag{29}$$

Inserting (29) into (9) gives

$$\hat{\alpha} = \frac{30\sum_{k=1}^{J} k^4\,\hat{\Phi}(k)}{J(J+1)(2J+1)(3J^2+3J-1)}. \tag{30}$$

The high SNR estimator

For high SNR, the elements of $\mathbf{R}$ are approximately equal to $r_{mn1}\mathrm{SNR}^{-1}$, where r_{mn1} is given by (27). Note that both this and the above estimator assign the largest weight to $\hat{\Phi}(J)$. This fact motivates the simple estimator proposed below.

A simple estimator

Consider the simple estimator

$$\hat{\alpha} = \hat{\Phi}(J) \tag{31}$$

where $\hat{\Phi}(J)$ is calculated, for example, using (6)-(7). For $J < N/4$, the asymptotic variance of $\hat{\alpha}$ for $\mathrm{SNR}/N \gg 1$ directly follows from (27)

$$\mathrm{var}[\hat{\alpha}] = \frac{r_{JJ1}}{\mathrm{SNR}} = \frac{1}{2\pi^2\mathrm{SNR}J^3(N-2J)^2} \tag{32}$$

The variance (32) is analytically minimized for $J = 3N/10$, for which the quotient of the error variance divided by the CRB is $115/90 \approx 1.27$.

Robust phase unwrapping

A significant improvement of the algorithm performance at low SNR scenarios can be achieved using an alternative robust unwrapping scheme in (6) and (7), that proceeds recursively: In (7), $\hat{\Phi}(1)$ is replaced with $\bar{\Phi}_m$, which is an average of the previous estimates $\{\hat{\Phi}(k)\}_{k=1}^{m-1}$. The quantity $\bar{\Phi}_m$ is calculated as the aritmetic mean of $\{\hat{\Phi}(k)\}_{k=1}^{m-1}$, for $k > 4$ with the addition

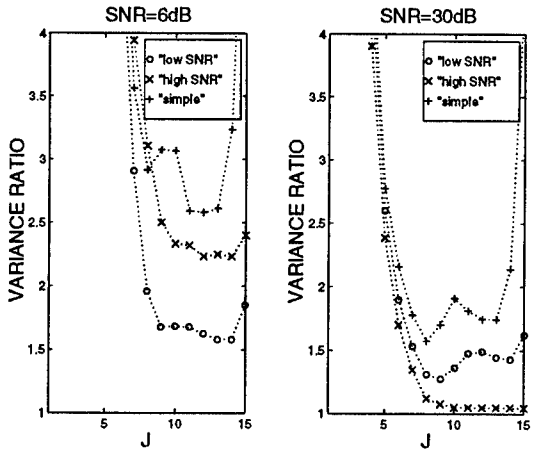


Figure 1. The quotient of the error variance divided by the CRB as a function of the design variable J.

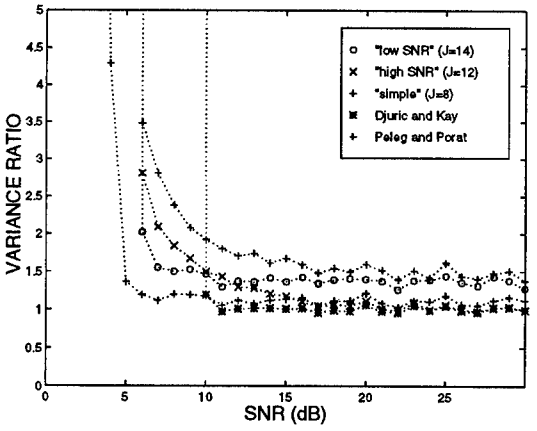


Figure 2. The quotient of the error variance divided by the CRB as a function of SNR.

that the largest and the smallest terms are excluded.

Numerical Examples

1000 realizations, each of length $N = 31$, are generated for the model (1) with $A = e^{i\mu}$ where μ is uniformly distributed in $[0, 2\pi]$, $f = 0.3$, and $\alpha = 0.1$. The performance for SNR=6dB and SNR=30dB is studied in Fig. 1, where the efficiency in terms of the quotient of the error variance divided by the CRB is shown versus J.

From the curves in Fig. 1 one can note that for low (high) SNRs the "low SNR" ("high SNR") estimator outperforms both the other methods in terms of efficiency. As expected from (32), the performance of the "simple" estimator is inferior compared with the efficiency of the other methods. For SNR=30dB the empirical efficiency is close to the predicted theoretical result r_{JJ1}/SNR. In general, the curves in Fig. 1 have two minima with respect to J.

The results from a similar experiment with performance versus SNR are given in Fig. 2. Here, $J = 14$ for the "low SNR" estimator, $J = 12$ for the "high SNR" estimator, and $J = 8$ for the "simple" estimator. Also, the frequency rate estimators in [3] and [4] are considered. The SNR threshold for the three estimators is 4dB below the threshold for the estimator in [3], however the latter estimator marginally performs better in the SNR range between 10dB and 20dB. For the estimator in [3] the threshold is sharp at 10dB, while for the estimators proposed in this paper there is a graceful degradation in performance for low SNRs above the threshold at 6dB. The SNR threshold for the method in [4] is 4dB and thus this method has the lowest threshold among the considered estimators. For high SNR its efficiency is close to the predicted theoretical result, that is slightly inferior performance compared to the "high SNR" estimator and the estimator in [3].

4. Conclusions

Three algorithms for estimating the frequency rate of a noisy complex-valued linear FM signal have been proposed, and their performances have been characterized. The methods rely on phase angle calculations of forth order sample moments of the noisy signal. It has been demonstrated that the performance of the proposed methods is nearly optimal for proper choices of the design variable J. The proposed algorithms require no numerical search, and can achieve lower SNR threshold than the algorithm in [3].

References

[1] T. Abatzoglou. Fast maximum likelihood joint estimation of frequency and frequency rate. *IEEE Trans. Aerospace and Electronic Systems*, 22:708–714, 1986.

[2] B. Boashash. Estimating and interpreting the instantanous frequency of a signal - part 2: algorithms and applications. *Proc. IEEE*, 80(4):540–568, 1992.

[3] P. Djurić and S. Kay. Parameter estimation of chirp signals. *IEEE Trans. ASSP*, 38(12):2118–2126, 1990.

[4] S. Peleg and B. Porat. Linear fm signal parameter estimation from discrete-time observations. *IEEE Trans. Aerospace and Electronic Systems*, 27(4):607–616, 1991.

[5] T. Söderström and P. Stoica. *System Identification.* Prentice-Hall International, Hemel Hempstead, UK, 1989.

[6] P. Tichavský and A. Swami. Statistical characterization of sample fourth-order cumulants of a noisy complex sinusoidal process. *IEEE Trans. Signal Processing*, 43(7):1620–1630, 1995.

Theoretical properties of the "Caterpillar" method of time series analysis

Vladimir Nekrutkin
Dept. of Mathematics, St.Petersburg University
Bibliotechnaya sq. 2, 198904, Russia
Fax +7(812)4286649, e-mail: com@trend.niimm.spb.su

Abstract

The work is devoted to the mathematical theory of the "Caterpillar" method which has proved to be a very powerful tool of analysis of time series. This method is based on the use of the principal component analysis technique applied to a multivariate sample which is obtained from the initial sample by the method of delays. A natural language used to analyse the method is the Hilbert-Schmidt operator theory. We give conditions when two deterministic functions are completely separated from each other for a finite period of observations. We also show that under mild conditions any deterministic function can be asymptotically separated from any ergodic random noise.

1. Introduction

The work is devoted to the mathematical theory of a method of time series analysis. A natural language used to analyse the method is the Hilbert-Schmidt operator theory.

Let f be a function on $[0,t]$. We assume that this function belongs to a Hilbert space H and is a realization of some, perhaps random, process, characteristics of which are unknown. We also assume that f can be represented as a sum of several functions f_i with every function related to a certain effect or to the noise component. We thus seek for an expansion

$$f = \sum_i f_i\,, \qquad (1)$$

where the terms f_i are "interpretable" and "independent".

We do not assume any parametric model for f and therefore the regression analysis technique can not be used to get (1). "Independence" of f_i in (1) can sometimes be achieved through an expansion of f with respect to an orthogonal basis, but the selection problem of the basis presents big difficulty and could not be uniquely resolved.

The main essence of the present approach is that the functions f_i are constructed through f itself. More precisely, in the case of discrete time, f_i are related to principal components of a multivariate sample generated from f by introducing the lag, or delay, variable.

2. Description of the Principal Scheme

Let $(f_i)_{i=1}^N$ be a numerical sequence, or time series, and let τ, $1 < \tau < N$, be an integer. Define a collection of τ-dimensional vectors $X^{(1)},\ldots,X^{(n)}$, $n = N-\tau+1$, by the formula $X^{(k)} = (x_{i,k})_{i=1}^{\tau}$, where $x_{i,k} = f_{k+i-1}$, and define the matrix

$$\mathbf{X} = (X^{(1)},\ldots,X^{(n)}) = \begin{pmatrix} f_1 & f_2 & \cdots & f_n \\ f_2 & f_3 & \cdots & f_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ f_\tau & f_{\tau+1} & \cdots & f_N \end{pmatrix}.$$

This matrix will be called the noncentered matrix of delay observations. Define the mean vector $\bar{X} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_\tau)^T$, where

$$\bar{x}_i = \frac{1}{n}\sum_{i=1}^{n} f_{s+i-1}, \quad i = 1,\ldots,\tau.$$

Subtracting this vector from each of $X^{(1)},\ldots,X^{(n)}$ we get the matrix $\mathbf{Y}$ of centered vectors $Y^{(1)},\ldots,Y^{(n)}$.

Consider now the covariance matrix of the vectors $Y^{(1)},\ldots,Y^{(n)}$ considered as a n-sample of τ-dimensional vectors and apply the principal component

0-8186-7576-4/96 $5.00

method to this sample. Let

$$\mathbf{V}_X = \left(v^{(1)}, v^{(2)}, \ldots, v^{(\tau)}\right) = \begin{pmatrix} v_1^{(1)} & v_1^{(2)} & \ldots & v_1^{(\tau)} \\ v_2^{(1)} & v_2^{(2)} & \ldots & v_2^{(\tau)} \\ \vdots & \vdots & \ddots & \vdots \\ v_\tau^{(1)} & v_\tau^{(2)} & \ldots & v_\tau^{(\tau)} \end{pmatrix}$$

be the matrix of eigen–vectors of the covariance matrix of $Y^{(1)}, \ldots, Y^{(n)}$.

The standard for the principal component analysis operations of computing principal components:

$$U_X = \mathbf{V}_X^T \mathbf{Y} = (U_1, \ldots, U_\tau)^T$$

and reconstruction of the initial (centered) sample based on a selected number r of principal components:

$$\tilde{\mathbf{Y}} = (v^{(i_1)}, \ldots, v^{(i_r)}) \begin{pmatrix} U_{i_1}^T \\ \vdots \\ U_{i_r}^T \end{pmatrix} = \sum_{s=1}^{r} v^{(i_s)} U_{i_s}^T,$$

where $0 < i_1 < i_2 < \ldots < i_r \leq \tau$, can be applied as usual. After reconstruction of the matrix $\tilde{\mathbf{X}} = \tilde{\mathbf{Y}} + \bar{\mathbf{X}}$, where $\bar{\mathbf{X}}$ - is a matrix with columns equals $\bar{X}$, the initial sequence is reconstructed by averaging over the diagonals of $\tilde{\mathbf{X}}$:

$$\tilde{f}_s = \begin{cases} \frac{1}{s} \sum_{i=1}^{s} \tilde{x}_{i,s-i+1} & 1 \leq s \leq \tau, \\ \frac{1}{\tau} \sum_{i=1}^{\tau} \tilde{x}_{i,s-i+1} & \tau \leq s \leq n, \\ \frac{1}{N-s+1} \sum_{i=1}^{N-s+1} \tilde{x}_{i+s-n,n-i+1} & n \leq s \leq N. \end{cases}$$

Thus, we have presented a method of analysis of time series which we call "caterpillar". The first reference to this notation and also the first numerical examples comes back to [1]. Note, that similar method was investigated from the geometrical point of view in [2]. Our assumption differ from the mentioned above by the successive application of the methods of functional analysis. We have applied this method to many practical problems and the method proved to be very powerful in analysis of time series, particularly nonstationary and short, with the value of N starting at 20. (It is well known that such time series are hard to analyse.) The method has been generalized to multivariate time series and random fields. In the next section we present some results concerning the theoretical properties of the method.

3. Some Results of Theoretical Study

Let us describe a method in the case of the functions of continuous argument. We first need to transform the interval $[0, t]$ into a rectangular π and "transfer" f from $[0, t]$ to π. Formally this operation can be defined with the help of a mapping $\theta : \pi \to [0, t]$ and consideration of the two-variate function $g = f \circ \theta$ instead of f. In the standard variant of the additive caterpillar, we have $\theta(x, s) = x + s$, $(x, s) \in \pi = [0, \tau] \times [0, t - \tau]$ and $g(x, s) = f(x + s)$.

Function g can often be considered as the kernel of an integral operator which happens to be the Hilbert-Schmidt operator and possesses therefore a number of attractive features. In particular g can be expanded with respect to two orthogonal sequences of the base functions.

$$g = \sum_n \psi_n \otimes \varphi_n \tag{2}$$

Selecting several terms in the expansion and projecting the two-dimensional function back to $[0, t]$ we thus get one or several terms in (1) which should then be interpreted and analysed.

Let us introduce one notation. Functions f_1 and f_2 are separeted in bi-ortogonal expansion (2) if the corresponding fields $g^{(1)}$ and $g^{(2)}$ are satisfying to the following conditions

$$\int g^{(1)}(x, s)\, g^{(2)}(y, s)\, ds = 0,$$

for almost all (x, y), and

$$\int g^{(1)}(x, u)\, g^{(2)}(x, v)\, dx = 0,$$

for almost all (u, v).

We formulate different features of the method, considering for simplicity only the additive case $\theta(x, s) = x + s$, $(x, s) \in \pi = [0, \tau] \times [0, t - \tau]$.

When f_1 and f_2 are continuous functions then the separability condition under mild conditions for functions f_1 and f_2 follows from the following two equalities:

$$\int_0^\tau f_1(\delta + z) f_2(z)\, dz = 0$$

for any $\delta \in [0, t - \tau]$,

$$\int_0^{t-\tau} f_1(\alpha + y) f_2(y)\, dy = 0$$

for every $\alpha \in [0, \tau]$, and two "periodicity conditions":

$$f_1(\delta + \tau + v) f_2(\tau + v) = f_1(\delta + v) f_2(v),$$

for all $v \in [0, t - \tau]$ and $\delta \in [-v, t - \tau - v]$.

$$f_1(\alpha + t - \tau + y) f_2(t - \tau + y) = f_1(\alpha + y) f_2(y)$$

for all $y \in [0, \tau]$ and $\alpha \in [-y, \tau - y]$.

Assume now that $(\Omega, \mathcal{F}, \mathbf{P})$ be some probability space, $f_1(s)$, $f_2(s)$ be two random processes defined for $s \geq 0$. Define $f = f_1 + f_2$ and letting $\tau = \tau(t) \in (0, t)$ define the random fields $g(x, y) = f(x+y)$, $g^{(1)}(x, y) = f_1(x + y)$ and $g^{(2)}(x, y) = f_2(x + y)$. Let us consider the noncentered correlation coefficients $\rho_1^{(t)}$ and $\rho_2^{(t)}$ between the random fields $g^{(1)}$ and $g^{(2)}$. We shall say that random processes f_1 and f_2 are stochastically separable when $t \to \infty$ if the correlation coefficients $\rho_1^{(t)}(u, v)$ and $\rho_2^{(t)}(x, y)$ converge in probability to 0 when $t \to \infty$ for almost all (u, v) and (x, y).

Let now $f(s)$, $s \geq 0$ be some deterministic function from $L^2[[0, \infty)]$ and $\xi(s)$ be a random process interpreted as a "pure noise". Let $\mathbf{E}\xi(s) = 0$ for any s and $R_\xi(x, y)$ be covariance function of $\xi(s)$. Assume (i) for any $s \in [0, \infty)$ there exists $\delta = \delta(s) > 0$ such that

$$\mathbf{P}\left(\int_0^T \xi^2(x + s)\, dx/T < \delta\right) \to 0$$

for $T \to \infty$; and (ii) for any $u, v \in R$

$$\frac{1}{T^2} \int_0^T dx \int_0^T dy\, \tilde{f}_T(x,u) \tilde{f}_T(y,u) R_\xi(x+v, y+v) \to 0 \qquad (3)$$

when $T \to \infty$, where

$$\tilde{f}_T(x, s) = f(x + s) / \sqrt{\frac{1}{T} \int_0^T f^2(y + s)\, dy}\,.$$

The following theorem holds.

Theorem. If $\tau \to \infty$, $t - \tau \to \infty$, and the conditions (i) and (ii) hold then f and ξ are stochastically separable when $t \to \infty$.

Note that (i) and (ii) hold if the process $\xi(s)$ is stationary,

$$\frac{1}{T} \int_0^T \xi^2(s)\, ds \xrightarrow{P} R_\xi(0) > 0$$

when $T \to \infty$,

$$\frac{1}{T^2} \int_0^T dx \int_0^T dy\, |R_\xi(x - y)| \to 0\,,$$

and f is bounded and satisfies

$$\liminf_{T \to \infty} \frac{1}{T} \int_0^T f^2(x + s)\, dx > 0.$$

If f is unbounded then the sufficient conditions become different. For example, if f is linear then (3) can be rewritten as

$$\frac{1}{T^2} \int_0^T dx \int_0^T dy\, (x + 1)\, (y + 1)\, |R_\xi(x - y)| = o(T^2)\,,$$

$T \to \infty$.

Theoretical results are in a very good agreement with the numerical results: when the number of observations is large then simple trend functions, like exponential and trigonometric functions, can usually be explicitly seen in the first components of (1).

4 Multiplicative "Caterpillar"

In this section we consider the multiplicative variant of the "Caterpillar" method. Although this case does not have a great partical value, nevertheless it shows that the choice of the function $\theta(x, s)$ differing from $x + s$ can also lead to interesting results.

Let f be given on the interval $[-t, t]$, the sets D_1 and D_2 have the form $D_1 = [\tau, \tau]$, $D_2 = [-t/\tau, t/\tau]$, with $\tau \in (0, t)$ and $\theta(x, s) = xs$, $x \in D_1$, $s \in D_2$. In the noncentered case, that is when $\mathbf{Y} = \mathbf{X}$, $g(x, s) = f(xs)$.

The separation conditions can be written as

$$\int_{-\tau}^{\tau} f_1(xu) f_2(xv)\, dx = 0 \qquad (4)$$

for any $u, v \in [-t/\tau, t/\tau]$ and

$$\int_{-t/\tau}^{t/\tau} f_1(xs) f_2(ys)\, ds = 0$$

for any $x, y \in [-\tau, \tau]$.

This implies that if one of functions f_1 or f_2 is an even function and another is an odd one, then these two functions are always separable. It is also possible to demonstrate that the above condition is not only sufficient, but also necessary.

References

[1] Belonin M.D., Tatarinov E.V., Kalinin O.M., Shimansky V.K., Beskrovnaya O.V., Gransky V.V., Pohitonova T.E., Factor analysis in oil geology, Moscow, VIZMC, 1971.

[2] Buchstaber, V.M., Time Series Analysis and Grassmannians. *Amer. Math. Soc. Trans.* (2) Vol. **162**, 1994.

Structurally Robust Weak Continuity

N. D. Sidiropoulos, J. S. Baras, and C. A. Berenstein
Institute for Systems Research
University of Maryland
College Park, MD 20742
nikos@glue.umd.edu

Abstract

We pose the following optimization: Given $\mathbf{y} = \{y(n)\}_{n=0}^{N-1} \in \mathbf{R}^N$, *find a finite-alphabet* $\widehat{\mathbf{x}} = \{\widehat{x}(n)\}_{n=0}^{N-1} \in \mathcal{A}^N$, *that minimizes* $d(\mathbf{x},\mathbf{y}) + g(\mathbf{x})$ *subject to:* $\mathbf{x}$ *satisfies a hard structural (syntactic) constraint, e.g.,* $\mathbf{x}$ *is piecewise constant of plateau run-length* $\geq M$, *or locally monotonic of lomo -degree* α. *Here,* $d(\mathbf{x},\mathbf{y}) = \sum_{n=0}^{N-1} d_n(y(n),x(n))$ *measures fidelity to the data, and is known as the* noise term, *and* $g(\mathbf{x}) = \sum_{n=1}^{N-1} g_n(x(n),x(n-1))$ *measures smoothness-complexity of the solution. This optimization represents the unification and outgrowth of several digital nonlinear filtering schemes, including, in particular, digital counterparts of* Weak Continuity (WC) *[6, 7, 2], and* Minimum Description Length (MDL) *[4] on one hand, and nonlinear regression, e.g.,* VORCA *filtering [11], and* Digital Locally Monotonic Regression *[10], on the other. It is shown that the proposed optimization admits efficient Viterbi-type solution, and, in terms of performance, combines the best of both worlds.*

1 Introduction

One of the classic problems in the true spirit of nonlinear filtering is that of detecting and estimating edges in noise. Among the great many approaches proposed so far, a particularly noteworthy one is the (nonconvex) variational *Weak Continuity (WC)* paradigm of Mumford-Shah [6, 7] and Blake-Zisserman [2] (see also the excellent recent book by Morel and Solimini [5]). Weak continuity is a rigorous paradigm for edge detection, which attempts to fit *piecewise*-smooth candidate "interpretations" to the observable data (thus the term *weak* continuity).

In real life we nowadays most often deal with digital data, i.e., sequences of finite-alphabet variables. Following Blake and Zisserman [2], we present a digital version of discrete-time WC. Given a (generally real-valued) sequence of finite extent $\mathbf{y} = \{y(n)\}_{n=0}^{N-1} \in \mathbf{R}^N$, the problem is to find a finite-alphabet sequence, $\widehat{\mathbf{x}} = \{\widehat{x}(n)\}_{n=0}^{N-1} \in \mathcal{A}^N$ (the "reproduction process"), that minimizes

$$\sum_{n=0}^{N-1}(y(n)-x(n))^2 + \sum_{n=1}^{N-1} h_{\alpha,\lambda_{WC}}(x(n)-x(n-1))$$

where

$$h_{\alpha,\lambda_{WC}}(t) = \begin{cases} \lambda_{WC}^2 t^2 & , t^2 < \frac{\alpha}{\lambda_{WC}^2} \\ \alpha & , otherwise \end{cases}$$

There exist essentially two ways to go about solving this problem: Dynamic Programming (DP) [1], and the so-called *Graduated Non Convexity (GNC)* algorithm [2]. For one-dimensional data, DP is probably the best way to go. According to Blake and Zisserman [2], Papoulias [8] was the first to implement a DP WC algorithm. The drawback of DP is that it does not generalize in higher dimensions, for lack of total ordering. The GNC, by comparison, carries over quite effortlessly in higher dimensions.

A related optimization has been advocated by Leclerc [4], based on the *Minimum Description Length (MDL)* principle of Rissanen; the goal is the minimization of:

$$\sum_{n=0}^{N-1}\frac{(y(n)-x(n))^2}{\sigma^2} + \sum_{n=1}^{N-1}\lambda_{MDL}\left[1-\delta(x(n)-x(n-1))\right]$$

where δ is the Kronecker delta function, and σ^2 is noise variance. Here, $\lambda_{MDL} \geq 0$.

2 Unification and Motivation

Both WC and MDL seek to minimize a nonconvex cost of the following general form

$$\mathcal{V}(\mathbf{y},\mathbf{x}) = \sum_{n=0}^{N-1} d_n(y(n),x(n)) + \sum_{n=1}^{N-1} g_n(x(n),x(n-1))$$

In the digital world, Leclerc's MDL formulation is a special case of WC. Indeed, if λ_{WC} is sufficiently large

0-8186-7576-4/96 $5.00

(i.e., $\lambda_{WC}^2 > \alpha$), then, t being integer, $h_{\alpha,\lambda_{WC}}(t) = \alpha\left[1 - \delta(t)\right]$. If, in addition, $\alpha = \lambda_{MDL}\sigma^2$, then WC reduces to Leclerc's MDL approach.

Both WC, and Leclerc's MDL approach are powerful and meritorious paradigms; however, both share a nontrivial shortcoming: they are not robust with respect to outliers, in the sense of being susceptible to noise-induced "impulses". Consider a single such outlier, i.e., a Kronecker delta of height Δ. If $(\frac{\Delta}{\sigma})^2 > 2\lambda_{MDL}$, then Leclerc's MDL approach will preserve this "impulse"; similarly, if $\Delta^2 > \frac{\alpha}{\lambda_{WC}^2}$, and $\Delta^2 > 2\alpha$, then WC will also preserve it. Observe that these statements should be interpreted as follows: for each given choice of respective optimization parameter(s), one can find a sufficiently large Δ which forces both "filters" to preserve "impulses" of height $\geq \Delta$. In the context of edge detection in impulsive noise, this behavior is undesirable; these "impulses", no matter how powerful, should not be preserved [12].

"Traditional" nonlinear filters (e.g., the root of the median) are robust with respect to outliers. This robustness stems from the fact that the implicit goal of these filters is to enforce (albeit suboptimally) "hard" structural (syntactic) constraints on the data, e.g., of the type $\mathbf{x}$ is piecewise constant of plateau run-length $\geq M$, or locally monotonic of lomo-degree α. How to optimally enforce such constraints has been the subject of previous work by the first author in so-called VORCA filtering [11] and digital locally monotonic regression [10]. VORCA filtering amounts to solving:

$$\textbf{minimize} \sum_{n=0}^{N-1} d_n(y(n), x(n))$$

$$\textbf{subject to} : \mathbf{x} = \{x(n)\}_{n=0}^{N-1} \in P_M^N$$

where P_M^N is the set of all sequences of N elements of $\mathcal{A}$ which are piecewise constant of plateau (run) length $\geq M$.

A real-valued sequence (string), $\mathbf{x}$, of length N, is *locally monotonic* of degree $\alpha \leq N$ (or *lomo-*α, or simply *lomo* in case α is understood) if each and every one of its substrings of α consecutive symbols is monotonic. Local monotonicity appears in the study of the set of root signals of the median filter [3]; it constraints the roughness of a signal by limiting the rate at which the signal undergoes changes of trend (increasing to decreasing or vice versa). In effect, it *limits the frequency of oscillations, without limiting the magnitude of jump level changes that the signal exhibits.* Local monotonicity implies a different notion of smoothness, as compared to e.g., limiting the support of the Fourier transform; the latter imposes a limit on *both* the frequency of oscillations, *and* the magnitude of jump level changes.

In [9], Restrepo and Bovik developed an elegant mathematical framework in which they studied locally monotonic regressions in $\mathbf{R}^N$. *Digital* locally monotonic regression has been proposed in [10], and it amounts to solving:

$$\textbf{minimize} \sum_{n=0}^{N-1} d_n(y(n), x(n))$$

$$\textbf{subject to} : \mathbf{x} = \{x(n)\}_{n=0}^{N-1} \in \Lambda(\alpha, N, \mathcal{A})$$

where $\Lambda(\alpha, N, \mathcal{A})$ is the set of all sequences of N elements of $\mathcal{A}$ which are locally monotonic of lomo-degree α [10]. Both approaches are robust, in the sense of suppressing impulse-like inputs, while retaining "true" (consistent) edge signals. However, both do not take into account the significance of level changes ("discontinuities") in the solution, i.e., they may declare an edge even when the two resulting levels are very close. This is often undesirable; and it happens exactly because the latter two approaches do not *explicitly* account for smoothness/complexity, i.e., unlike WC, they do not incorporate a "soft" smoothness/complexity penalty into the cost function.

3 Structurally Robust Weak Continuity

It appears quite natural, then, to combine the power of WC with the appeal and demonstrated effectiveness of "hard" structural constraints, and propose the minimization of:

$$\sum_{n=0}^{N-1} d_n(y(n), x(n)) + \sum_{n=1}^{N-1} g_n(x(n), x(n-1))$$

$$\textbf{subject to} : \mathbf{x} \in \mathcal{S}$$

where $\mathcal{S}$ is the set of all sequences of N elements of $\mathcal{A}$ satisfying some local "hard" structural constraint. Here, again, $d(\mathbf{x}, \mathbf{y}) = \sum_{n=0}^{N-1} d_n(y(n), x(n))$ is a fidelity measure, and $g(\mathbf{x}) = \sum_{n=1}^{N-1} g_n(x(n), x(n-1))$ is a smoothness-complexity measure. We will refer to this optimization as *Structurally Robust Weak Continuity (SR-WC).* When $\mathcal{S} = P_M^N$, *Runlength-Constrained Weak Continuity (RC-WC)* results; similarly, if $\mathcal{S} = \Lambda(\alpha, N, \mathcal{A})$, then *Locally Monotonic Weak Continuity (LM-WC)* results. Both retain the unique merits of WC, are robust with respect to outliers, take complexity into consideration, and admit efficient Viterbi-type solution. In fact, RC-WC, and LM-WC can be solved using exactly the same resources and computational structures as VORCA, and digital locally monotonic regression, respectively [12]. The extension to weak continuity (i.e., the incorporation of the first-order smoothness-complexity measure $g(\mathbf{x}) = \sum_{n=1}^{N-1} g_n(x(n), x(n-1))$ into the cost functional) essentially comes "for free", due to the structure of the Viterbi Algorithm. The resulting complexity of RC-WC, LM-WC is $O((|\mathcal{A}|^2 + |\mathcal{A}|(M-1))N)$, $O(|\mathcal{A}|^2 \alpha N)$, respectively.

By virtue of the above, efficient computation of SR-WC can be taken for granted. What is intriguing and unexplored is how to go about choosing fidelity and smoothness/complexity measures. We know that, at least for some specific choices, e.g., "classic" WC, MDL, or VORCA, we may expect very good nonlinear filtering results. The question is, can we make even better choices, and in what sense. This is partially explored in the following.

4 Example

This particular example demonstrates the effectiveness of simple RC-WC. Figure 1 depicts a typical input sequence. This particular input has been generated by adding i.i.d. noise on some artificial "true" noise-free test data. The noise has been generated according to a mixture of a uniform distribution and an "outlier" distribution, the mixture being heavily weighted in favor of the uniform distribution, and most of the data points are contaminated. It should be stressed that we do not utilize our exact knowledge of the noise model to fully match the optimization to the noise characteristics, which is certainly a possibility [11, 10, 9]. Instead, as it will be explained shortly, we only use some crude noise measurements to help us pick reasonable values for two optimization parameters. The noise-free test data has not been reproduced on its own, due to space limitations; instead, it has been overlaid on the restoration plots, using a dashed line. This is meant to help the reader judge filtering "quality".

Here we take $d_n(y(n),x(n)) = |y(n) - x(n)|$, $\forall n \in \{0,1,\cdots,N-1\}$, and $g_n(x(n),x(n-1)) = \lambda^2_{WC}\left[1-\delta(x(n)-x(n-1))\right]$, $\forall n \in \{0,1,\cdots,N-1\}$, $\mathcal{A} = \{0,\cdots,99\}$, $N = 512$, and $\mathcal{S} = P^N_M$.

For $M = 1$, we obtain "plain" WC, and the result for $\lambda^2_{WC} = 25$ is depicted in Figure 2. This is excellent filtering, yet powerful outliers are preserved. We could, in principle, further increase λ^2_{WC}, thereby eventually eliminating outliers, but, at the same time, also "mending" true edges. Clearly, this is not the way to go about ameliorating this problem, for, no matter what our choice of λ^2_{WC} is, one can always find a sufficiently powerful outlier that will fool WC.

For $\lambda^2_{WC} = 0$, we obtain "plain" VORCA, and the result for $M = 15$ is depicted in Figure 3. This too is excellent filtering, the outliers have been effectively eliminated, yet some undesirable "weak" edges still remain. For $\lambda^2_{WC} = 25$, and $M = 15$ we have "true" hybrid RC-WC, and the result is depicted in Figure 4. It is obvious that RC-WC combines the power of both methods: this is, indeed, almost perfect filtering.

One obvious objection may be anticipated: one may wonder about how we came up with the particular choices of M, λ that led to these results. In the following, we address this subject.

4.1 Selection of Optimization Parameters

We will use the following definitions. A *feature (outlying burst) of width* $w < M$ is a "short" arbitrary deviation from a plateau, consisting of a total of w perturbed samples. A *constant segment of saliency (width-strength product)* $\mu = w \cdot H$ is a (potentially long) equidistant deviation from a plateau, i.e., a string of w equal samples which differ by H from the plateau level.

The following two claims refer to *this particular instance of RC-WC*, i.e., $d_n(y(n),x(n)) = |y(n) - x(n)|$, $\forall n \in \{0,1,\cdots,N-1\}$, and $g_n(x(n),x(n-1)) = \lambda^2_{WC}\left[1-\delta(x(n)-x(n-1))\right]$, $\forall n \in \{0,1,\cdots,N-1\}$. Proofs can be found in [12].

Theorem 1 *Assume that M is odd. RC-WC eliminates all features (outlying bursts) of width $w \leq \frac{M-1}{2}$, regardless of λ^2_{WC}, and the same is true for $\lambda^2_{WC} = 0$, i.e., "plain" VORCA filtering with respect to the above choice of $d_n(\cdot,\cdot)$.*

Theorem 2 *RC-WC suppresses all constant segments of saliency (width-strength product) $\mu = w \cdot H < 2\lambda^2_{WC}$, i.e., "mends" the "weak" edges at the endpoints of such segments, and the same holds for $M = 1$, i.e., "plain" WC with respect to the above choice of $d_n(\cdot,\cdot)$, $g_n(\cdot,\cdot)$.*

The overall conclusion is that this particular instance of RC-WC suppresses features of either (i) width $w \leq \frac{M-1}{2}$ (M: odd), *regardless of strength*, or (ii) saliency (width-strength product) $\mu = w \cdot H < 2\lambda^2_{WC}$. This allows us to essentially separately fine-tune two important aspects of filter behavior. In a nutshell, M controls outlier rejection, whereas λ^2_{WC} controls residual ripple.

5 Conclusions

Motivated by the power of WC-based methods [6, 7, 2, 4], "complementary" previous work by the first author in optimal nonlinear filtering under "hard" structural (so-called syntactic) constraints [11, 10], and realizing that a potential shortcoming of WC could be ameliorated by introducing "hard" structural constraints, whereas a drawback of the methods of [11, 10] could be rectified by introducing "soft" weak continuity constraints, we have posed, solved, and demonstrated the effectiveness of a novel hybrid optimization, dubbed Structurally Robust Weak Continuity, combining the advantages while avoiding the shortcomings of its constituent elements. SR-WC includes its constituent elements as special cases, and inherits efficient Viterbi implementation from [11, 10]. What is most intriguing is how to go about choosing fidelity and smoothness/complexity measures. This deserves further investigation, and long-term research in this direction is currently underway.

References

[1] R. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, N.J., 1957.

[2] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, Mass., 1987.

[3] N. Gallagher. Median filters: a tutorial. In *Proc. IEEE Int. Symp. Circ., Syst., ISCAS-88*, pages 1737–1744, 1988.

[4] Y. Leclerc. Constructing Simple Stable Descriptions for Image Partitioning. *Int. J. Computer Vision*, 3(1):73–102, 1989.

[5] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Birkhauser, Boston-Basel-Berlin, 1994.

[6] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, 1985.

[7] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Math.*, 42:577–685, 1989.

[8] A. Papoulias. Curve Segmentations Using Weak Continuity Constraints. M.Sc. thesis, Univ. of Edinburgh, 1985.

[9] A. Restrepo and A. C. Bovik. Locally Monotonic Regression. *IEEE Trans. Signal Processing*, 41(9):2796–2810, Sep. 1993.

[10] N. Sidiropoulos. Fast Digital Locally Monotonic Regression. Submitted for publication, IEEE Trans. Signal Processing. Sum mary to appear in Proc. 1996 IEEE Int. Symp. on Circuits and Systems, May 12-15, Atlanta, GA.

[11] N. Sidiropoulos. The Viterbi Optimal Runlength-Constrained Approximation Nonlinear Filter. *IEEE Trans. Signal Processing*, 44(3), March 1996.

[12] N. Sidiropoulos, J. Baras, and C. Berenstein. Structurally Robust Weak Continuity. Submitted for publication, IEEE Trans. Signal Processing.

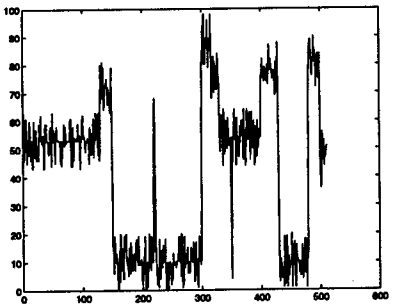


Figure 1. Input sequence, $\{y(n)\}_{n=0}^{511}$

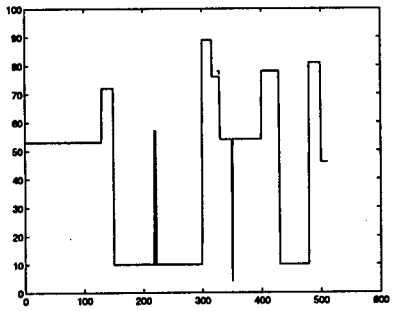


Figure 2. Output of digital WC, $\lambda_{WC}^2 = 25$

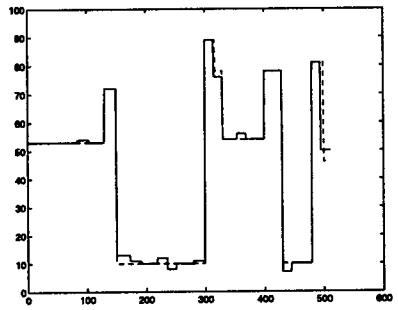


Figure 3. Output of VORCA, $M = 15$

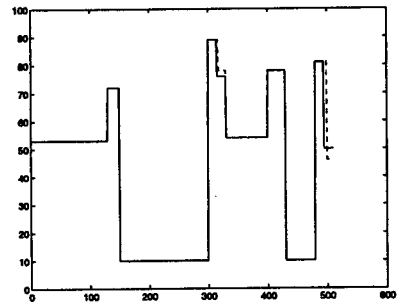


Figure 4. Output of RC-WC, $M = 15$, $\lambda_{WC}^2 = 25$

Continuous and Discrete Time Filters for Markov Jump Linear Systems with Gaussian Observations

Vikram Krishnamurthy and Jamie Evans *
Department of Electrical and Electronic Engineering
University of Melbourne
Parkville, Victoria 3052, Australia

vikram@mullian.ee.mu.oz.au, jamse@mullian.ee.mu.oz.au

Abstract

We present new finite dimensional filters for estimating the state of Markov jump linear systems, given noisy measurements of the Markov chain. Discrete time as well as continuous time models are considered. A robust version of the continuous time filters is used to derive a discretization which links the continuous and discrete time results. Simulations compare the robust discretization with direct numerical solutions of the filtering equations. The new filters have applications in the passive tracking of maneuvering targets and speech coding.

1. Introduction

Consider a discrete-time Markov jump linear system whose (vector) state equation evolves as:

$$s_n = C(X_{n-1})\, s_{n-1} + v_n$$

where X_n denotes a finite state homogeneous Markov chain and v_n is a zero mean stochastic process which is independent of the process X_n. Assume that we have noisy measurements y_n of the Markov chain X_n in white Gaussian noise. In this paper we show how to compute filtered estimates $\hat{s}_n$ of the state s_n, i.e., $\hat{s}_n = \mathbf{E}\{s_n|\mathcal{Y}_n\}$ where $\mathcal{Y}_n$ denotes the filtration generated by the observations.

Instead of noisy measurements of the Markov chain X_n, suppose that only noisy measurements of s_n are available. In such a case, it is well known that the optimal state filter is infinite dimensional [1]. Indeed the optimal state estimates would involve a computational cost that is exponential in the data length. Sub-optimal finite dimensional approximations are given in [1]. *However, as we show in this paper, given noisy observations y_n of the Markov chain, the optimal state filter for s_n is finite dimensional.* We also derive continuous-time versions of the filters.

The key contributions of this paper can be summarized as follows:

*This work was partially supported by ATERB and ARC grants, the Cooperative Research Centre for Sensor, Signal and Information Processing (CSSIP) and a Telstra Research Laboratories Postgraduate Fellowship.

1. **Finite Dimensional Filters**: In Sec. 2 we derive finite dimensional filters for state estimation of discrete-time Markov jump linear systems given noisy observations of the Markov chain. These derivations are based on the reference probability method and thus lead to filtering equations in unnormalized or Zakai form. Finite dimensional filters are presented for the state estimation problem in continuous-time in Sec. 3.
2. **Robust Discretization**: Having derived both continuous and discrete-time filters independently, our next contribution is to show that an appropriate *robust* discretization of the continuous-time filters results in the discrete-time filters. This is the subject of Sec. 4.
3. **Numerical Examples**: Using computer simulations in Sec. 5, we compare the performance of robust discretized filters with two standard numerical approximations, namely the Euler-Maruyama and Milstein algorithms. The robust scheme is seen to outperform these methods as the discretization step size is increased.

2. Discrete-Time Filters

2.1. Signal Model and Aim

Let X_l, $l \in Z^+ = \{1, 2, \dots, \}$ denote a S-state discrete-time Markov chain defined on a probability space $(\Omega, \mathcal{F}, P)$ with state space $\{e_1, \dots, e_S\}$ where e_i denotes the unit S-vector with 1 in the ith position. Denote the transition probabilities $a_{ji} = P(X_n = e_j|X_{n-1} = e_i)$ and A for the $S \times S$ matrix (a_{ji}), $1 \le i, j \le S$. Note that $\sum_{j=1}^{S} a_{ji} = 1$ for $1 \le i \le S$.

Consider the following jump linear system driven by X_n:

$$s_n = C(X_{n-1})\, s_{n-1} + v_n \tag{1}$$

where $s_n, v_n \in \mathcal{R}^M$ and v_n is a is a zero mean process independent of the Markov chain X_n. Assume that X_n is observed indirectly via the scalar process y_n as follows:

$$y_n = \langle g, X_n \rangle + w_n \tag{2}$$

where $g = (g_1\ g_2 \dots g_S)'$ is the vector of levels of the Markov chain. Also $\langle \cdot, \cdot \rangle$ denotes the scalar product in $\mathcal{R}^S$. w_n is

0-8186-7576-4/96 $5.00 © 1996 IEEE

white Gaussian noise with variance κ^2 independent of the processes X_n and v_n.

For any $n \in Z^+$, let $\mathcal{F}_n$ denote the sigma field generated by X_m, $m \leq n$. Let $\mathcal{Y}_n$ denote the sigma field generated by y_m, $m \leq n$. Let $\mathcal{G}_n = \mathcal{F}_n \bigvee \mathcal{Y}_n$, i.e., the sigma field generated by $\{X_m, y_m\}$, $m \leq n$.

For any measurable process $\{\phi_n\}$, let $\hat{\phi}_n = \mathbf{E}\{\phi_n|\mathcal{Y}_n\}$ where $\mathbf{E}$ denotes expectation under measure P.

Aim: For fixed known values of A, g and κ, and of the initial state s_0, compute the filtered estimates $\hat{s}_n = \mathbf{E}\{s_n|\mathcal{Y}_n\}$.

2.2. Zakai State Filter for Jump Linear System

Define the probability measure P_0 such that the $\mathcal{G}_{n-1}$ restriction of the Radon-Nikodym derivative of P with respect to P_0 is

$$\frac{dP}{dP_0}\Big|_{\mathcal{G}_{n-1}} = \Lambda_n = \prod_{m=1}^{n} \exp\left(\frac{-(\langle g, X_m\rangle^2 - 2\, y_m \langle g, X_m\rangle)}{2\,\kappa^2}\right)$$

If ϕ_n, $n \in Z^+$ is a measurable sequence, then an abstract version of Bayes theorem states

$$\hat{\phi}_n = \mathbf{E}\{\phi_n|\mathcal{Y}_n\} = \frac{\mathbf{E}_0\{\Lambda_n\, \phi_n|\mathcal{Y}_n\}}{\mathbf{E}_0\{\Lambda_n|\mathcal{Y}_n\}}$$

where $\mathbf{E}_0$ denotes expectation with respect to P_0. Define the *un-normalized conditional expectation* $\sigma_n(\phi_n) = \mathbf{E}_0\{\Lambda_n\, \phi_n|\mathcal{Y}_n\}$ and let

$$b_i(y_m) = \exp\left(-\frac{1}{2\,\kappa^2}(g_i^2 - 2\, y_m\, g_i)\right), \quad i = 1, \ldots, S.$$

Theorem 2.1 *The filtered state is given by* $\sigma_n(s_n) = \sum_{i=1}^{S} \sigma_n(s_n\, X_n(i))$ *where*

$$\sigma_n(s_n\, X_n(i)) = b_i(y_n) \sum_{j=1}^{S} C(e_j)\, a_{ij}\, \sigma_{n-1}(s_{n-1}\, X_{n-1}(j)) \quad (3)$$

Proof See [2] □

To compute $\hat{s}_n$ we use Thm. 2.1 and the normalization:

$$\hat{s}_n = \sigma_n(s_n)/\sigma_n(1) \text{ where } \sigma_n(1) = \sum_{j=1}^{S} \sigma_n(X_n(j))$$

where the un-normalized state estimate $\sigma_n(X_n(j))$ is computed using the standard HMM state filter [3]

$$\sigma_n(X_n(j)) = b_j(y_n) \sum_{i=1}^{S} a_{ji}\, \sigma_{n-1}(X_{n-1}(i)) \quad (4)$$

3. Continuous-time Filters

3.1. Signal Model and Aim

Let X_t, $t \geq 0$ be a continuous-time Markov chain defined on a probability space $(\Omega, \mathcal{F}, P)$ with state space $\{e_1, \ldots, e_S\}$. Let the transition rate matrix (infinitesimal generator) be A. That is, defining $p_t^i = P(X_t = e_i)$, $1 \leq i \leq S$, the probability distribution $p_t = (p_t^1\ p_t^2 \cdots p_t^S)'$ satisfies the forward equation $dp_t/dt = A\, p_t$. Also note that $\sum_{i=1}^{S} a_{ij} = 0$ for $1 \leq j \leq S$.

Consider the following Markov jump linear system

$$s_t = \int_0^t C(X_r)\, s_r\, dr + v_t \quad (5)$$

where $s_t, v_t \in \mathbf{R}^M$, s_0 is known and v_t is a zero mean process independent of $\mathcal{F}_t$. Also for each given X_r, $C(X_r)$ is a $M \times M$ known matrix.

Assume that X_t is observed indirectly via the process y_t where

$$y_t = \int_0^t \langle g, X_r\rangle\, dr + w_t$$

where w_t is a standard Wiener process independent of the processes X_t and v_t.

Let $\mathcal{F}_t$ and $\mathcal{Y}_t$ denote respectively, the sigma-algebras generated by X_s, $s \leq t$ and y_s, $s \leq t$. Also let $\mathcal{G}_t = \mathcal{Y}_t \bigvee \mathcal{F}_t$.

Aim: Compute the filtered estimate $\hat{s}_t = \mathbf{E}\{s_t|\mathcal{Y}_t\}$ a.s. where $\mathbf{E}$ denotes expectation under measure P.

3.2. Zakai State Filter for Jump Linear System

Define the probability measure P_0 such that the $\mathcal{F}_t$ restriction of the Radon-Nikodym derivative of P with respect to P_0 is

$$\frac{dP}{dP_0}\Big|_{\mathcal{F}_t} = \Lambda_t = \exp\left(\int_0^t \langle X_r, g\rangle\, dy_r - \tfrac{1}{2}\int_0^t \langle X_r, g\rangle^2\, dr\right)$$

Note that under P_0, y_t is a standard Wiener process independent of the process X_t [4].

Now for any measurable process H_t we write $\sigma_t(H_t) = \mathbf{E}_0\{\Lambda_t\, H_t|\mathcal{Y}_t\}$ where $\mathbf{E}_0$ denotes expectation with respect to P_0. An abstract version of Bayes' theorem then states that

$$\hat{H}_t = \mathbf{E}\{H_t|\mathcal{Y}_t\} = \sigma_t(H_t)/\sigma_t(1)$$

Theorem 3.1 *The Zakai filter for* s_t *defined in (1) is* $\sigma_t(s_t) = \sum_{i=1}^{S} \sigma_t(s_t\, X_t(i))$ *where*

$$\sigma_t(s_t\, X_t(i)) = \widehat{s_0\, X_0(i)} + C(e_i)\int_0^t \sigma_r(s_r\, X_r(i))\, dr$$
$$+ \sum_{j=1}^{S}\int_0^t \sigma_r(s_r\, X_r(j))\, a_{ij}\, dr + \int_0^t g_i \sigma_r(s_r\, X_r(i))\, dy_r \quad (6)$$

Proof See [2] □

To obtain $\hat{s}_t$ from (6) we use

$$\hat{s}_t = \sigma_t(s_t)/\sigma_t(1) \text{ where } \sigma_t(1) = \sum_{j=1}^{S} \sigma_t(X_t(j))$$

where the un-normalized state estimate $\sigma_t(X_t(j))$ is computed using the standard HMM state filter [4],

$$\sigma_t(X_t(j)) = \widehat{X_0(i)} + \sum_{j=1}^{S}\int_0^t \sigma_r(X_r(j))\, a_{ij}\, dr$$
$$+ \int_0^t g_i \sigma_r(X_r(i))\, dy_r \quad (7)$$

4. Rapprochement of Continuous and Discrete-time Filters

Given a continuous-time Markov jump linear system and a realization of the observation process, we are interested in obtaining a computable approximation of the continuous time filters. We consider two approaches:

1. One way to proceed is to discretize robust versions of the continuous-time filters derived in Sec. 3. This is discussed in Sec. 4.1 and 4.2.
2. Alternatively, the continuous-time jump linear system can be approximated by a discrete-time system and the discrete-time filters of Sec. 2 applied. This is discussed in Sec. 4.3.

The aim of this section is to establish the equivalence of these two approaches. In particular, we will show that a standard first-order discretization of the robust filter is identical to the discrete-time filter of Sec. 2 applied to a discrete-time approximation of the continuous-time Markov jump linear system.

4.1. Robust Continuous-time Filters

In this subsection we derive a version of the continuous-time filter which depends continuously on the observation path. This so called *robust filter* [5] involves the solution of an ordinary differential equation as opposed to the stochastic differential equation of (6). This robust reformulation of the filtering equations is also applied in [6].

Let $\phi_t^i = \exp\left\{g_i y_t - \frac{1}{2} g_i^2 t\right\}$. Then we can re-express the Zakai filter (6) in robust form as follows:

Theorem 4.1 *Suppose $\bar{\sigma}_t(X_t(i))$ and $\bar{\sigma}_t(s_t X_t(i))$ are the solutions of the ordinary linear differential equations*

$$\frac{d}{dt}\bar{\sigma}_t(X_t(i)) = \frac{1}{\phi_t^i}\sum_{j=1}^{S} a_{ij}\,\phi_t^j\,\bar{\sigma}_t(X_t(j)) \tag{8}$$

$$\frac{d}{dt}\bar{\sigma}_t(s_t X_t(i)) = C(e_i)\,\bar{\sigma}_t(s_t X_t(i)) + \frac{1}{\phi_t^i}\sum_{j=1}^{S} a_{ij}\,\phi_t^j\,\bar{\sigma}_t(s_t X_t(j)) \tag{9}$$

respectively.

Then for all $0 \le t \le T$

$$\pi_t(s_t X_t(i)) \triangleq \frac{\phi_t^i\,\bar{\sigma}_t(s_t X_t(i))}{\sum_{i=1}^{S}\phi_t^i\,\bar{\sigma}_t(X_t(i))}$$

defines a locally Lipschitz continuous version of $\mathbf{E}[s_t X_t(i)|\mathcal{Y}_t]$. That is,

$$|\pi_t(s_t X_t(i))[y_1] - \pi_t(s_t X_t(i))[y_2]| \le K\|y_1 - y_2\|$$

where

$$\|y\| \triangleq \sup_{0\le t\le T}|y(t)|,$$

$|\cdot|$ is the Euclidean norm of a vector, and K depends on $\|y_1\|$ and $\|y_2\|$.

Proof See [2] □

4.2. Explicit Time Discretization of Robust Filters

In what follows we consider a regular partition of the interval $[0,T]$ into N intervals of length Δ with $t_n = n\,\Delta$, $n = 0,\ldots,N$.

If we use an explicit first-order Euler approximation for (9) and transform back to the standard unnormalized conditional expectations, we arrive at the following approximation for (6)

$$\begin{aligned}\sigma_n(s_n X_n(i)) =& \psi_n^i\,\sigma_{n-1}(s_{n-1}X_{n-1}(i)) \\ &+ \Delta\psi_n^i\,C(e_i)\,\sigma_{n-1}(s_{n-1}X_{n-1}(i)) \\ &+ \Delta\psi_n^i\sum_{j=1}^{S} a_{ij}\,\sigma_{n-1}(s_{n-1}X_{n-1}(j))\end{aligned} \tag{10}$$

where $\psi_n^i = \phi_n^i/\phi_{n-1}^i$.

A similar procedure leads to a robust, explicit discretization for (7) as given in [5, 6]

$$\begin{aligned}\sigma_n(X_n(i)) =& \psi_n^i\,\sigma_{n-1}(X_{n-1}(i)) \\ &+ \Delta\psi_n^i\sum_{j=1}^{S} a_{ij}\,\sigma_{n-1}(X_{n-1}(j))\end{aligned} \tag{11}$$

4.3. Discrete-Time Approximate Model

We now wish to consider a discrete-time Markov jump linear system that approximates the continuous-time one. We use superscripts c and d to distinguish between discrete and continuous-time parameters and signals.

Consider the discrete-time Markov jump linear system with

$$\begin{aligned}A^d &= I + \Delta\,A^c \\ C^d(\cdot) &= I + \Delta\,C^c(\cdot) \\ y_n^d &= (y_n^c - y_{n-1}^c)/\Delta \\ \kappa^2 &= 1/\Delta\end{aligned}$$

The discrete-time filter equations of Sec. 2 become

$$\begin{aligned}\sigma_n(s_n X_n(i)) =& \sum_{j=1}^{S}(I + \Delta\,C^c(e_j))(\delta_{ij} + \Delta\,a_{ij}^c) \\ & b_j\left(\frac{y_n^c - y_{n-1}^c}{\Delta}\right)\sigma_{n-1}(s_{n-1}X_{n-1}(j))\end{aligned}$$

$$\sigma_n(X_n(i)) = (\delta_{ij} + \Delta\,a_{ij}^c)\,b_j\left(\frac{y_n^c - y_{n-1}^c}{\Delta}\right)\sigma_{n-1}(X_{n-1}(j))$$

Note that $b_j\left(\frac{y_n^c - y_{n-1}^c}{\Delta}\right) = \psi_n^j$. Finally expanding the above equations and neglecting the $O(\Delta^2)$ terms, we obtain identical filters to those obtained via explicit discretization in Sec. 4.2. *The important conclusion then is that a first order discretization of the robust continuous-time filter is equivalent to the discrete-time filter derived in Sec. 2.*

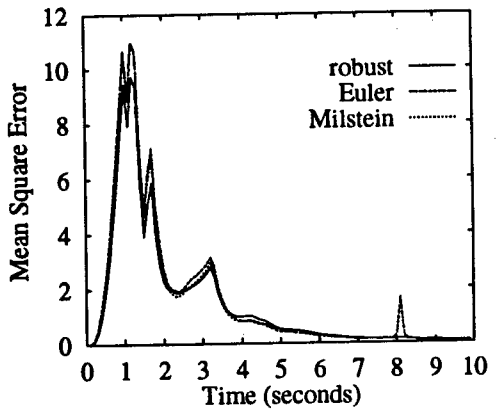


Figure 1. Mean Square Error: $\Delta = 0.1$

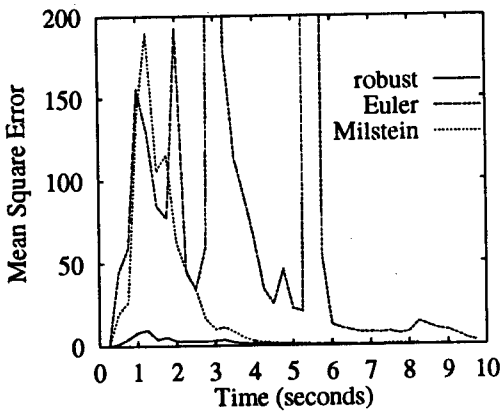


Figure 2. Mean Square Error: $\Delta = 0.25$

5. Numerical Examples

Rather than using this robust discretization scheme, the filtering equations (7) and (6) may be directly discretized using standard techniques for the numerical solution of stochastic differential equations.

In this section we compare the robust discretization with two direct techniques: the Euler-Maruyama and Milstein schemes. Roughly speaking, the Euler scheme is a first order approximation (more precisely it is an order 0.5 strong Ito-Taylor approximation) while the Milstein scheme is a second order discretization scheme (an order 1 strong Ito-Taylor approximation) [7, Chapter 10].

We consider a two-dimensional continuous-time jump linear system driven by a two state continuous-time Markov chain. The system parameters A, g, $C(e_1)$ and $C(e_2)$ used in the simulations are given by

$$A=\begin{pmatrix}-2 & 2\\ 2 & -2\end{pmatrix}\quad g=\begin{pmatrix}1\\ -1\end{pmatrix}\quad C(e_1)=\begin{pmatrix}0 & -1\\ 2 & -3\end{pmatrix}\quad C(e_2)=\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}$$

For all results, the simulation period was 10 seconds and the fast-sampled versions of the continuous-time sample paths were generated using a time step of 10^{-4} seconds. We assume perfect knowledge of the initial state of the Markov chain and jump linear system. In what follows, we assume each component of v_t is an independent Wiener process with zero drift and diffusion coefficient 0.01.

Fig. 1 and 2 illustrate the performance of the robust, Euler and Milstein discretizations of the continuous-time filter.

For each of the discretization step sizes, $\Delta = 0.1$ (Fig. 1), $\Delta = 0.25$ (Fig. 2), we plot of the evolution of the mean square error. The mean square error values were calculated based on 100 sample path runs. For ease of comparison each run was performed using the same realization of the Markov state.

With a small discretization step size the performance of all schemes is comparable. However as the discretization step size is increased we notice that the behaviour of the Euler (first order) and Milstein (second order) schemes becomes quite erratic. In contrast, the robust discretization (first order) continues to track satisfactorily.

6. Conclusion

In this paper we have derived *finite dimensional optimal recursive filters* for estimating the state of a Markov jump linear system given noisy observations of the underlying Markov chain.

We then presented a *robust discretization* of the continuous time filters which is based on the discretization of a *robust* version of the stochastic filtering equations. The robust discretization leads to a difference equation which is equivalent to that obtained using the discrete-time filters on a discrete approximation of the continuous-time model.

Simulations illustrated the advantages of the robust discretization over techniques based on the direct numerical solution of the stochastic filtering equations.

References

[1] J. Tugnait, "Adaptive estimation and identification for discrete systems with Markov jump parameters," *IEEE Trans. Auto. Control*, vol. AC-27, Oct. 1982.

[2] V. Krishnamurthy and J. Evans, "Finite dimensional filters for passive tracking of Markov jump linear systems," 1996. Submitted.

[3] R. J. Elliott, "Exact adaptive filters for Markov chains observed in Gaussian noise," *Automatica*, vol. 30, no. 9, pp. 1399–1408, 1994.

[4] R. J. Elliott, "New finite dimensional filters and smoothers for noisily observed Markov chains," *IEEE Trans. Inform. Theory*, vol. 39, pp. 265–271, Jan. 1993.

[5] J. Clarke, "The design of robust approximations to the stochastic differential equations of nonlinear filtering," in *Communication Systems and Random Processes Theory, Darlington 1977* (J. Skwirzynski, ed.), Alphen aan den Rijn: Sijthoff and Noordhoff, 1978.

[6] M. James, V. Krishnamurthy, and F. L. Gland, "Time discretization of continuous-time filters and smoothers for HMM parameter estimation," *IEEE Trans. Inform. Theory*, 1995. To appear.

[7] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*. Berlin Heidelberg: Springer-Verlag, 1992.

NUMBER THEORETIC METHODS IN PARAMETER ESTIMATION

Stephen D. Casey

Institute for Systems Research
University of Maryland
College Park, MD 20742

Brian M. Sadler

Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783

ABSTRACT

Recent developments in the theory on the zeta function, algorithms on generalizations of Euclidean domains, and variations on equidistribution theory have led to algorithms for several classes of problems in parameter estimation that are general and very efficient. We present the theoretical justifications for these algorithms, and discuss their use in the analysis of periodic pulse trains.

1. INTRODUCTION

[1]Problems in harmonic analysis and synthesis are intertwined with their applications in signal and image processing. Some recent advances in this analysis have used number theory to extend existing theories (e.g., sampling theory, fast transform computations) and develop new approaches to problems (e.g., interpolation). Number theoretic methods have also been successfully applied to the analysis of periodic point processes. The purpose of this note is to discuss several recent developments in which number theory has been used to develop algorithms for several classes of parameter estimation problems.

We first present modifications of the Euclidean algorithm which determine the period from a sparse set of noisy measurements [1, 2]. The elements of the set are the noisy occurrence times of a periodic event with (perhaps very many) missing measurements. The proposed algorithms are computationally straightforward and converge quickly. A robust version is developed that is stable despite the presence of arbitrary outliers. The Euclidean algorithm approach is justified by a theorem which shows that, for a set of randomly chosen positive integers, the probability that they do not all share a common prime factor approaches one quickly as the cardinality of the set increases. The theorem is in essence a probabilistic interpretation of the Riemann Zeta Function. In the noise-free case this implies convergence with only ten data samples, independent of the percentage of missing measurements. In the case of noisy data simulation results show, for example, good estimation of the period from one hundred data samples with fifty percent of the measurements missing and twenty five percent of the data samples being arbitrary outliers.

We then use these algorithms in the analysis of periodic pulse trains, getting an estimate of the underlying period [6, 7]. This estimate, while not maximum likelihood, is used as initialization in a three-step algorithm that achieves the Cramer-Rao bound for moderate noise levels, as shown by comparing Monte Carlo results with the Cramer-Rao bounds. An approach using multiple independent data records is also developed that overcomes high levels of contamination.

We close by discussing our work on the deinterleaving of multiple periodic pulse trains. Here we give a variation on Weyl's Equidistribution Theorem, which shows that noisy phase-wrapped data is equidistributed on $[0, 1)$ almost surely. We then use periodogram-like operators in a multistep procedure to isolate fundamental periods.

2. MODIFIED EUCLIDEAN ALGORITHMS

Our problem begins with a set of noisy occurrence times of a periodic event with (perhaps very many) missing measurements. We have developed modifications of the Euclidean algorithm for determining the period from this set [1], [2]. This problem arises in radar pulse repetition interval (PRI) analysis, in bit synchronization in communications, in biomedical applications, and other scenarios. We assume our data is a finite set of real numbers

$$S = \{s_j\}_{j=1}^{n}, \text{ with } s_j = k_j\tau + \phi + \eta_j, \qquad (1)$$

where τ (the period) is a fixed positive real number, the k_j's are non-repeating positive integers, ϕ (the phase)

[1]First author on sabbatical leave from American University. Research partially supported by AFOSR Grant F49620-94-1-0196.

is a real random variable uniformly distributed over the interval $[0, \tau)$, and the η_j's are zero-mean independent identically distributed (iid) error terms. We assume that the η_j's have a symmetric probability density function (pdf), and that $|\eta_j| < \frac{\tau}{2}$ for all j. We develop an algorithm for isolating the period of the process from this set, which we shall assume is (perhaps very) sparse. In the noise-free case our basic algorithm, given below, is equivalent to the Euclidean algorithm and converges with very high probability given only $n = 10$ data samples, independent of the number of missing measurements. We assume that the original data set is in descending order, i.e., $s_j \geq s_{j+1}$.

Modified Euclidean Algorithm

1.) After the first iteration, append zero.
2.) Form the new set with elements $s_j - s_{j+1}$.
3.) Sort in descending order.
4.) Eliminate elements in $[0, \eta_0]$ from end of the set.
5.) Algorithm is done if left with a single element. Declare $\widehat{\tau} = s_1$. If not done, go to **(1)**.

Here, $0 < \eta_0 < \tau$ is a noise threshold. Noise-free simulation examples demonstrate successful estimation of τ for $n = 10$ with 99.99% of the possible measurements missing. In fact, with only 10 data samples, it is possible to have the percentage of missing measurements arbitrarily close to 100%. There is, of course, a cost, in that the number of iterations the algorithm needs to converge increases with the percentage of missing measurements. In the presence of noise and false data (outliers), there is a tradeoff between the number of data samples, the amount of noise, and the percentage of outliers. The algorithm will perform well given low noise for $n = 10$, but will degrade as noise is increased. However, given more data, it is possible to reduce noise effects and speed up convergence by binning the data, and averaging across bins. Binning can be effectively implemented by using an adaptive threshhold with a gradient operator, allowing convergence in a single iteration in many cases. Simulation results show, for example, good estimation of the period from one hundred data samples with fifty percent of the measurements missing and twenty five percent of the data samples being arbitrary outliers [1], [2].

Our algorithm is based on several theoretical results, which we now present. The first leads to a modification of the basic Euclidean algorithm, allowing a reformulation using subtraction rather than division.

Proposition 1 ([1])

$$\gcd(\tau k_1, \ldots, \tau k_n) = \tau \gcd((k_1 - k_2), \ldots, (k_{n-1} - k_n), k_n) . \quad (2)$$

We then show that our procedure almost surely converges to the period by proving the following result. The Riemann Zeta Function is defined on the complex half space $\{z \in \mathbf{C} : \Re(z) > 1\}$ by $\zeta(z) = \sum_{n=1}^{\infty} n^{-z}$. Euler demonstrated the connection of ζ with number theory by showing (in 1736) that

$$\zeta(z) = \prod_{j=1}^{\infty} \frac{1}{1 - (p_j)^{-z}} , \Re(z) > 1 ,$$

where $\mathbf{P} = \{p_1, p_2, p_3, \ldots\} = \{2, 3, 5, \ldots\}$ is the set of all prime numbers. In the following, we let $P\{\cdot\}$ denote probability, $\text{card}\{\cdot\}$ denote the cardinality of the set $\{\cdot\}$, and let $\{1, \ldots, \ell\}^n$ denote the sublattice of positive integers in $\mathbf{R}^n$ with coordinates c such that $1 \leq c \leq \ell$. Therefore, $N_n(\ell) = \text{card}\{(k_1, \ldots, k_n) \in \{1, \ldots, \ell\}^n : \gcd(k_1, \ldots, k_n) = 1\}$ is the number of relatively prime elements in $\{1, \ldots, \ell\}^n$.

Theorem 1 ([1]) *For $n \geq 2$, we have that*

$$\lim_{\ell \to \infty} \frac{N_n(\ell)}{\ell^n} = [\zeta(n)]^{-1} . \quad (3)$$

Therefore, given n ($n \geq 2$) randomly chosen positive integers $\{k_1, \ldots, k_n\}$,

$$P\{\gcd(k_1, \ldots, k_n) = 1\} = [\zeta(n)]^{-1} . \quad (4)$$

Also,

$$\lim_{n \to \infty} [\zeta(n)]^{-1} = 1 , \quad (5)$$

converging to 1 from below faster than $1/(1 - 2^{1-n})$.

Thus, from (4) and (5), as n grows it quickly becomes very likely that n randomly selected integers have a gcd of 1. This fact, together with Proposition 1, make estimation of τ via our algorithm possible.

3. PRI ANALYSIS

The parameter estimation techniques given above lead to an effective method for periodic pulse interval analysis (see [6], [7]). We assume time is highly resolved and ignore any time quantization error. We are primarily concerned with a single periodic pulse train with (perhaps very many) missing observations that may be contaminated with outliers. Our data model for this case, in terms of the arrival times t_j, is given by (1), with the additional assumption that η_j is zero-mean additive white Gaussian noise. Outliers are included as arbitrary arrival times. The problem, again, is to recover the period τ and possibly the phase ϕ. With Gaussian noise the minimum variance unbiased estimates for this linear regression problem take a least-squares form. However, this requires knowledge

of the k_j's. We therefore propose a multi-step procedure that proceeds by (i) estimating τ directly, (ii) estimating the k_j's, and (iii) refining the estimate of τ using the estimated k_j's in the least-squares solution. This estimate is shown to perform well, achieving the Cramer-Rao bound in many cases, despite many missing observations and contaminated data. The direct estimate of τ (step (i)) is obtained using the modified Euclidean algorithms described above. While not maximum-likelihood (ML), the modified Euclidean algorithm performs well under difficult conditions.

We now give the maximum likelihood solution and Cramer-Rao bounds for estimating τ and ϕ. Our analysis has led us to work with the data set $\{t_{j+1}-t_j\}_{j=1}^{n-1}$, so as to avoid estimating ϕ (which can be unreliable). Given the sample data set S from (1) we may write

$$\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} 1 & k_1 \\ 1 & k_2 \\ \vdots & \vdots \\ 1 & k_n \end{bmatrix} \begin{bmatrix} \phi \\ \tau \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}, \quad (6)$$

where $k_{j+1} > k_j$. In compact form this is

$$\mathbf{t} = X\beta + \eta, \quad (7)$$

where $\beta = [\phi, \tau]^T$ and the definitions of $\mathbf{t}$, η, and X follow from (6). We eliminate ϕ by forming the differences $y_j = t_{j+1} - t_j = (k_{j+1} - k_j)\tau + (\eta_{j+1} - \eta_j)$, yielding

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \end{bmatrix} = \begin{bmatrix} k_2 - k_1 \\ k_3 - k_2 \\ \vdots \\ k_n - k_{n-1} \end{bmatrix} \tau + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_{n-1} \end{bmatrix}, \quad (8)$$

where $\delta_j = \eta_{j+1} - \eta_j$. Similar to (7) we may write (8) compactly as

$$\mathbf{y} = X_d\tau + \delta. \quad (9)$$

Equations (7) and (9) are linear regression problems whose least squares solutions yield the minimum-variance unbiased estimate when the noise is zero-mean Gaussian, e.g., see Kay [4]. Generally, use of (9) is preferred for estimating τ, avoiding estimation of ϕ which has high variance. The solution to (9) corresponds to ML estimation and takes the form of a least squares estimate

$$\widehat{\tau} = (X_d^T R_\delta^{-1} X_d)^{-1} X_d^T R_\delta^{-1} \mathbf{y}, \quad (10)$$

where $R_\delta = E[\delta\delta^T]$. We have assumed white noise so $R_\delta = \sigma_\eta^2 \tilde{R}_\delta$ where $\tilde{R}_\delta$ has 2's on the main diagonal, -1's on the first upper and lower diagonals, and zeros elsewhere. In general R_δ is full rank and its inverse can be expressed element-wise as $[R_\delta^{-1}]_{ij} = \min(i,j) - ij/n$, and is therefore easily computed. Although optimal, use of (10) requires knowledge of X_d. This is not a problem if there are no missing observations for then $k_j = j$ for $j = 1, 2, \ldots n$. However, when observations are arbitrarily missing then the k_j's are not known in general, and one is faced with more unknowns than equations in (9).

The pdf of the noise η in (7) is multivariate Gaussian, leading to the Cramer-Rao bound (CRB) for (10)

$$\mathrm{var}\{\tau - \widehat{\tau}\} \geq \sigma_\delta^2 (X_d^T \tilde{R}_\delta^{-1} X_d)^{-1}, \quad (11)$$

with $\sigma_\delta = 2\sigma_\eta$. Generally, the CRB is reduced for smaller σ_η^2. Also, for fixed n, it is reduced when the spread of the k_j's increases.

Now, if τ were known then X_d could be estimated using $(1/\tau)$ y. Ideally, this estimate is composed of positive integers, but imperfect knowledge of τ and the presence of noise will generally yield an estimate of X_d that has non-integer components. We therefore propose to estimate X_d via

$$\widehat{X}_d = \mathrm{round}\left[\frac{1}{\widehat{\tau}_{MEA}}\,\mathbf{y}\right], \quad (12)$$

where $\widehat{\tau}_{MEA}$ is the estimate of τ obtained via the modified Euclidean algorithm, and $\mathrm{round}[\cdot] = \lfloor \cdot + \frac{1}{2} \rfloor$ is rounding to the nearest integer. A refined estimate of τ is then obtained by using $\widehat{X}_d$ in (10) yielding

$$\widehat{\tau} = (\widehat{X}_d^T R_\delta^{-1} \widehat{X}_d)^{-1} \widehat{X}_d^T R_\delta^{-1} \mathbf{y}. \quad (13)$$

This result approaches the optimal minimum variance performance when $\widehat{X}_d$ is close to X_d. The refinement algorithm is summarized as follows.

Refined Estimation Algorithm

1.) Estimate τ via the modified Euclidean algorithm, calling this estimate $\widehat{\tau}_{MEA}$.
2.) Estimate $\widehat{X}_d$ via (12).
3.) Refine the estimate of τ using $\widehat{X}_d$ in (13), calling this estimate $\widehat{\tau}$.

Performance analysis of the estimate $\widehat{\tau}_{MEA}$ depends not only on the distribution of the noise η_j, but also on the distribution of the k_j's. We have completed this analysis for some specific cases in [6]. We also compare the estimates to Cramer-Rao bounds via Monte Carlo simulation, revealing the very good performance of the algorithm with many missing observations and contaminated data (see [6], [7]). We can also apply our estimation procedures to estimation of the frequency of a single sinusoid in Gaussian noise. We address the problem [8], using only very sparse noisy zero-crossings with the presence of outliers.

4. DEINTERLEAVING

We close by discussing our work on deinterleaving. Our data model is the union of M copies of (1), each with different periods or "generators" $\Gamma = \{\tau_i\}$, k_{ij}'s and phases. Let $\tau = \max_i\{\tau_i\}$. Then our data is

$$\Delta = \bigcup_{i=1}^{M} \{\phi_i + k_{ij}\tau_i + \eta_{ij}\}_{j=1}^{n_i} , \qquad (14)$$

where n_i is the number of elements from the i^{th} generator, $\{k_{ij}\}$ is a linearly increasing sequence of natural numbers with missing observations, ϕ_i is a random variable uniformly distributed in $[0, \tau_i)$, and the η_{ij}'s are zero-mean iid Gaussian with standard deviation $3\sigma_{ij} < \tau/2$. We think of the data as events from M periodic processes, and represent it, after reindexing, as $\Delta = \{\alpha_l\}_{l=1}^{N}$. Assuming only minimal knowledge of the range of $\{\tau_i\}$, namely bounds T_L, T_U such that $0 < T_L \leq \tau_i \leq T_U$, we phase wrap the data by the mapping

$$\Phi_\rho(\alpha_l) = \left\langle \frac{\alpha_l}{\rho} \right\rangle = \frac{\alpha_l}{\rho} - \left\lfloor \frac{\alpha_l}{\rho} \right\rfloor , \qquad (15)$$

where $\rho \in [T_L, T_U]$, and $\lfloor\cdot\rfloor$ is the floor function. Thus $\langle\cdot\rangle$ is the fractional part, and so $\Phi_\rho(\alpha_l) \in [0, 1)$.

Definition 1 *A sequence of real random variables* $\{x_j\} \subset [0, 1)$ *is* essentially uniformly distributed in the sense of Weyl *if given* a, b, $0 \leq a < b < 1$,

$$\frac{1}{n}\text{card}\{1 \leq j \leq n : x_j \in [a, b]\} \longrightarrow (b - a) \qquad (16)$$

as $n \longrightarrow \infty$ *almost surely.*

Weyl's Theorem is presented in [3]. For our variation, we assume that for each i, $\{k_{ij}\}$ is a linearly increasing infinite sequence of natural numbers with missing observations such that $k_{ij} \longrightarrow \infty$ as $j \longrightarrow \infty$. We must make this assumption because the result is only approximately true for a finite length sequence.

Theorem 2 *For almost every choice of* ρ *(in the sense of Lebesgue measure)* $\Phi_\rho(\alpha_l)$ *is essentially uniformly distributed in the sense of Weyl.*

Moreover, the set of ρ's for which this is not true are rational multiples of $\{\tau_i\}$. Therefore, except for those values, $\Phi_\rho(\alpha_{ij})$ is essentially uniformly distributed in $[0, 1)$. Moreover, the values at which $\Phi_\rho(\alpha_{ij}) = 0$ almost surely are $\rho \in \{\tau_i/n : n \in \mathbf{N}\}$. These values of ρ cluster at zero, but spread out for lower values of n.

We then map the phase wrapped data by non-linear variations on the periodogram,

$$F(\alpha_l, \rho) = \frac{1}{N}\sum_l \cos^{2r-1}(2\pi\frac{\alpha_l}{\rho}) + i\frac{1}{N}\sum_l \sin^{2r-1}(2\pi\frac{\alpha_l}{\rho}) , \qquad (17)$$

for $r = 2, 3, \ldots$. Now, the periodicity of sin and cos gives us that $\cos^{2r-1}(2\pi\Phi_\rho(\alpha_l)) = \cos^{2r-1}(2\pi\frac{\alpha_l}{\rho})$ and $\sin^{2r-1}(2\pi\Phi_\rho(\alpha_l)) = \sin^{2r-1}(2\pi\frac{\alpha_l}{\rho})$. By Theorem 2, the random variables $\Phi_\rho(\alpha_l)$ are uniformly distributed on $[0, 1)$ for almost every choice of ρ. We can then compute the distributions of the real and imaginary parts of F. The "noise-like" behavior of $\Phi_\rho(\alpha_l)$ for *a.e.* ρ leads to a "flat" range for F. However, at $\rho \in \{\tau_i/n : n \in \mathbf{N}\}$, we have increasingly strong peaks as n decreases. In turn, this gives the following. Let i_0 denote the index of the most prolific generator, and $\Re$, $\Im$ denote the real and imaginary parts.

Theorem 3

$$\max_\rho (\Re F - |\Im F|) = \tau_{i_0} . \qquad (18)$$

We then isolate the data generated by τ_{i_0} by convolution with a pulse train of width τ_{i_0}, and subtract it out. We then repeat the process, terminating when Δ equals the empty set.

5. REFERENCES

[1] Casey, S. D., and Sadler, B. M., "Modifications of the Euclidean algorithm for isolating periodicities from a sparse set of noisy measurements," *IEEE Trans. Sig. Proc.*, to appear.

[2] Casey, S. D., and Sadler, B. M., "A Modification of the Euclidean algorithm for isolating periodicities from a sparse set of noisy measurements," *IEEE Int. Conf. Acoustics, Speech, and Sig. Proc. (ICASSP '95)*, Vol. 3, pp. 1764-1767, 1995.

[3] Dym, H., and McKean, H. P., *Fourier Series and Integrals*, Academic Press, Orlando, Florida, 1972.

[4] Kay, S. M., *Fundamentals of Statistical Signal Processing*, Prentice-Hall, New Jersey, 1993.

[5] Knuth, D. E., *The Art of Computer Programming, Volume 2: Seminumerical Algorithms (Second Edition)*, Addison-Wesley, Reading, Massachusetts, 1981.

[6] Sadler, B. M., and Casey, S. D., "PRI analysis from sparse data via a modified Euclidean algorithm," *29th Annual Asilomar Conf. on Sig., Syst., and Computers* (invited), 1995.

[7] Sadler, B. M., and Casey, S. D., "On periodic pulse interval analysis with outliers and missing observations," submitted to *IEEE Trans. Sig. Proc.*, 1996.

[8] Sadler, B. M., and Casey, S. D., "Frequency estimation via sparse zero crossings," *IEEE Int. Conf. Acoustics, Speech, and Sig. Proc. (ICASSP '96)*, 1996.

[9] Wiley, R. G., *Electronic Intelligence: The Analysis of Radar Signals (Second Edition)*, Artech House, Norwood, Massachusetts, 1993.

TP-2: Nonstationary Signal Processing

A TEST FOR SPATIAL STATIONARITY AND APPLICATIONS

A. Ephraty, J. Tabrikian and H. Messer
Department of Electrical Engineering - Systems, Tel Aviv University
Tel-Aviv 69978, Israel
Tel. +972-36408119, Fax. +972-36407095
E-mail: messer@eng.tau.ac.il

Abstract

This paper presents a test which accepts or rejects, based on the data collected by an array of N sensors, the hypothesis that the sampled tempo-spatial field is spatially stationary. The proposed test is applicable to arrays in an arbitrary but known 3-dimensional geometry. It is based on the estimated second order *spatial cumulant spectrum matrix, which is theoretically diagonal for a stationary spatial field. We show how the proposed test can be used for robust detection of a source in shallow water.*

1. Introduction

This paper presents a test which accepts or rejects, based on the data collected by an array of N sensors, the hypothesis that the sampled tempo-spatial field is spatially stationary. Unlike the test proposed in [1] which is applicable only to uniform spatial sampling (i.e., only to linear, uniform arrays), our test is applicable to arrays in *arbitrary* but known 3-dimensional geometry. Similarly to the test for temporal stationarity of time series [3], our test uses the *cumulant spectra* of the tempo-spatial field. In particular, it is based on the estimated *second order* spatial cumulant spectrum.

It can be shown that the necessary and sufficient conditions for the spatial fields to be spatially stationary are:

1. The sources are uncorrelated and are located in the far field zone.
2. The sources are zero mean, temporally stationary.
3. The additive noise is spatially stationary.

The last two conditions are satisfied in most applications[1] and therefore the test can be used for studying the physical scenarios related to condition 1. For example, if the propagation medium is known to be bounded (as in shallow water), spatial non-stationarity indicates existence of one or more sources in the sampled tempo-spatial field.

The proposed test (detector) does not employ any prior knowledge about the source, the number of the sources, the noise and/or the bounded propagation environment. Therefore, it is robust to modeling uncertainties or mismatches, which are a major problem in underwater acoustics. The performance degradation due to this robustness is studied by comparing the performance of the proposed test to those of two generalized likelihood test ($GLRT$) detectors:

- $GLRT1$, in which only the spatial spectrum of the additive noise is assumed to be known, and
- $GLRT2$, in which the spatial spectrum of the additive noise *and* the propagation channel are assumed to be known.

We show that, as expected, if there is no mismatch in propagation channel, the $GLRT2$ outperforms $GLRT1$ which outperforms the proposed test. In the presence of modeling mismatches the performances of the two $GLRT$ detectors reduce dramatically. However, as the proposed test reflects the degree of stationarity, its performance does not significantly vary.

2. The proposed test

Assume an array of sensors with N elements located at $\{\mathbf{x}_i\}_{i=1}^N$. Let $y_i(t)$ denote the measured random field by sensor i at time t: $y_i(t) = y(t, \mathbf{x}_i), \quad i = 1, \cdots, N$. If the random field is zero mean, the entries of the $N \times N$ spatial covariance matrix are given by the samples of the cross-correlation function $\mathbf{R}_y(t, \mathbf{x}_1, \mathbf{x}_2)$:

$$[\mathbf{R}_y(t)]_{ij} = \mathbf{R}_y(t, \mathbf{x}_i, \mathbf{x}_j) = E\{y(t, \mathbf{x}_i) y^*(t, \mathbf{x}_j)\} \quad i, j = 1, ..., N \tag{1}$$

[1] It can be shown that while with real data collected in shallow water, condition 3 is not always satisfied, a certain operation on the data can impose the additive noise to be spatially stationary.

0-8186-7576-4/96 $5.00 © 1996 IEEE

If the field is temporally stationary, then $[\mathbf{R}_y(t)]_{ij} = [\mathbf{R}_y]_{ij} = \mathbf{R}_y(\mathbf{x}_i, \mathbf{x}_j)$.

The second order cumulant spectrum is defined by a two dimensional Fourier transform of the cross-correlation function:

$$\mathbf{S}_y(\mathbf{k}_1, \mathbf{k}_2) = \int_{\mathbf{x}_1} \int_{\mathbf{x}_2} \mathbf{R}_y(\mathbf{x}_1, \mathbf{x}_2) e^{-j(\mathbf{k}_1^t \mathbf{x}_1 + \mathbf{k}_2^t \mathbf{x}_2)} d\mathbf{x}_1 d\mathbf{x}_2 \ . \tag{2}$$

where $\mathbf{k}$ is the 3-dimensional vector of the wave numbers at the three directions. If the field is spatially stationary, then $\mathbf{R}_y(\mathbf{x}_1, \mathbf{x}_2) = func(\mathbf{x}_1 - \mathbf{x}_2)$. In [2], it is shown that this condition leads to:

$$\mathbf{S}_y(\mathbf{k}_1, \mathbf{k}_2) = func(\mathbf{k}_1)\delta(\mathbf{k}_1 + \mathbf{k}_2) \ . \tag{3}$$

We construct the stationarity test by studying the estimated second order spatial cumulant matrix $[\hat{S}]_{ij} = \hat{\mathbf{S}}_y(\mathbf{k}_{1i}, \mathbf{k}_{2j})$. Theoretically, if the field is spatially stationary, only the diagonal of the matrix $\hat{S}$ consists of non-zero elements. Our proposed test is based on this property.

The test is based on accepting or rejecting the null hypotheses:
H_0: spatially stationary field: $\mathbf{S}_y(\mathbf{k}_1, \mathbf{k}_2) = f(\mathbf{k}_1)\delta(\mathbf{k}_1 + \mathbf{k}_2)$.
If the received signal at the array is (temporally) ergodic, then:

$$\mathbf{R}_y(\mathbf{x}_1, \mathbf{x}_2) = \lim_{T \to \infty} \frac{1}{T} \int_T y(t, \mathbf{x}_1) y^*(t, \mathbf{x}_2) dt \tag{4}$$

and the spatial covariance matrix can be estimated by integration over sufficiently large observation time T. Hence, the samples obtained from an array of sensors are used to estimate the covariance matrix. The second order cumulant spectrum, $\hat{S}$, is calculated by applying the discrete Fourier transform.

Our test is an ad-hoc one, and it suggests to evaluate the intensity of the off-diagonal entries[2] of $\hat{S}$ and to compare it to a given threshold:

$$\zeta = \sum_{i \neq j} |\hat{S}_{ij}|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \ . \tag{5}$$

Selection of the threshold γ determines the probability of "false alarm," P_{fa}. For example, if the additive noise is zero mean, i.i.d. Gaussian noise such that its time samples at different sensors satisfy:

$$E\{n_m(l) n_p^*(l)\} = R_n(m,p) = \sigma_n^2 \delta(m-p) \tag{6}$$

then, the asymptotic distribution of the test statistic ζ under H_0 is complex Wishart [2] with degrees of freedom $\frac{N}{2}(N-1)$ and covariance $\frac{(N\sigma_n^2)^2}{L}$:

$$\zeta|_{H_0} \sim W^c(\frac{N}{2}(N-1), \frac{(N\sigma_n^2)^2}{L}) = \frac{(N\sigma_n^2)^2}{2L} \chi^2(N(N-1)) \tag{7}$$

L is the number of time samples, which is roughly the time-bandwidth product, and $\chi^2(M)$ is the Chi squares distribution with M degrees of freedom. That is, for any give false alarm probability, $P_{fa} = Prob\{\zeta|_{H_0} > \gamma\}$, the threshlod γ can be set.

3. The alternative tests

While the proposed test only uses the fact that under H_0 the field is stationary, both alternative tests employ specific prior information about the propagation field. For $GLRT1$ only the noise covariance matrix, $R_n(m,p) = E\{n_m(l) n_p^*(l)\}$, $m,p = 1,...,N$, is assumed known. Therefore, the test accepts H_0 (no signal) if the measured covariance matrix is R_n and rejects it otherwise. Since there is a one-to-one correspondence between the covariance matrix and the cross-spectral matrix (the second order spatial cumulant spectrum), the test can similarly be put in terms of the estimated spectral matrix $\hat{S}$. Since $\hat{S}$ is Hermitian, it is sufficient to look at either the upper or the lower off-diagonal entries of $\hat{S}$ and the diagonal terms. Putting these $N + \frac{N}{2}(N-1)$ entries in a vector $\hat{\mathbf{s}}$, this vector is asymptotically complex Gaussian. Under H_0 its mean, μ_0, and covariance, Λ_0, are assumed known while under H_1 its mean vector, μ_1, and its covariance matrix, Λ_1, are free. Thus, the $GLRT$ takes the form:

$$\tilde{\zeta} = \frac{f(\hat{\mathbf{s}}|H_0)}{\max_{\mu_1, \Lambda_1} f(\hat{\mathbf{s}}|H_1)} = \frac{f(\hat{\mathbf{s}}|H_0)}{f(\hat{\mathbf{s}}|\mu_1^{ML}, \Lambda_1^{ML})} \underset{H_1}{\overset{H_0}{\gtrless}} \tilde{\gamma} \tag{8}$$

where μ_1^{ML} and Λ_1^{ML} are the maximum likelihood estimates of μ_1 and Λ_1 (the sample mean and covariance of $\hat{S}$). For large observation time (where the Gaussian assumption holds) the test is equivalent to:

$$\zeta_1 = (\hat{\mathbf{s}} - \mu_0)^H \Lambda_0^{-1} (\hat{\mathbf{s}} - \mu_0) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_1 \tag{9}$$

For the case of white, i.i.d. noise the test gets a form similar to (7) and the threshold γ_1 can be evaluated for any given P_{fa}.

Note that $GLRT1$ can be also phrased in the domain of the covariance matrix, as in [1].

The other alternative test, $GLRT2$, employs even more prior information than $GLRT1$: it assumes also that the propagation channel is modeled with unknown parameters. The test decides on one of the two hypotheses:

$$\begin{aligned} H_0: \ & \mathbf{y}(l) = \mathbf{n}(l) \\ H_1: \ & \mathbf{y}(l) = \mathbf{g}(\theta) s(l) + \mathbf{n}(l) \end{aligned} \tag{10}$$

[2] Since $\hat{S}$ is an Hermitian matrix, the test uses either the upper or the lower off-diagonal matrix.

where $l = 1, ..., L$ are the time samples. $\mathbf{n}$ is the noise vector, s is the source signal of unknown power σ_s^2 and $\mathbf{g}(\cdot)$ is a known function which characterizes the propagation channel. θ is the vector of the unknown source location and channel parameters. $\mathbf{n}$ and s are assumed to be zero mean, uncorrelated Gaussian processes of known covariance Λ_n and 1, respectively. The distribution of the data under H_0 is completely known (as in $GLRT1$). Under H_1 it is known up to few parameters (σ_s and θ) which appear in the covariance of a Gaussian distribution, and therefore the $GLRT$ can be formed. The resulting test is:

$$\zeta_2 = \min_{\theta}\{ln\phi(\mathbf{y},\theta) - \phi(\mathbf{y},\theta)\} \underset{H_1}{\overset{H_0}{\gtrless}} \gamma_2 \qquad (11)$$

where for white noise of variance σ_n^2, $\phi(\mathbf{y},\theta) = \frac{\mathbf{g}^H(\theta)\hat{R}_y\mathbf{g}(\theta)}{\sigma_n^2\mathbf{g}^H(\theta)\mathbf{g}(\theta)}$ and $\hat{R}_y = \frac{1}{L}\sum_{l=1}^{L}\mathbf{y}(l)\mathbf{y}^H(l)$. For this test, however, it is hard to establish the threshold γ_2 analytically for a given P_{fa} even when the noise is white.

Equation (11) shows that unlike $GLRT1$ and the proposed test, $GLRT2$ involves a multidimensional search procedure over the unknown vector parameter θ. Furthermore, it requires knowledge of the number of sources under hypothesis H_1.

Since a test which employs more prior information performs better, we expect that the $GLRT2$ would outperform the $GLRT1$ which would outperform our ad-hoc stationarity test. In the next section we consider a practical problem where we demonstrate that this is indeed the case when the propagation channel corresponds to the modeling assumption. If, however, there are modeling mismatches, as is usually the case, e.g. in underwater acoustics, the difference in performance of the three tests becomes negligible.

4. Simulation results and conclusions

In this section, we focus on one of the applications of the test in a shallow water waveguide. Fig. 1 describes the environmental scenario which was one of the more complex benchmarks used in the May 1993 NRL Workshop on Acoustic Models in Signal Processing [5]. We apply our test to detect a source in the channel. Using a normal mode propagation program, KRAKEN [4], we simulated the channel in which a narrow band point source at frequency $100Hz$ in a temporally and spatially white, zero mean, Gaussian noise is assumed. The source was located at depth $z_0 = 50m$ and range of $r_0 = 7000m$ from the array. 200 snapshots of the received field were obtained by a uniform, vertical array of 13 sensors whose aperture is the depth of the propagation channel.

Figure 2 compares the receiver operation characteristic (ROC) of the tests presented in this paper, assuming no mismatch in the channel environmental parameters. Fig. 3 presents the probability of detection of a source in the channel as a function of the signal-to-noise-ratio (SNR) per sensor per snapshot, with false alarm 10^{-3}.

In the second experiment mismatches were induced in the environmental parameters according to Fig. 1. The performance of the test is depicted, in its two representations of Fig. 2 and Fig. 3, in Fig. 4 and in Fig. 5, respectively. The performance of the GLRTs reduces, while that of the proposed stationarity test even improves slightly.

These results show that the proposed test for source detection in shallow water is robust while the other tests, specially the $GLRT2$ are very sensitive to errors in the assumed environmental parameters. In addition it is computationally simple and does not involve a search procedure over the unknown parameters.

References

[1] S. Bose and A. O. Steinhardt. Invariant tests for spatial stationarity using covariance structure. *Proc. IEEE ICASSP*, 4:41–44, May 1993.

[2] D. R. Brillinger. *Time series: Data Analysis and Theory.* Holden-Day, San Fransisco, CA, 1981.

[3] M. J. Hinich. Detecting a transient signal by bispectral analysis. *IEEE Trans. Acoust., Speech, Signal Processing*, 38:1277–1283, July 1990.

[4] M. B. Porter and E. L. Reiss. A numerical method for ocean acoustic normal modes. *J. Acoust. Soc. Am.*, 76:244–252, 1984.

[5] M. B. Porter and A. Tolstoy. The matched field processing benchmark problems. *Journal of Computational Acoustics*, 2:161–185, 1994.

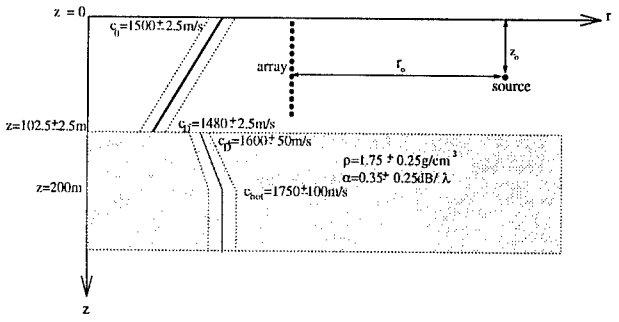


Figure 1. The NRL workshop "genlmis" scenario configuration

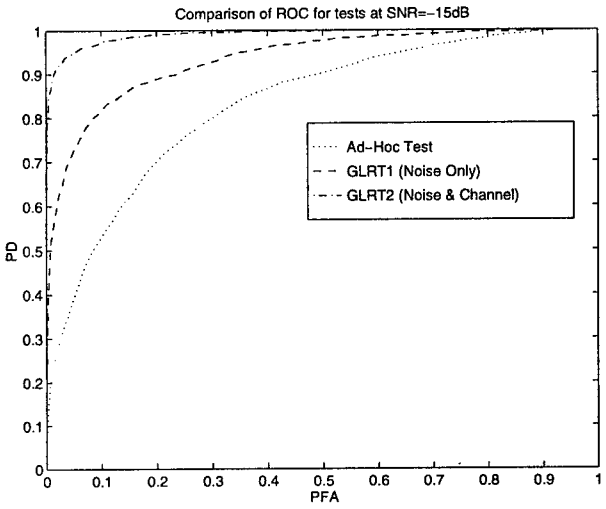


Figure 2. Receiver operation characteristic for source detection with no environmental mismatch.

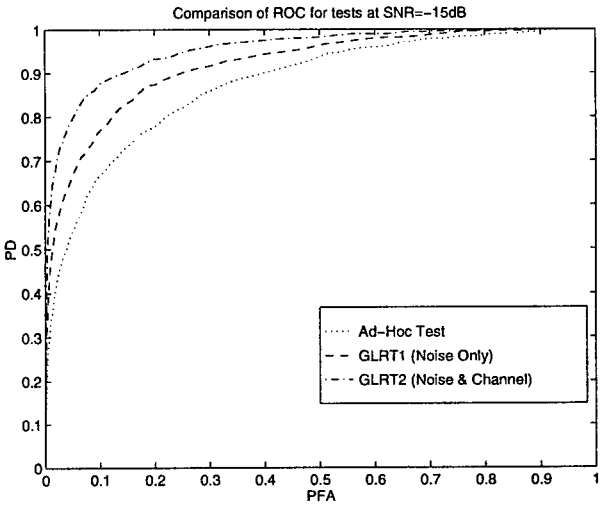


Figure 4. Receiver operation characteristic for source detection under environmental mismatch.

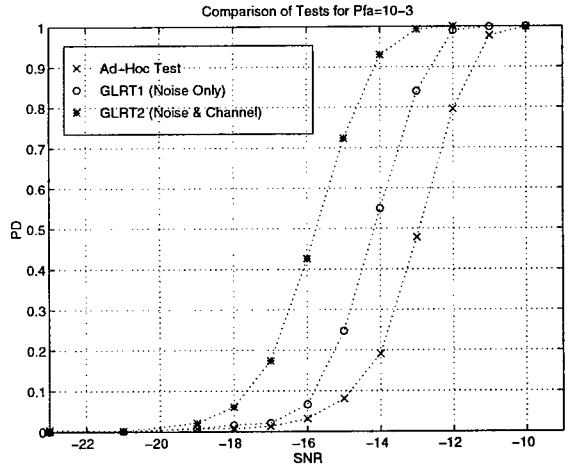


Figure 3. Probability of source detection vs. SNR with no environmental mismatch.

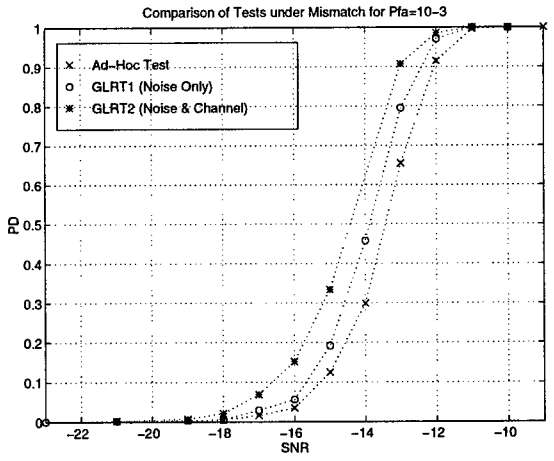


Figure 5. Probability of source detection vs. SNR under environmental mismatch

Measure of Cyclostationarity for Gaussian Processes Based on the Likelihood Ratio Test.

Christophe ANDRIEU - Patrick DUVAUT
ENSEA-ETIS Groupe Signal, 6 avenue du Ponceau 95014 Cergy Cedex - France
andrieu@ensea.fr - duvaut@ensea.fr

Abstract

The problem addressed in this paper is the detection of cyclostationarity, and the measurement of the trend of a process to have this property. This problem is of great importance, because in applications algorithms using the property of cyclostationarity assume the periodicities of the statistics to be known. Thus the periodicities need to be detected/estimated, and furthermore, a measure must be given in order to qualify the trend of a process to have a given periodicity. This measure will give information about the opportuneness of using a cyclostationary modelization instead of a stationary one.

1. Introduction

As pointed out by the number of publication on the subject, cyclostationary processes (processes whose statistics depend periodically on time) are of great interest in many fields (communications, signal processing, hydrology, multivariate analysis, array processing...) and have shown to provide a better modelization in many cases of interest than the stationary "syndrome".

The problem addressed in this paper is the hypothesis testing, of cyclostationarity versus stationarity for Gaussian processes. Indeed, althought interesting and useful tests for the presence of cyclostationarity have been introduced in [7][8][10][5], we got interested in developing another test relying on the theory of detection [2]. As it will be stated in this paper, the test we develope, will lead to a natural "measure" of cyclostationarity, in close connection with the Kullback-Leibler information.

We first recall some properties of second order cyclostationarity and notations useful in a cyclostationary framework. We then briefly recall the works of [8] [10] and [5] on the problem of detection. We then present our test, and results concerning the cyclostationary version of Whittle's approximant, and the asymptotic power and false alarm probability of our test.

1.1. Cyclostationarity

A real valued time series $(x_k)_{k\in\mathbb{Z}}$ is said to be cyclostationary when the following properties are respected for all n and an integer T [9]:

$$\begin{aligned} \mathbb{E}\left[x_n^2\right] &< +\infty \\ \mathbb{E}[x_{n+T}] &= \mathbb{E}[x_n] \\ r_x[n,\tau] \stackrel{def}{=} \mathbb{E}[x_n x_{n+\tau}] &= \mathbb{E}[x_{n+T}x_{n+\tau+T}] \end{aligned} \quad (1)$$

The correlation matrix Σ_n of size nT of the process $(x_k)_{k\in\mathbb{Z}}$ is block-Toeplitz in the general case, and Toeplitz when the process is stationary, that is $T = 1$.

Due to periodicity in time, the correlation can be written as follows [9]:

$$r_x[n,\tau] = \sum_{k=0}^{T-1} r_x^k[\tau] \exp\left(\frac{2i\pi kn}{T}\right) \quad (2)$$

The $\left(r_x^k[\tau]\right)_{k=0,\ldots,T-1,\tau\in\mathbb{Z}}$ are called the cyclocorrelations. One can associate to those cyclocorrelations the cyclospectra, defined as[9]:

$$r_x^k[\tau] = \int_0^{2\pi} \exp(i\lambda\tau)\, \mathcal{F}_{r_x}^k(\lambda)\, d\lambda \quad (3)$$

1.2. The Zivanovic-Gardner's Degree of Cyclostationarity

Zivanovic and Gardner have defined a degree of cyclostationarity by a distance between the correlation of the cyclostationary process to the correlation of the

0-8186-7576-4/96 $5.00

"closest" stationary process [8]. This definition leads to the usefull following expression:

$$DCS_x = \frac{\sum_{\alpha \neq 0} \int |\mathcal{F}_x^\alpha(\lambda)|^2 \, d\lambda}{\int |\mathcal{F}_x^0(\lambda)|^2 \, d\lambda} \quad (4)$$

which is none but the ratio of energy of the non stationary part of the process with the stationary one.

1.3. Hurd-Gerr's test

This test relies on the property of correlation of the spectral measure of a cyclostationary process. We thus calculate a normalized spectral correlation:

$$\gamma(\lambda_j, h, M) = \frac{\left| \sum_{m=0}^{M-1} \widehat{I}_N(\lambda_{j+m}) \widehat{I}_N^*(\lambda_{j+h+m}) \right|^2}{\sum_{m=0}^{M-1} \left| \widehat{I}_N(\lambda_{j+m}) \right|^2 \sum_{m=0}^{M-1} \left| \widehat{I}_N(\lambda_{j+h+m}) \right|^2} \quad (5)$$

where $\widehat{I}_N(\lambda) = \frac{1}{\sqrt{2\pi N}} \sum_{n=0}^{N-1} x_n \exp(-i\lambda n)$, $\lambda_k = 2\pi k / N$ and M is a smoothing parameter. Then the result of Goodman is used: $P(\gamma > c) = (1-c)^{M-1}$ under given conditions [10].

1.4. Dandawaté-Giannakis's test

This test uses the asymptotic properties of the estimators of cumulants, and the test is formulated as follows: for a candidate cycle α one makes the following hypothesis testing:

$$\begin{aligned} H_0 &: \widehat{C}_{k,x}^N = \varepsilon_{k,x}^N \text{ for all arguments.} \\ H_1 &: \widehat{C}_{k,x}^N = C_{k,x}^N + \varepsilon_{k,x}^N \text{ for some arguments.} \end{aligned} \quad (6)$$

where $C_{k,x}^N$ is nonzero (it is the k^{th} cyclic-cumulant of cycle α of the process x), and $\varepsilon_{k,x}^N$ is a zero mean random variable. The asymptotic statistics of $\varepsilon_{k,x}^N$ are a classical result, from which an hypotheses test is built, allowing one to take a statistical decision.

2. Testing Cyclostationarity versus Stationarity

2.1. Expression of the test

We consider a Gaussian random process $(x_k)_{k\in\mathbb{Z}}$ which can be cyclostationary of period T or stationary. Without any loss of generality, we assume the process to be 0 mean. The aim of the test will be to decide wether:

$$\begin{aligned} H_0 &: (x_k)_{k\in\mathbb{Z}} \text{ is stationary} \\ H_1 &: (x_k)_{k\in\mathbb{Z}} \text{ is cyclostationary} \end{aligned} \quad (7)$$

We build our test on the basis of the likelihood ratio [2], whose optimal properties are well known:

$$\theta = \frac{P_{\Sigma_\omega}\left((x_k)_{k=1,\dots,n}\right)}{P_{\Sigma_\Omega}\left((x_k)_{k=1,\dots,n}\right)} \quad (8)$$

where Σ_ω is Toeplitz, and Σ_Ω is T-block-Toeplitz, which corresponds to the two hypothesis we are testing.

The log-likelihood of a Gaussian random process takes the well known following form:

$$\begin{aligned} L_n &= \log\left(P_n\left((x_k)_{k=1,\dots,n}\right)\right) \\ &= -\frac{1}{2}\left\{n\log(2\pi) + \log(|\Sigma|) + \underline{x}_n^\dagger \Sigma^{-1} \underline{x}_n\right\} \end{aligned} \quad (9)$$

Studying the asymptotic behaviour of the test is not straightforward, and an alternative consists in developping the "principal part" of the log-likelihood.

2.2. Approximant of the log-likelihood

We introduce the following notations:

$$\begin{aligned} Z_n &= L_{nT}(\mathcal{F}) + \frac{n}{4\pi}\left[\int_0^{2\pi} \log\left(\det\left(\mathcal{F}(\lambda)\right)\right)\right. \\ &\quad \left. + Tr\left(\widehat{\mathcal{I}}_n(\lambda)\mathcal{F}^{-1}(\lambda)\right) d\lambda\right] + \frac{nT\log(2\pi)}{2} \end{aligned} \quad (10)$$

$$\begin{aligned} \left[\widehat{\mathcal{I}}_n(\lambda)\right]_{(a,b)} &= \widehat{I}_{nT}\left(\tfrac{\lambda - 2\pi a}{T}\right) \widehat{I}_{nT}^*\left(\tfrac{\lambda - 2\pi b}{T}\right) \\ \left[\mathcal{F}(\lambda)\right]_{(a,b)} &= \mathcal{F}_{r_x}^{a-b}\left((\lambda - 2\pi b)/T\right) \end{aligned} \quad (11)$$

We will call $\mathcal{F}$, the Gladysev's spectral matrix associated with $(x_k)_{k\in\mathbb{Z}}$. The following scalar product will be required:

$$\langle F_0, F_1 \rangle_{1/2} = \sum_{k\in\mathbb{Z}} |k| \, Tr\left(F_0^\dagger[k] F_1[k]\right) \quad (12)$$

on the Banach space $\mathcal{B}_T$ of $T \times T$ matrix valued periodic functions $F \in \mathcal{L}^1$ whose associated series converge to F and $\|F\|_{1/2} < +\infty$. ($F[k]$ is the k^{th} Fourier coefficient of $F(.)$). We will also use the same notation for the scalar case: more precisely, when being concerned with scalar products containing $Tr(A)$ or $\det(A)$ the same notation will be used.

One can extend the results, available in the scalar Gaussian stationary case [3], concerning the asymptotic behaviour of the log-likelihood, of a cyclostationary Gaussian process [1], which leads to the following:

Theorem 1 *If $\mathcal{F}_0$ and $\mathcal{F}_1$ are the Gladysev's spectral matrices of two Gaussian cyclostationary random*

processes, such that $\mathcal{F}_0$ *and* $\mathcal{F}_1$ *belong to* $\mathcal{B}_T$ *and* $\det(\mathcal{F}_1) > 0$, *then*

$$\begin{aligned}\mathbb{E}_{\mathcal{F}_0}[Z_n(\mathcal{F}_1)] &= -\frac{1}{2}\left[\frac{1}{2}\left\|\log\left(\det(\mathcal{F}_1)\right)\right\|_{1/2}^2\right. \\ &\quad + \left\langle \log\left(\det(\mathcal{F}_1)\right), Tr\left(\mathcal{F}_0\mathcal{F}_1^{-1}\right)\right\rangle_{1/2} \\ &\quad \left. + \left\langle Tr\left(\mathcal{F}_0, Tr\left(\mathcal{F}_1^{-1}\right)\right)\right\rangle_{1/2}\right] + o(1) \\ \mathbb{V}ar_{\mathcal{F}_0}[Z_n(\mathcal{F}_1)] &= O(1) \end{aligned} \quad (13)$$

and when $\mathcal{F}_1 = I_T$ *then,*

$$\begin{aligned}\mathbb{V}ar_{\mathcal{F}_0}[Z_n(\mathcal{F}_1)] &= -\left\langle \log\left(\det(\mathcal{F}_1)\right), Tr\left(\mathcal{F}_1^{-2}\right)\right\rangle_{1/2} \\ &\quad -2\left\|Tr\left(\mathcal{F}_1^{-1}\right)\right\|_{1/2}^2 + o(1)\end{aligned}$$

Similar results have already been demonstrated in the multivariate case in several papers (see references in [6]), but the results presented here are more precise. These results will allow in practice, under given hypothesises, to obtain the precise rate of convergence to the approximation of the likelihood ratio.

2.3. Asymptotic behaviour of the test

This section holds with a direct application of the result obtained in the preceeding section, and allows one to define a degree of cyclostationarity relying on the likelihood ratio test, and has interpretation in terms of the Kullback-Leibler information.

The log likelihood ratio takes the following well known form:

$$\begin{aligned}\Theta_{NT} &= \log\left(p_0\left((x_n)_{n=1,\dots,NT}\right)\right) \\ &\quad -\log\left(p_1\left((x_n)_{n=1,\dots,NT}\right)\right)\end{aligned} \quad (14)$$

and we introduce the asymptotic form of the likelihood ratio:

$$\begin{aligned}\widetilde{\Theta}_{NT} &= \frac{N}{4\pi}\int_0^{2\pi}\left[\log\left(\det\left(\mathcal{F}(\lambda)\right)/\left(\mathcal{F}_{r_x}^0(\lambda)\right)^T\right)\right. \\ &\quad \left. + Tr\left\{\widehat{\mathcal{I}}_N(\theta)\left(\mathcal{F}^{-1}(\lambda) - \left(\mathcal{F}_{r_x}^0(\lambda) I_s\right)^{-1}\right)\right\}\right] d\lambda\end{aligned} \quad (15)$$

The results introduced in the preceeding section shows that:

$$\left|\mathbb{E}_{0/1}\left(\widetilde{\Theta}_{NT} - \Theta_{NT}\right)\right| \overset{N\to+\infty}{=} O(1) \quad (16)$$

Furthermore, one has the following Lemma (see references in [6]), whose scalar proof was given by Grenander and Rosenblatt:

Lemma 2 *Let* F *be a bounded odd matrix valued functions, and* $\widehat{I}_n$ *the multidimensionnal periodogram associated with a realisation of a Gaussian processes of spectral* Φ *density then, as* $n \to +\infty$:

$$\mathbb{E}_\Phi\left(\int_0^{2\pi} Tr\left(\widehat{I}_n F\right) d\lambda\right) \to \int_0^{2\pi} Tr\left(\Phi F\right) d\lambda \text{ and,}$$

$$\mathbb{V}ar_\Phi\left(\int_0^{2\pi} Tr\left(\widehat{I}_n F\right) d\lambda\right) \sim \frac{4\pi}{n}\int_0^{2\pi} Tr\left((F\Phi)^2\right) d\lambda$$

This Lemma can be straightforwardly extended to the cyclostationary case [1], where F is replaced by the Gladysev's spectral matrix $\mathcal{F}$.

Then, using this Lemma, and considering $-\frac{1}{N}\widetilde{\Theta}_{NT}$ as $N \to +\infty$ under hypotheses H_1, one obtains:

$$-\frac{1}{N}\widetilde{\Theta}_{NT} \overset{q.m.}{\to} \mu \quad (17)$$

where

$$\begin{aligned}\mu &= -\frac{1}{4\pi}\int_0^{2\pi}\left[\log\left(\det\left(\mathcal{F}\right)/\left(\mathcal{F}_{r_x}^0\right)^T\right)\right. & (18)\\ &\quad \left. + T - Tr\left(\mathcal{F}/\mathcal{F}_{r_x}^0\right)\right] d\lambda & (19)\end{aligned}$$

which can be interpreted according to the Kullback-Leibler information of the corresponding cyclostationary and stationary processes that is

$$\mu = K\left(\mathcal{F}, \mathcal{F}_{r_x}^0 I_T\right) \quad (20)$$

3. Asymptotic probabilities

In this section, we study the asymptotic power α and the asymptotic false alarm probability β of the likelihood ratio test. These results have been developped in the framework of stationary scalar Gaussian processes [4]. As shown below, the proof of the results in the cyclostationary case, is not different from what can be found in [4]. The results had only to be checked and written once for all.

3.1. Preliminary

Here we recall the following theorem:

Theorem 3 *If* $\log(\det(\mathcal{F}_0))$ *and* $\log(\det(\mathcal{F}_1))$ *belong to* $\mathcal{L}^1$ *then for* $t \in [0,1]$

$$\mathcal{Q}_{nT}(t) \overset{n\to+\infty}{\to} \mathcal{Q}_0(t)$$

where

$$\begin{aligned}\mathcal{Q}_0(t) &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[t\log\left(\det\left(\mathcal{F}_0\right)\right)\right. \\ &\quad + (1-t)\log\left(\det\left(\mathcal{F}_1\right)\right) \quad (21) \\ &\quad \left. - \log\left(t\det\left(\mathcal{F}_0\right) + (1-t)\det\left(\mathcal{F}_1\right)\right)\right] d\lambda\end{aligned}$$

$$\mathcal{Q}_{nT}(t) = \frac{1}{nT}\log\left[\mathbb{E}_0\left(\exp\left(t\log\left(\frac{dP_1^{\otimes nT}}{dP_0^{\otimes nT}}\right)\right)\right)\right]$$

The calculus for the proof are the same matrix manipulations in the block-Toeplitz case as in [4] for the stationary case, and Szëgo's theorem for the cyclostationary case (which is required for the proof), is implicitely shown in [1].

The series $\left(P_0^{\otimes n}, P_1^{\otimes n}\right)$ of probabilities associated with the spectral matrices $\mathcal{F}_0$ and $\mathcal{F}_1$, is said in that case to be of Chernoff.

3.2. Formula "à la Chernoff"

The following theorem is demonstrated in [4] for the stationary scalar case, and the extension to the cyclostationary case is not difficult [1].

Theorem 4 *Let us consider a series* $\left(P_0^{\otimes n}, P_1^{\otimes n}\right)$ *of Chernoff, we further suppose that:*

1. $K\left(P_0, P_1\right)$ *and* $K\left(P_1, P_0\right)$ *exist.*
2. *We define* $a \in \left[-K\left(P_0, P_1\right), K\left(P_1, P_0\right)\right]$

then, as $n \to +\infty$

$$\frac{1}{n}\log\left(P_0^{\otimes n}\left(\frac{1}{n}\log\left(\frac{dP_1^{\otimes n}}{dP_0^{\otimes n}}\right) \geq a\right)\right) \quad \to \quad -h_0\left(a\right)$$

$$\frac{1}{n}\log\left(P_1^{\otimes n}\left(\frac{1}{n}\log\left(\frac{dP_1^{\otimes n}}{dP_0^{\otimes n}}\right) \leq a\right)\right) \quad \to \quad a + h_0\left(a\right)$$

where

$$h_0\left(a\right) = \sup_{\theta \in \mathbb{R}} \left(\theta a - \mathcal{Q}_0\left(\theta\right)\right) \tag{22}$$

and $K\left(P_0, P_1\right)$ is the Kullback information between the two processes.

Thus in the cyclostationary case, in the framework of our test, the asymptotic power and false alarm of the Neyman-Pearson test can be written in the interesting following manner:

$$\begin{aligned} \alpha\left(a\right) &= -h\left(a\right) \text{ and } \beta\left(a\right) = a - \alpha\left(a\right) \\ h_1\left(a\right) &= \sup_{\theta \in \mathbb{R}} \left(\theta a - \mathcal{Q}_1\left(\theta\right)\right) \end{aligned} \tag{23}$$

4. Conclusion

We propose a test for cyclostationarity based on the theory of decision which gives birth to a measure of cyclostationarity which can be connected to the Kullback-Leibler information of information theory. This test allows one to test wether a frequency is a cyclic frequency of a given process, and gives an indicator of the degree of cyclostationarity according to this cyclic frequency. This work has two main interests: we provide a statistical test, a strong theoretical background to the notion of degree of cyclostationarity first introduced by Zivanovic and Gardner, and a new measure of cyclostationarity. Application of this test will appear in a later paper.

References

[1] C. Andrieu, P. Duvaut, "Asymptotic Behaviour of the likelihood of a Multivariate Stationary Gaussian Time-Series - Application to Cyclostationary Gaussian Processes" Technical Report from ENSEA-ETIS Groupe-Signal.

[2] A. Borovkov, Statistique mathématique, Édition Mir (1987) (french translation of Mathematical Statistics Nauka-Moscow).

[3] J. Coursol, D. Dacunha-Castelle, "Remarks on the Approximation of the Likelihood Function of Stationary Gaussian Processes", Theory of Probabilities and its Applications 1982.

[4] D. Dacunha-Castelle, "Séminaire d'Orsay sur les grandes déviations", Astérisque 68, S.M.F.

[5] A. V. Dandawaté, G. B. Giannakis, "Statistical Tests for the Presence of Cyclostationarity", IEEE Transactions on Signal Processing, Vol. 42, No. 9, September 1994.

[6] K. Dzhaparidze, Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series Springer-Verlag 1986.

[7] W.A. Gardner, "Measurement of Spectral Correlation", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 11, No. 1, July 1986, North-Holland.

[8] W.A. Gardner, G. Zivanovic, "Degrees of Cyclostationary and their Application to Signal Detection and Estimation", Signal Processing Vol. 22, No. 3, March 1991, Elsevier.

[9] E. G. Gladysev, "Periodically Correlated Random Sequences", Soviet Mathematics 1961, vol.2.

[10] H. L. Hurd, N. L. Gerr, "Graphical Methods for Determining the Presence of Periodic Correlation", Journal of Time Series Analysis, Vol. 12, No. 4.

Empirical Determination of the Frequencies of an Almost Periodic Sequence

D. Dehay
IRMAR
University of Rennes
35042 Rennes
FRANCE
Dehay@Uhb.Fr

H.L. Hurd
Center for Stochastic Processes
University of North Carolina
Chapel Hill, NC 27599-2630
hurd@stat.unc.edu

Abstract

This note is concerned with the problem of determination of the countable set $\Lambda = \{\lambda_1, \lambda_2, ...\}$ *of frequencies belonging to an almost periodic sequence by methods in which a finite number of frequencies* $\{\lambda_1^{(n)}, \lambda_2^{(n)}, ..., \lambda_{K_n}^{(n)}\} = \Lambda_n$ *are are produced at each stage n. We seek algorithms for which* Λ_n *converges to* Λ *but yet each* Λ_n *is not too big.*

1. Introduction

Almost periodic sequences share many properties of AP (almost periodic) functions. [1] For example, the Fourier coefficient

$$a(\lambda) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} x_k \exp(-i\lambda k) \qquad (1)$$

exists for every λ and Parseval's equality

$$\sum_{\lambda\in\Lambda} |a(\lambda)|^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} |x_k|^2$$

implies that the set of frequencies $\Lambda = \{\lambda : a(\lambda) \neq 0\}$ belonging to the AP sequence, is countable. In the case of AP sequences the frequencies may all be taken in the interval $[0, 2\pi)$. The frequencies Λ and coefficients $\{a(\lambda), \lambda \in \Lambda\}$ are uniquely determined by the AP sequence X and for this reason it is said that each AP sequence has an associated (unique) Fourier series,

$$x_k \sim \sum_{j=-\infty}^{\infty} a(\lambda_j) \exp(i\lambda_j k) \qquad (2)$$

where $\sim$ is not to be taken as equality except under additional assumptions, for example, when Λ is finite.

This work is motivated by some problems of estimation [4, 3] and detection [2] for stochastic processes connected with almost periodicity. Here we begin to treat a simpler problem, the empirical determination of the frequencies Λ of an AP sequence, to illustrate a problem involving finite computation. That is, suppose we are given the sequence $\{x_k\}$ one element at a time; in other words, we are given the finite sequences $X_n = \{x_k, k = 0, 1, ..., n\}$ for $n = 0, 1, ...$ Now if we are told, a priori, that some $\lambda \in \Lambda$, then from (1), the sequence

$$a_n(\lambda) = \frac{1}{n} \sum_{k=0}^{n-1} x_k \exp(-i\lambda k) \qquad (3)$$

converges to $a(\lambda) \neq 0$, and if $\lambda \notin \Lambda$, then $a_n(\lambda) \to 0$. But if Λ is unknown a-priori, we are faced with performing the limit[1] for an uncountable collection of λ to determine the countable set Λ on which $a(\lambda) \neq 0$.

So given the practical constraint of finite computations, we wish to determine Λ by some sort of limit of a sequence of operations, each of which involves only a finite number of calculations. We address the problem in the following manner: at each computation stage n we compute a finite number of frequencies

$$\{\lambda_1^{(n)}, \lambda_2^{(n)}, ..., \lambda_{K_n}^{(n)}\} = \Lambda_n \qquad (4)$$

using only the subsequence X_n. We seek algorithms for determining Λ_n from X_n that have the following properties:

[1] In the context where the index k corresponds to time, we must know the sequence for infinite time.

0-8186-7576-4/96 $5.00

1. Λ_n converges to Λ in the sense that for every $\lambda \in \Lambda$ there is a sequence $\{\lambda_n\}$ with $\lambda_n \in \Lambda_n$ and $\lambda_n \to \lambda$;

2. Λ_n is not too big in the sense that convergent sequences taken from the Λ_n converge only to elements of Λ; this additional constraint is needed because the finite set $\{2\pi j/n, j = 0, 1, ..., n-1\}$ satisfies the property 1., but yet is too big in the sense that some elements may not be near an element of Λ;

2'. a sufficient condition for 2. is that there exists a resolution function $\epsilon(n) \to 0$ such that for every $\lambda' \in \Lambda_n$ there is a $\lambda \in \Lambda$ with

$$|\lambda' - \lambda| < \epsilon(n). \tag{5}$$

Our approach is through the *weighted* Fourier coefficient estimator

$$a_n^w(\lambda) = \frac{\exp(i\lambda n)}{2n} \sum_{k=0}^{2n} x_k w_{k+n}^{(n)} \exp(-i\lambda k) \tag{6}$$

where $w_k^{(n)}$ is the *Bartlett* weight sequence that has considerable application in spectral density estimation. It is given by

$$w_k^{(n)} = \begin{cases} 2(1 - |k/n|) & |k| < n \\ 0 & |k| \geq n \end{cases} \tag{7}$$

and in order to keep the estimation procedure one-sided as in (6), we center the sequence at n. The discrete Fourier transform of (7) is

$$\begin{aligned} W_n(\lambda) &= \sum_{k=-(n-1)}^{n-1} w_k \exp(-i\lambda k) \\ &= \frac{2}{n} \frac{sin^2(n\lambda/2)}{sin^2(\lambda/2)} \end{aligned} \tag{8}$$

and this is related to (6) by

$$\exp(i\lambda n) \sum_{k=0}^{2n} w_{k+n}^{(n)} \exp(-i\lambda k) = W_n(\lambda). \tag{9}$$

The factor of 2 appears in (7) in order to ensure $lim_{n\to\infty} a_n^w(\lambda) = a(\lambda)$ and the $\exp(i\lambda n)$ is needed to account for the centering of the window in $[0, 2n]$ rather than $[-n, n]$. The fraction $1/2n$ appearing in (6) is used in place of $1/(2n+1)$ for ease in computation and in the presentation.

Note that $W_n(\lambda)$ is periodic with period 2π, $W_n(\lambda) = W_n(\lambda + 2\pi)$, and $W_n(0) = 2n$.

The fact $W_n(0) = 2n$ and the observation that $|W_n(\pi)|$ does not exceed $2/n$ together motivate the following lemma.

Lemma 1 *For the Bartlett kernel $W_n(\lambda)$, given $\delta > 0$ there exists a number $K > 0$ and an integer n_0 for which*

$$\frac{W_n(\lambda)}{W_n(0)} < \frac{K}{n^2} \tag{10}$$

for all $|\lambda| > \delta$ and $|\lambda - 2\pi| > \delta$, provided $n > n_0$.

Proof. Choose $n_0 > \pi/\delta$, then

$$\frac{W_n(\lambda)}{W_n(0)} = \frac{2}{n^2} \frac{sin^2(n\lambda/2)}{sin^2(\lambda/2)} \leq \frac{2}{n^2} \frac{1}{sin^2(\delta/2)} \tag{11}$$

and the result follows by the identification $K = 2/sin^2(\delta/2)$. ∎

2. Λ is a singleton

In this case,

$$x_k = a_0 \exp(i\lambda_0 k) \tag{12}$$

and for any λ we have

$$a_n^w(\lambda) = \frac{a_0}{2n} W_n(\lambda - \lambda_0). \tag{13}$$

It may be easily shown that $W_n(\lambda - \lambda_0)$ has a global maximum at λ_0 and is locally maximum in the neighborhood of λ_0 given by $|\lambda - \lambda_0| < \pi/n$.

The following procedure meets the constraint of finite computation at each stage n and produces the required convergence of Λ_n to Λ. First, compute $a_n^w(\lambda)$ for the distinct values $\lambda_j^{(n)} = j2\pi/n, j = 0, 1, ...n-1$ and put $\lambda_{j_0}^{(n)}$ into Λ_n provided it maximizes $|a_n^w(\lambda_j^{(n)})|$ for $j = 0, 1, ..., n-1$.

The localization properties of the function $W_n(\lambda)$ show that each set Λ_n will contain at most two elements, the two $\lambda_j^{(n)}$ surrounding λ_0. That is, if $\lambda_{j_0}^{(n)} \leq \lambda_0 \leq \lambda_{j_0+1}^{(n)}$, then one or both of $\lambda_{j_0}^{(n)}, \lambda_{j_0+1}^{(n)}$ will be elements of Λ_n and so every element $\lambda^{(n)} \in \Lambda_n$ satisfies $|\lambda^{(n)} - \lambda_0| < \frac{1}{n}$.

3. Λ is finite

When Λ is finite, say $card(\Lambda) = J$,

$$x_k = \sum_{j=1}^{J} a_j \exp(i\lambda_j k) \tag{14}$$

and note that there exists a $\delta > 0$ for which $|\lambda_j - \lambda_{j'}| > \delta$ for all j, j' with $j \neq j'$.

We will choose the elements of Λ_n to be the *strong local* peaks among the finite collection of $a_n^w(\lambda)$ computed at each stage.

To complete the argument, first take n sufficiently large, say $n > n_0$, so that $\frac{(\lambda''-\lambda')}{2\pi/n} > 10$. For any fixed $n > n_0$, let

$$\lambda_{j^*}^{(n)} = \arg \max_{\lambda_j^{(n)} \in (\lambda', \lambda'')} |a_n^w(\lambda_j^{(n)})|.$$

But for there to be a strong local peak at $\lambda_{j^*}^{(n)}$) also requires

$$|a_n^w(\lambda_{j^*}^{(n)})| > 10|a_n^w(\lambda_j^{(n)})|$$

for $3C \leq |j^* - j| \leq 4C$, which thus requires

$$\frac{|a_n^w(\lambda_{j^*}^{(n)})|}{|a_n^w(\lambda_{j^*+4C}^{(n)})|} > 10$$

From the preceeding discussion, for sufficiently large n, this ratio must converge to 1 and hence will not exceed, say 2. It may be seen that for sufficiently large n, no $\lambda_j^{(n)} \in (\lambda', \lambda'')$ will be selected as a strong local peak.

Finally we note again that we do not know Λ or the numbers $\{a(\lambda), \lambda \in \Lambda\}$ a-priori and so the value of n_0 for which $n > n_0$ produces these results is not known to us. One cannot say when these events occur, only that they will.

4. Λ has isolated cluster points

Here we consider the class of AP sequences for which $\sum_\Lambda |a(\lambda)| < \infty$. Suppose λ_0 is the cluster point of Λ and $a_0 = a(\lambda_0)$. We shall assume λ_0 is an interior point of $[0, 2\pi)$, and omit the adjustments for the case when $\lambda_0 = 0$. For any $\epsilon > 0$ there is a deleted neighborhood $D(\lambda_0, \delta) = (\lambda_0 - \delta, \lambda_0 + \delta) \backslash \{\lambda_0\}$ of λ_0 for which

$$\sum_{\lambda_j \in D(\lambda_0, \delta)} |a_j| < \epsilon/2.$$

Take $\epsilon = |a_0|/10$. The examination of $a_n^w(\lambda)$ at the point λ_0 will help us understand the new situation. Consider then

$$a_n^w(\lambda_0) = a_0 + \frac{1}{2n} \sum_{\lambda_j \in D(\lambda_0, \delta)} a_j W_n(\lambda_0 - \lambda_j) + \frac{1}{2n} \sum_{\lambda_j \notin D(\lambda_0, \delta)} a_j W_n(\lambda_0 - \lambda_j) \qquad (22)$$

Because the last sum is finite we know from the previous section it is $O(\frac{1}{n^2})$ so there is an n_0 for which $n > n_0$ implies this quantity will not exceed $\epsilon/2$. As for the middle term,

$$|\frac{1}{2n} \sum_{\lambda_j \in D(\lambda_0, \delta)} a_j W_n(\lambda_0 - \lambda_j)|$$

$$\leq |\frac{1}{2n} \sum_{\lambda_j \in D(\lambda_0, \delta)} |a_j||W_n(\lambda_0 - \lambda_j)|$$

$$\leq \sum_{\lambda_j \in D(\lambda_0, \delta)} |a_j| \leq \epsilon/2 \qquad (23)$$

Of course we already know that $a_n^w(\lambda_0) \to a(\lambda_0)$ but this permits us to see what happens when we evaluate $a_n^w(\lambda)$ at the nearest point $\lambda_j^{(n)*}$ to λ_0. For n sufficiently large, the points not in $D(\lambda_0, \delta)$ will contribute to $a_n^w(\lambda_j^{(n)*})$ as described in the finite case. The middle term contributes at most $\epsilon/2$ (or $|a_0|/10$) and so $a_n^w(\lambda_j^{(n)*})$ will eventually dominate the rest (in this example, by a factor of 10).

The demonstration that Λ_n is not too big is in progress; we expect to follow as in the finite case. The idea is to show that any interval (λ', λ'') containing no points of Λ, $a_n^w(\lambda)$ will ultimately not have any strong local peaks. Even though Λ is infinite, there are only a finite number of frequencies outside a neighborhood of λ_0, and those inside have summable amplitudes, so their contribution to $a_n^w(\lambda)$ for $\lambda \in (\lambda', \lambda'')$ will eventually become negligible.

5. Acknowledgements

This work was supported by the Office of Naval Research under contract N00014-95-C-0093.

References

[1] C. Corduneanu, *Almost Periodic Functions*, Chelsea: New York, 1989.

[2] A. V. Dandawate and G. B. Giannakis, "Statistical Tests for the Presence of Cyclostationarity", *IEEE Trans. on Signal Processing*, vol. 42, no. 9, pp. 2355-2369, 1994.

[3] D. Dehay and H.L. Hurd , "Representation and Estimation for Periodically and Almost Periodically Correlated Random Processes", in "Cyclostationarity in Communications and Signal Processing", W.A. Gardner, Ed., IEEE Press, 1993.

[4] H. L. Hurd and J. Leskow, "Estimation of the Fourier Coefficient Functions and Their Spectral Densities for ϕ−Mixing Almost Periodically Correlated Processes," *Statistics and Probability Letters*, vol. 14, No. 4, pp. 299-306, 1992.

To be more specific, we compute $a_n^w(\lambda)$ for the values $\lambda_j^{(n)} = j2\pi/Cn, j = 0, 1, ...Cn-1$ where C is a positive integer. For some arbitrary $\lambda_{j_0} \in \Lambda$ denote $\lambda_{j,j_0}^{(n)}$ as the $\lambda_j^{(n)}$ closest to λ_{j_0} and consider the expression

$$a_n^w(\lambda_{j,j_0}^{(n)}) = \frac{1}{2n} a_{j0} W_n(\lambda_{j,j_0}^{(n)} - \lambda_{j_0}) + \frac{1}{2n} \sum_{\substack{j=1 \\ j \neq j_0}}^{J} a_j W_n(\lambda_{j,j_0}^{(n)} - \lambda_j). \quad (15)$$

If the $\lambda_j^{(n)}$ are sufficiently dense with respect to the kernel $W(\cdot)$, then the first term on the right is close to a_{j_0}, say

$$|\frac{1}{2n} a_{j0} W_n(\lambda_{j,j_0}^{(n)} - \lambda_{j_0})| \geq .9|a_{j0}| \quad (16)$$

and the second term is $O(\frac{1}{n^2})$ because all the remaining λ_j are "far" from $\lambda_{j,j_0}^{(n)}$. We are lead to conclude that there exists an interval $I_n(j_0, j_1)$ between λ_{j_0} and the next largest λ_j (call it λ_{j_1}) for which

$$\left| \frac{a_n^w(\lambda_j^{(n)})}{a_n^w(\lambda_{j,j_0}^{(n)})} \right| = \frac{O(\frac{1}{n^2})}{O(a_{j0}) + O(\frac{1}{n^2})} = O(\frac{1}{n^2}) \quad (17)$$

for all $\lambda_j^{(n)}) \in I_n(j_0, j_1)$. In other words, for sufficiently large n, the value of $|a_n^w(\lambda_{j,j_0}^{(n)})|$ will begin to dominate, as n^2, the values taken in a nearby interval. Immediately neighboring values may have to be excluded because of the shape of $W(\cdot)$.

A frequency index j^* is said to produce a *strong local* peak with parameters K_1, K_2, K_3 if

$$|a_n^w(\lambda_{j^*}^{(n)})| > |a_n^w(\lambda_j^{(n)})|$$

for $|j^* - j| \leq K_1$, and

$$|a_n^w(\lambda_{j^*}^{(n)})| > K_3 |a_n^w(\lambda_j^{(n)})|$$

for $K_2 \leq |j^* - j| \leq K_1$. So a *strong local* peak is at least K_3 times larger than it's neighbors except for those nearby ($|j^* - j| < K_2$). The elements of Λ_n are the frequencies λ_{j^*} associated with the *strong local* peaks.

In general, the value of K_3 is to be considered large (e.g. $K_3 = 10$ or $K_3 = 100$), and the values of K_1 and K_2 will depend on the Fourier transform of the weight sequence w_k.

For some arbitrary $\lambda_{j_0} \in \Lambda$, if we take $C = 4$, $K_1 = 16, K_2 = 12$ and $K_3 = 10$ with the Bartlett sequence, then for n sufficiently large we will obtain a $\lambda_{j,j_0}^{(n)}$ that satisfies $|\lambda_{j,j_0}^{(n)} - \lambda_{j_0}| < 2\pi/Cn$. Hence for each $\lambda_{j_0} \in \Lambda$ one may find a sequence $\{\lambda_{j_0}^{(n)}\}$ with $\lambda_{j_0}^{(n)} \in \Lambda_n$ and $\lambda_{j_0}^{(n)} \to \lambda_{j_0}$.

Now we must show that the Λ_n are not too big. We will show that in any open interval (λ', λ'') containing no points of Λ, $a_n^w(\lambda)$ will ultimately not have any strong local peaks.

Consider the expression

$$a_n^w(\lambda) = \frac{1}{2n} \sum_{j'=1}^{J} a_{j'} W_n(\lambda - \lambda_{j'}) \quad (18)$$

for λ in (λ', λ''). Although the values $|a_n^w(\lambda)|$ can vary significantly for $\lambda \in (\lambda', \lambda'')$, we will now show that

$$\lim_{n \to \infty} \frac{|a_n^w(\lambda + 2\pi/n)|}{|a_n^w(\lambda)|} \to 1 \quad (19)$$

for λ in this interval. Clearly $\lambda + 2\pi/n \to \lambda$ as $n \to \infty$.

The ratio in (19) may be expressed as

$$\frac{\sum_{j'=1}^{J} \frac{A_{j'}^{(n)}(\lambda + 2\pi/n)}{B_{j'}(\lambda + 2\pi/n)}}{\sum_{j'=1}^{J} \frac{A_{j'}^{(n)}(\lambda)}{B_{j'}(\lambda)}} \quad (20)$$

where $A_{j'}^{(n)}(\lambda) = a_{j'} sin^2[n(\lambda - \lambda_{j'})/2]$ and $B_{j'}(\lambda) = sin^2[(\lambda - \lambda_{j'})/2]$. Because $A_{j'}^{(n)}(\lambda + 2\pi/n) = A_{j'}^{(n)}(\lambda)$ is bounded with respect to n, the following Lemma may be applied.

Lemma 2 *If $\{\alpha_j^{(n)}\}_1^\infty, j = 1, 2, ..., J$ are each bounded sequences and if $\{\beta_j^{(n)}\}_1^\infty, j = 1, 2, ..., J$ are each convergent sequences with $\beta_j^{(n)} \to \beta_j$, then*

$$lim_{n \to \infty} \frac{\sum_{j'=1}^{J} \alpha_{j'}^{(n)} \prod_{\substack{k=1 \\ k \neq j'}}^{J} \beta_k^{(n)}}{\sum_{j'=1}^{J} \alpha_{j'}^{(n)} \prod_{\substack{k=1 \\ k \neq j'}}^{J} \beta_k} = 1. \quad (21)$$

The proof follows from setting $\beta_j^{(n)} = \beta_j + \eta_j^{(n)}$ where $\eta_j^{(n)} \to 0$ for all j.

In our current problem we set $\alpha_j^{(n)} = A_j^{(n)}(\lambda)$, $\beta_j^{(n)} = B_j(\lambda + 2\pi/n)$ and $\beta_j = B_j(\lambda)$, and we note that $\beta_j^{(n)} \to \beta_j$ because $B_j(\lambda)$ are continuous functions of λ. Thus by application of the Lemma we obtain (19). Furthermore, considered as functions of λ, the convergence is uniform because the α's are uniformly bounded and the β's come from shifts of the single continuous function $1/sin^2(x)$ for x bounded away from 0 and 2π. And all this is also true when $\lambda + 2\pi/n$ is replaced with $\lambda \pm K2\pi/n$ for finite (fixed) K.

Detection of Affine Time-Delay Multipath Striations using a Generalized Cyclostationary Model

Douglas E. Lake
Office of Naval Research, Code 313
800 N. Quincy St.
Arlington, VA 22217
laked@onrhq.onr.navy.mil

Abstract

An important feature available in certain scenarios for the underwater sonar detection of broadband signals is the formation of "striated" patterns in spectrograms. The observed striations can be modeled by a broadband multipath signal whose time delay between arrivals is slowly varying linearly. This model leads to generalized notion of cyclostationarity where the cyclic frequency varies linear with frequency. In this paper, noncoherent and coherent methods to detect these broadband energy patterns based on this model are presented and demonstrated on a nontrivial example. While noncoherent methods cannot distinguish between the positive and negative delay rates, coherent methods determine both the sign in addition to an estimate of the initial multipath delay.

1. INTRODUCTION

Very low frequency (VLF) underwater signals can be exploited at long ranges because of their propagation characteristics. In certain scenerios, an important characteristic of VLF broadband energy propagation is the formation of "striated" patterns in the time-frequency domain (i.e., spectrograms) caused by multipath interference. This phenomenon is sometimes referred to as "Lloyd's Mirror" by the sonar community. In [3], acoustic propagation models were demonstrated to predict these VLF multipath interference patterns and a variety of detection methods were developed. In this paper, methods to detect these broadband energy patterns based on a generalized notion of cyclostationarity are presented and demonstrated on realistic simulated data.

2. Affine Time Delay Multipath Model

The observed striation patterns can be modeled by a broadband multipath signal whose time delay between arrivals is varying linearly. This linear relationship could perhaps arise from a variety of environmental factors, but in practice the cause is the motion of the source. In a multipath environment, the received signal $x(t)$ in the time domain can be written as

$$x(t) = s(t) + \gamma s(t - \tau(t)) + n(t), \quad 0 \le t \le T \quad (1)$$

where $s(t)$ is the direct path stationary signal $s(t - \tau(t))$ is a delayed version of the signal from an alternative path (possibly with a different amplitude), and $n(t)$ is a stationary process representing the aggregate effects of all noise factors.

A first order approximation to the delay function τ is linear so that $\tau(t) = \alpha t + \beta$. The delay is assumed to be slowly varying so that the delay rate α is small. In order to distinguish the two paths in the model, the initial multipath delay β is assumed to be strictly positive. With this simplification, (1) becomes

$$x(t) = s(t) + \gamma s(t - \alpha t - \beta) + n(t), \quad 0 \le t \le T \quad (2)$$

which will be referred to as the *affine time delay model.* In [2], several applications are discussed where the affine time delay model has been used to model the propagation effects on signals from sources in motion.

We assume the delay rate is sufficiently small that the signal is approximately stationary over short time windows. In particular, the expected amount of energy at frequency ω at time t, denoted by $P(t,\omega)$, is well-approximated by

$$P(t,\omega) = (1+\gamma^2+2\gamma\cos(\omega\alpha t+\omega\beta))S(\omega)+N(\omega) \quad (3)$$

where $S(\omega)$ and $N(\omega)$ are the power spectral densities of the signal and noise, respectively. For simplicity, this can

0-8186-7576-4/96 $5.00

be written alternatively as

$$P(t,\omega) = A(\omega) + B(\omega)\cos(\omega\alpha t + \omega\beta) \qquad (4)$$

where $A(\omega)$ represents the energy of the agregate stationary components and $B(\omega)$ represents the peak energy of the nonstationary or *striating* components.

In our discussion $P(t,\omega)$ is the spectrogram, that is, the short-time Fourier transform (STFT). In principle, however, P could be any time-frequency distribution and could be optimally matched to this model (as is discussed in the conclusions section). This spectrogram model contains interference patterns with peaks along the following family of hyperbolic curves

$$\omega\alpha t + \omega\beta = k, \quad k = 0, \pm 1, \pm 2, \ldots \qquad (5)$$

as is shown in the example in Figure 1 with Nyquist frequency normalized to 1.

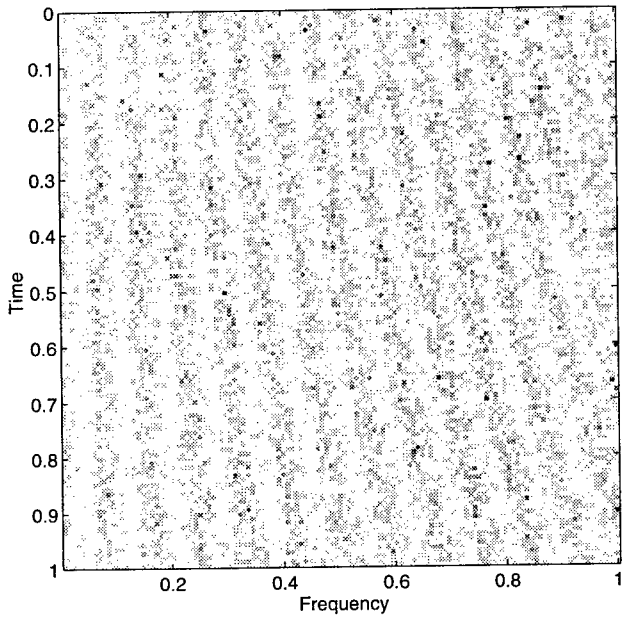


Figure 1. Spectrogram for Affine Multipath Delay Model with White Signals and Noise

The signal $s(t)$ in this example is band-limited white noise with a cutoff frequency equal to Nyquist. The delayed signal is obtained by an initial interpolation using the MATLAB function INTERP with an oversampling factor of 4 and additional resolution using linear interpolation. The results do not change significantly using higher initial oversampling factors. The noise $n(t)$ is white and the SNR is 0dB, i.e., the noise energy equals that of $s(t)$. The received signal time series $x(t)$ consists of $N = 2^{15} = 32768$ samples over $T = 1$ second. The spectrogram is calculated with 128 nonoverlapping FFT's of length 256 resulting in a 128×128 image. The multipath parameters are $\gamma = 1$, $\alpha = -8$ samples/second and $\beta = 32$ samples. So the multipath delay varies from an initial value of $\tau(0) = \beta = 32$ samples to a final value of $\tau(1) = 24$ samples over the second interval. Note in Figure 1 that there are $\tau(0)/2 = 16$ peaks at time 0, $\tau(1)/2 = 12$ peaks at time 1, and $|\alpha|/2 = 4$ peaks for frequency 1.

3. Generalized Cyclostationary Approaches

Although the process $x(t)$ is nonstationary, it does have some stationarity properties that allow the application of traditional signal processing methods. For each fixed frequency ω, consider the marginal process $Q(t) = Q_\omega(t) = P(t,\omega)$. This process is stationary with significant spectral energy at *cyclic frequency* $\lambda = \alpha\omega$. Moreover, the *cyclic phase* of Q is $\beta\omega$. Because the cyclic frequency (or one over the cyclic period) is not constant, but varies with frequency (in this case linearly), $x(t)$ can be called a *generalized cyclostationary process.*

3.1. NonCoherent Detector

A simple noncoherent detector of the delay rate can be constructed based on these generalized cyclostationary properties. The first step is to estimate the stationary component $A(\omega)$ in (4) and subtract it out of each frequency column. Prewhitening of the received signal can be important in practice, but in our example we assume the signal is already white for the sake of simplicity. The power spectral density can estimated for each fixed frequency giving a two dimensional function of frequency ω and cyclic frequency λ. Energy integrated over frequency and the cyclic frequency bins corresponding to $\lambda = \alpha\omega$ (i.e., lines through the origin) provide a measure of the marginal likelihood for a given delay rate. These steps are simplified notationally by first defining the *cyclostationary gram* $C(\alpha,\omega)$ by

$$C(\alpha,\omega) = \int_0^T e^{-i\alpha\omega t} P(t,\omega)dt \qquad (6)$$

which effectively accounts for the linear relationship between cyclic frequency and frequency (i.e., concentrates all the energy along a vertical line). Figure 2 shows the cyclostationary gram for our example. The noncoherent delay rate likelihood function $L(\alpha)$ can now be written as

$$L(\alpha) = \int_0^\pi |C(\alpha,\omega)|^2 d\omega \qquad (7)$$

which is shown in Figure 3.

3.2. Two-parameter Coherent Detector

Note that $L(\alpha)$ is symmetric about 0. This points out one inadequacy of this approach, that is, only absolute

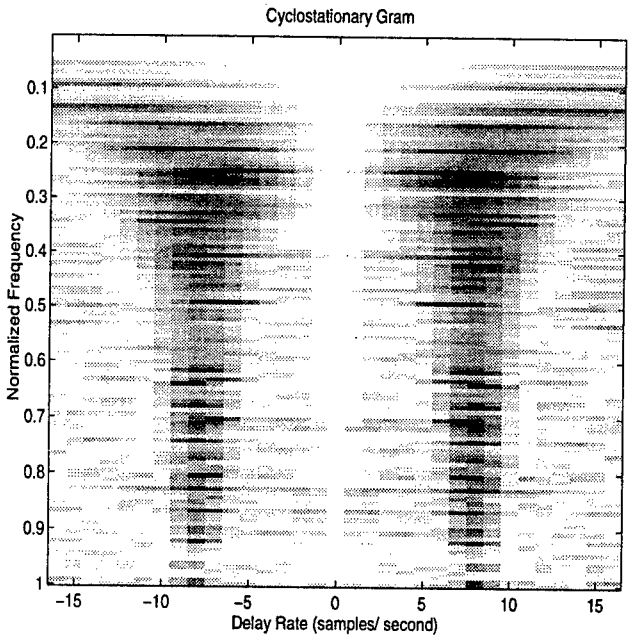


Figure 2. Cyclostationary gram taking into account linear dependency of cyclic frequency on frequency.

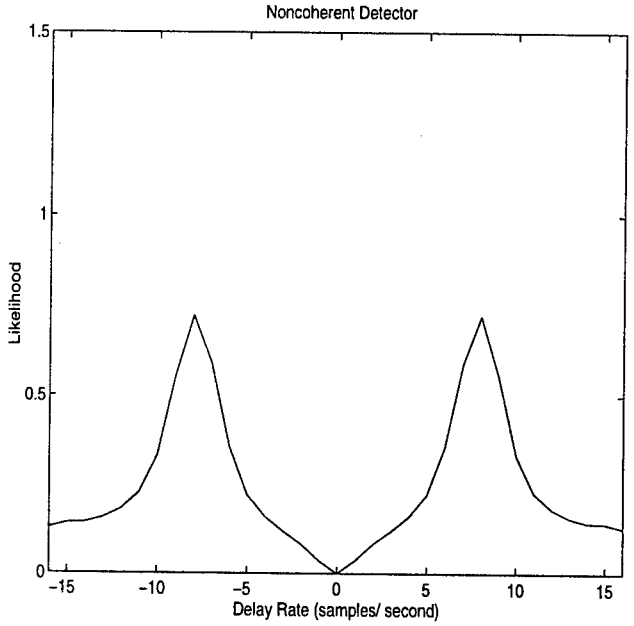


Figure 3. Noncoherent likelihood function for multipath delay rate parameter α.

value of the delay rate is being detected and there is no discriminatin between sources that are opening and closing the receiver at the same speed. We therefore consider a coherent detection approach with a two-parameter joint likelihood function $L(\alpha, \beta)$ which incorporates the cyclic phase

$$L(\alpha, \beta) = \left| \int_0^{\pi} e^{-i\beta\omega} C(\alpha, \omega) d\omega \right|^2 \qquad (8)$$

which can be interpreted as the Fourier transform of the coulumns of the complex cyclostationary gram. The joint likelihood for our example is shown in the right image of Figure 4. The peak is clearly evident at $(-8, 32)$ corresponding to the corect values of our model.

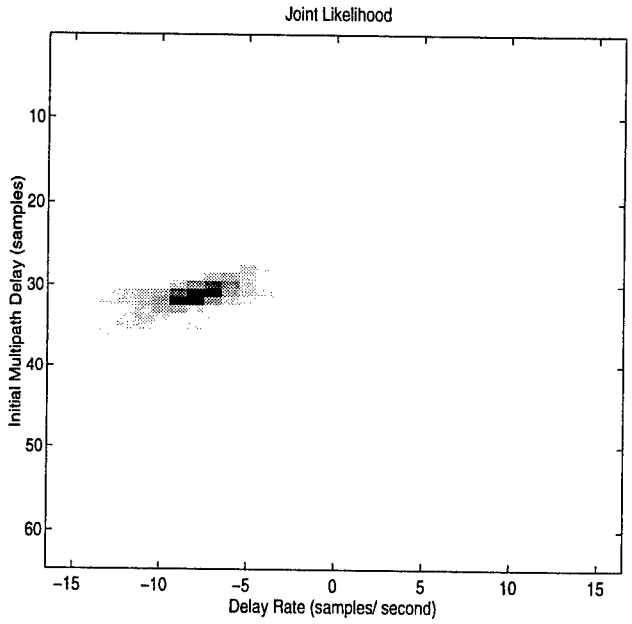


Figure 4. Joint two-parameter likelihood function for α and β.

A coherent marginal likelihood function for the delay rate can be found by integrating out the β parameter and is shown in Figure 5. The coherent detector is now able to distinguish between positive and negative values of the delay rate.

4. Conclusions and Future Directions

Coherent and noncoherent methods based on a generalized notion of cyclostationarity can be constructed to detect broadband multipath signals with affine time delay. A result demonstrated in this paper (and widely known in radar and other similar applications) is that while noncoherent methods cannot distinguish between the positive and nega-

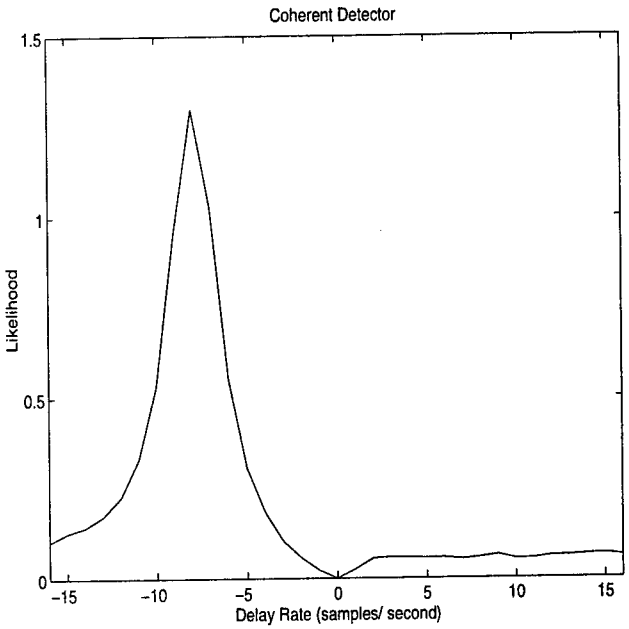


Figure 5. Coherent detector obtained by integrating joint likelihood function over β.

tive delay rates, coherent methods determine both the sign in addition to an estimate of the initial multipath delay.

Several possible research directions for this model and these detection methods are apparent. The affine time delay model can be generalized by letting τ be an arbitrary slowly varying function (i.e., small τ'). The use of alternative time-frequency distributions, especially ones matched to the hyperbolic patterns (see [4] and [1]) are worth investigating. The detection performance of these methods also need to be assessed for varying parameters(.e.g., λ) and multiple signals with possibly more than two paths.

References

[1] R. G. Baraniuk and D. L. Jones. Unitary equivalence: A new twist on signal processing. *to appear in the IEEE Trans. Signal Processing*.

[2] S. Chiu. Statistical Estimation of the Parameters of a Moving Source from Array Data. *Annals of Statistics*, 14(7):559-578,1986.

[3] D. E. Lake and G. Pei. Estimation of Very Low Frequency (VLF) Energy Distributions. MITRE Technical Report 92W0000218, October,1992.

[4] A. Papandreaou, F. Hlawatsch, and G. F. Boudreaux-Bartels. The hyperbolic class of time-frequency representations part I: Constant-Q warping, the hyperbolic paradigm, properties, and members. *IEEE Trans. Signal Processing*, 41:3425-3444, December 1993.

DIFFERENTIATING RANDOM AMPLITUDE HARMONICS FROM CONSTANT AMPLITUDE HARMONICS

Guotong Zhou
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250

ABSTRACT

Periodogram is a useful tool to reveal hidden periodicities in a given time series but the resulting spectral lines have often been associated with constant amplitude harmonics. Possibilities exist where the harmonics actually have non-zero mean random (as opposed to constant) amplitudes because the two can have identical periodograms. Applications exist to support the random amplitude models. Cyclic statistics are employed here as effective tools to distinguish constant from random amplitude harmonic models. The algorithms are FFT based and are easy to implement as illustrated by numerical examples.

1. INTRODUCTION

Detection of hidden periodicities embedded in a random process has been a concern over 100 years. Schuster in 1894 devised the periodogram as a means of searching for hidden periodicities. It has had much success in many areas ranging from seasonal and economic time series, seismology, geophysics, spectroscopy, and communications to sonar and radar signal processing (see e.g., [2], [1], [5] and references therein). In this paper, we consider a single harmonic $e^{j(\omega_0 t+\phi_0)}$ which may be "hidden" in a discrete-time series $\{x(t)\}_{t=0}^{T-1}$. The periodogram of $x(t)$ is defined as

$$I_{2x}(\alpha) \triangleq \frac{|X_T(\alpha)|^2}{T}, \tag{1}$$

$$X_T(\alpha) \triangleq \sum_{t=0}^{T-1} x(t)\, e^{-j\alpha t}, \tag{2}$$

where (2) is simply the DFT of the data $\{x(t)\}_{t=0}^{T-1}$.

If the periodogram shows a peak at ω_0, one tends to believe that $x(t)$ is of the form

$$x(t) = A\, e^{j(\omega_0 t+\phi_0)} + v(t), \tag{3}$$

where A, ω_0, ϕ_0 are deterministic constants and $v(t)$ is stationary additive noise. On the other hand, if (1) does not show any peak, one is tempted to say that $x(t)$ is stationary.

These are the pitfalls that researchers are easily subject to, and they are the interest of this paper. Our purpose here is to clarify that **(i)** when the periodogram exhibits a peak at ω_0, $x(t)$ can also be of the form

$$x(t) = s(t)\, e^{j(\omega_0 t+\phi_0)} + v(t), \tag{4}$$

where $s(t)$ is an ergodic random process with mean $m_s \triangleq E[s(t)] \neq 0$ and is assumed to be uncorrelated with $v(t)$. We refer to (3) as the constant amplitude harmonic model, and (4) as the random amplitude harmonic model. Alternatively, we will also call (4) a harmonic in multiplicative and additive noise. Note that (3) can be regarded as a special case of (4) with $s(t) = A$.

Another point that we wish to clarify is **(ii)** when the periodogram does not show any peak, it is still possible for $x(t)$ to obey (4) but with $m_s = 0$. The goal of this paper is to provide tools that can distinguish stationary processes such as $v(t)$, constant amplitude harmonics (3), and random amplitude harmonics (4), using cyclic statistics.

Random amplitude harmonics show up in a variety of applications. In radar processing, when a non-point target is fast maneuvering or scintillating, the resulting harmonic (due to Doppler shift) carries a random amplitude [6]. In underwater acoustics, when the medium (the ocean) is dispersive or fluctuating, the sonar return also experiences some random amplitude effect [3]. The model in (4) is also appropriate for Doppler weather radar/lidar returns, where $s(t)$ is due to the randomness in the scatterers (hydrometeors or areasol particles). Due to carrier modulation, (4) is suitable for communications signals as well.

We wish to point out that it is important to identify the correct model at least for the following reasons: 1) Whether the harmonic has random or constant amplitude reveals partial information about the source (target) such as scattering or fading; 2) The Cramér-Rao bounds on the parameter estimates are different for the two models [7]; 3) The corresponding maximum likelihood (ML) estimates are also different. For example, in the nonzero mean ($m_s \neq 0$) case and when $v(t)$ is zero-mean white Gaussian, estimates $\hat{m}_s$ (or $\hat{A}$), $\hat{\omega}_0$, $\hat{\phi}_0$ obtained by minimizing the mean square error between $x(t)$ and $m_s e^{j(\omega_0 t+\phi_0)}$ are ML when $s(t) = A$ but are not when $s(t)$ is random [8]. Therefore by assuming the wrong model, one may intend to obtain ML estimates but cannot.

We will show that (3) and (4) can have identical periodograms when $m_s \neq 0$. However, when a peak is detected in the periodogram, one easily tends to believe that the true model is (3), and the possibility of (4) being present is usually overlooked.

The rest of the paper is organized as follows: in Section 2 we examine the cyclic statistics of (3) and (4) and devise algorithms to distinguish the two. Some practical aspects of the algorithms are discussed in Section 3. We use simulated data to illustrate the procedures in Section 4 and draw conclusions in Section 5.

2. RETRIEVAL USING CYCLIC STATISTICS

The processes in (3) and (4) are called wide-sense cyclostationary because their mean or variance are periodic functions of time. The mean of (3) is given by

$$m_{1x}(t) \triangleq E[x(t)] = A\, e^{j(\omega_0 t+\phi_0)} + m_v, \tag{5}$$

0-8186-7576-4/96 $5.00

whereas the mean of (4) is

$$m_{1x}(t) = m_s\, e^{j(\omega_0 t+\phi_0)} + m_v, \qquad (6)$$

with $m_v \triangleq E[v(t)]$. If $\omega_0 \neq 0 \bmod (2\pi)$, we realize from both (5) and (6) that $\lim_{T\to\infty} T^{-1} \sum_{t=0}^{T-1} m_{1x}(t) = m_v$, hence one can always remove the time average of $x(t)$ to equivalently remove the mean of $v(t)$. W.l.o.g, we henceforth assume that $m_v = 0$ and rewrite

$$m_{1x}(t) = A\, e^{j(\omega_0 t+\phi_0)} \quad \text{for (3)}, \qquad (7)$$

$$m_{1x}(t) = m_s\, e^{j(\omega_0 t+\phi_0)} \quad \text{for (4)}. \qquad (8)$$

Since (7) and (8) are periodic functions of t, we consider their Fourier Series (FS) coefficients, which we call the cyclic mean [4],

$$M_{1x}(\alpha) \triangleq \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} m_{1x}(t)\, e^{-j\alpha t}. \qquad (9)$$

For (7) and (8), they are given by

$$M_{1x}(\alpha) = A\, e^{j\phi_0}\, \delta(\alpha-\omega_0), \qquad (10)$$

$$M_{1x}(\alpha) = m_s\, e^{j\phi_0}\, \delta(\alpha-\omega_0), \qquad (11)$$

respectively, where $\delta(\cdot)$ is the Kronecker delta function. Consistent sample estimate of $M_{1x}(\alpha)$ is given by [4]

$$\hat{M}_{1x}(\alpha) \triangleq \frac{1}{T} \sum_{t=0}^{T-1} x(t)\, e^{-j\alpha t}, \qquad (12)$$

which is simply the normalized DFT of the data.

From (10) and (11), we see that if $m_s \neq 0$, a plot of $|\hat{M}_{1x}(\alpha)|$ will show a peak at $\alpha = \omega_0$ for both (3) and (4), the location of which provides an estimate of ω_0. The phase at the peak, $\arg[\hat{M}_{1x}(\hat{\omega}_0)]$, gives $\hat{\phi}_0$, and the peak strength $|\hat{M}_{1x}(\hat{\omega}_0)|$ yields an estimate for A or m_s. We proved in [8] that these estimates are consistent with the following variance rates: $\mathrm{var}(\hat{\omega}_0) = O(T^{-3})$, $\mathrm{var}(\hat{m}_s) = O(T^{-1})$, $\mathrm{var}(\hat{A}) = O(T^{-1})$, and $\mathrm{var}(\hat{\phi}_0) = O(T^{-1})$.

It is easy to see that the periodogram in (1) is related to (12) as follows: $I_{2x}(\alpha) = T\, |\hat{M}_{1x}(\alpha)|^2$. Hence, a peak in $|\hat{M}_{1x}(\alpha)|$ is equivalent to a peak in $I_{2x}(\alpha)$ at the same location. In this sense, cyclic mean and periodogram are equivalent. However, the latter does not contain phase information.

Because when $m_s \neq 0$, cyclic mean peaks at ω_0 for both (3) and (4), the two models cannot be distinguished. However, their respective variance tells the difference: $\sigma_x^2(t) = \sigma_v^2$, for (3), and $\sigma_x^2(t) = \sigma_s^2 \exp\{2j(\omega_0 t+\phi_0)\} + \sigma_v^2$, for (4), where σ_v^2 denotes the variance of $v(t)$ and similarly for σ_s^2. We term the FS coefficient of $\sigma_x^2(t)$ as the cyclic variance of $x(t)$ and it is given by

$$C_{2x}(\alpha) \triangleq \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \sigma_x^2(t)\, e^{-j\alpha t} \qquad (13)$$

which equals $\sigma_v^2\, \delta(\alpha)$ for (3), and $\sigma_s^2\, e^{2j\phi_0}\, \delta(\alpha - 2\omega_0) + \sigma_v^2\, \delta(\alpha)$ for (4). It is this quantity that reveals the difference between the two models (3) and (4): a peak at $\alpha \neq 0$ hints towards the random amplitude model (4).

Note that we can also use the cyclic covariance of $x(t)$, defined as the FS coefficient of $\mathrm{cov}\{x(t),\, x(t+\tau)\}$ w.r.t. t, at lag $\tau \neq 0$, to achieve similar results. But cyclic variance is slightly easier to implement.

Under (4) and for (13), the peak at $\alpha = 2\omega_0$ relies on $\sigma_s^2 \neq 0$ to be visible. We always have $\sigma_s^2 > 0$ when $s(t)$ is a real random process. However, when $s(t)$ is complex, $\sigma_s^2 = 0$ may happen – this is the case for QAM processes for example. The fourth-order cyclic statistic proposed in [4] resolves the problem.

Sample estimate of (13) is given by

$$\hat{C}_{2x}(\alpha) \triangleq \frac{1}{T} \sum_{t=0}^{T-1} \left[x(t) - \hat{m}_s\, e^{j(\hat{\omega}_0 t+\hat{\phi}_0)}\right]^2 e^{-j\alpha t} \qquad (14)$$

which is the normalized DFT of the square of the mean-compensated process. For constant amplitude harmonics, we simply replace the above formula by $m_s = A$.

Therefore we obtain the first result: If the cyclic mean of $x(t)$, or the periodogram of $x(t)$, shows a peak at ω_0, we need to further compute the cyclic variance of $x(t)$, or the periodogram of $\left[x(t) - m_s e^{j(\omega_0 t+\phi_0)}\right]^2$ in order to distinguish (3) and (4). If the resulting quantity shows a peak at $2\omega_0$, then the model is (4); otherwise, (3) is in force.

Now let us see what happens if $M_{1x}(\alpha)$ does not show any peak at all. This implies that the possibility for the constant amplitude harmonic model (3) is ruled out. Our task here becomes deciding whether $x(t)$ is a purely stationary process or model (4) with $m_s = 0$.

To resolve this problem, we again compute the cyclic variance. The cyclic variance of a stationary process shows a single peak at $\alpha = 0$, whereas (4) shows an additional peak at $\alpha = 2\omega_0$. The following observation is made: If the cyclic mean of $x(t)$, or the periodogram of $x(t)$, shows no peak, then we rule out the possibility of (3). We then compute the cyclic variance of $x(t)$, or the periodogram of $x^2(t)$. If the latter shows a peak at $\alpha \neq 0$, then $x(t)$ is due to (4).

We want to point out however, that it is possible to design a rigorous statistical test to decide on the peaks, and this constitutes an interesting future research direction.

When there are multiple harmonics present and we wish to decide between the models $x(t) = \sum_{l=1}^{L} A_l\, e^{j(\omega_l t+\phi_l)} + v(t)$ and $x(t) = \sum_{l=1}^{L} s_l(t)\, e^{j(\omega_l t+\phi_l)} + v(t)$, where A_l, ω_l, ϕ_l are deterministic constants and $\{s_l(t)\}_{l=1}^{L} \cup \{v(t)\}$ are mutually uncorrelated, the cyclic mean exhibits peaks at $\{\omega_l\}_{l=1}^{L}$ for both models if $E[s_l(t)] \neq 0$. However, the cyclic variance only peaks at $\alpha = 0$ for the constant amplitude model but shows peaks at $\{0\} \cup \{2\omega_l\}_{l=1}^{L}$ for the random amplitude model instead. Hence the cyclic algorithms of this section still apply for multicomponent harmonics.

3. PRACTICAL CONSIDERATIONS

When $x(t)$ is zero-mean, the cyclic variance can be easily computed as the normalized DFT of the data squared. However, the nonzero-mean case is more interesting because this is where one could be confused between random and constant amplitude harmonic models. It is also more cumbersome since one has to first estimate the time varying mean $m_{1x}(t)$, remove it, and then compute the cyclic variance. Estimates $\hat{m}_s$ (or $\hat{A}$), $\hat{\omega}_0$ and $\hat{\phi}_0$ can all be computed from $\hat{M}_{1x}(\alpha)$ to form $\hat{m}_{1x}(t)$. But if the data length is short, these estimates may not be very accurate. As a result, spurious peaks may occur in the cyclic variance of (3) at $\alpha \neq 0$ due to the "residue" harmonic not completely removed. This may hamper the performance of the detection scheme based on the cyclic variance.

The following alternative can be considered which avoids the aforementioned problem and requires somewhat less computation. We first compute the sample cyclic mean and if we detect a peak at $\alpha \neq 0$, we record that peak strength and denote it as $\hat{m}_s$. Next, instead of the variance, we consider the mean square of $x(t)$,

$$m_{2x}(t) \triangleq E[x^2(t)] = m_{2s}\, e^{2j(\omega_0 t+\phi_0)} + \sigma_v^2, \qquad (15)$$

where $m_{2s} \triangleq E[s^2(t)] = \sigma_s^2 + m_s^2$. When $s(t) = A$, we simply replace m_{2s} by A^2.

Since (15) is a periodic function of t, we consider its FS coefficient which we term the cyclic mean square,

$$M_{2x}(\alpha) \triangleq \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} x^2(t)\, e^{-j\alpha t}, \qquad (16)$$

$$= m_{2s}\, e^{j2\phi_0}\, \delta(\alpha - 2\omega_0) + \sigma_v^2\, \delta(\alpha), \qquad (17)$$

whose consistent sample estimate is given by

$$\hat{M}_{2x}(\alpha) \triangleq \frac{1}{T} \sum_{t=0}^{T-1} x^2(t)\, e^{-j\alpha t}. \qquad (18)$$

Now for both (3) and (4) with $m_s \neq 0$, (18) will show peaks at $\alpha = 0$ and $\alpha \neq 0$. Denote the peak strength at $\alpha \neq 0$ as $\hat{m}_{2s}$ and compute $\hat{\sigma}_s^2 \triangleq \hat{m}_{2s} - \hat{m}_s^2$. If $\hat{\sigma}_s^2$ is close to zero, we decide that (3) is more appropriate; otherwise, we choose (4). The rationale is that $\sigma_s^2 = m_{2s} - m_s^2$ is nonzero for $s(t)$ random, and $\sigma_s^2 = 0$ for $s(t) = A$. Of course, an interesting research problem here is to develop a statistical test in order to decide on the zeroness of the random variable $\hat{\sigma}_s^2$.

We wish to point out that the value of σ_s^2, obtained either from $m_{2s} - m_s^2$ using the cyclic mean and mean square, or from the peak strength of the cyclic variance at $\alpha = 2\omega_0$, may be used as a measure of dispersion or fading in Doppler radar or sonar applications.

The advantage of the cyclic mean square approach is that one can avoid estimating ω_0 and ϕ_0. However, the difference between the models (3) and (4) is revealed more numerically than graphically.

4. SIMULATIONS

We illustrate the algorithms proposed in this paper using simulated data. The following specifications apply to all examples: $T = 512$, $\omega_0 = 1$, $\phi_0 = 0.6$. In addition, additive noise $v(t)$ is a zero-mean uniformly distributed process with variance $\sigma_v^2 = 0.5$.

Example 1: Consider $x_1(t)$ which is given by (4) with i.i.d. Gaussian $s(t)$ having $m_s = 1.2$ and $\sigma_s^2 = 0.4$, and $x_2(t)$ which is given by (3) with $A = 1.2$. The real parts of $x_1(t)$ and $x_2(t)$ are shown in Figs. 1a and 1b, and the sample cyclic means are shown in Figs. 1c and 1d, respectively. It is difficult to classify $x_1(t)$ and $x_2(t)$ into (3) or (4) using the figures obtained so far. From $\hat{C}_{1x_1}(\alpha)$ and $\hat{C}_{1x_2}(\alpha)$ we obtained $\hat{m}_s = 1.1902$, $\hat{\omega}_0 = 0.9999$ for both $x_1(t)$ and $x_2(t)$, $\hat{\phi}_0 = 0.6371$ for $x_1(t)$ and $\hat{\phi}_0 = 0.6367$ for $x_2(t)$. We then subtracted the respective $\hat{m}_{1x}(t) = \hat{m}_s\, e^{j(\hat{\omega}_0 t+\hat{\phi}_0)}$ from $x_1(t)$ and $x_2(t)$, squared the resulting quantities, and took their normalized DFT, the magnitudes of which are plotted in Figs. 1e and 1f. A distinct extra peak occurred at $\alpha = 2$ in Fig. 1e, and we therefore subscribe $x_1(t)$ to (4) and $x_2(t)$ to (3).

Alternatively, we can bypass the estimation of ω_0 and ϕ_0 by adopting the approach in Section 3. From Figs. 1c and 1d, we estimated the peak strength at $\alpha \neq 0$ to be $\hat{m}_s = 1.1902$. We then computed the sample cyclic mean square $\hat{M}_{2x}(\alpha)$, the magnitudes of which are shown in Figs. 1g and 1h for $x_1(t)$ and $x_2(t)$ respectively. The peak strength at the nonzero cycle yielded $\hat{m}_{2s} = 1.7798$ for $x_1(t)$ and $\hat{m}_{2s} = 1.4234$ for $x_2(t)$, from which we inferred $\hat{\sigma}_s^2 = 0.3632$ for $x_1(t)$ and $\hat{\sigma}_s^2 = 0.0068$ for $x_2(t)$. Since the latter can be regarded as statistically zero, $x_2(t)$ is attributed to (3) and $x_1(t)$ to (4).

Example 2: We consider here the case with $m_s = 0$. All other parameters remain the same as in Example 1. The real parts of the time series are plotted in Figs. 2a and 2b. The sample cyclic mean does not show a dominant peak for $x_1(t)$ (Fig. 2c) but does so for $x_2(t)$ (Fig. 2d). The sample cyclic variance magnitudes are plotted in Figs. 2e and 2f for $x_1(t)$ and $x_2(t)$ respectively. The extra peak at $\alpha \neq 0$ in Fig. 2e distinguishes $x_1(t)$ from $x_2(t)$. The sample cyclic mean square magnitudes are plotted in Figs. 2g and 2h, from which $\hat{\sigma}_s^2 = 0.4211$ and $\hat{\sigma}_s^2 = 0.0086$ were estimated for $x_1(t)$ and $x_2(t)$ respectively. Since the latter is statistically zero, we decide that $x_2(t)$ came from (3).

5. CONCLUSIONS

Our focus here has been on random and constant amplitude harmonics. Traditionally, one examines the periodogram and based on the presence or absence of a peak, decides whether the process contains a constant amplitude harmonic or is purely stationary. We argue here that for both cases, a random amplitude harmonic could be present, either with non-zero mean or with zero-mean random amplitude. By employing the cyclic variance or cyclic mean square, one can easily tell the difference between the following pairs: random (with $m_s \neq 0$) vs. constant amplitude harmonics, and random amplitude harmonics (with $m_s = 0$) vs. a purely stationary process. Simulation studies corroborate these findings. Rigorous statistical tests lie ahead as further studies and the results can be easily extended to multicomponent processes.

REFERENCES

[1] O. Besson and P. Stoica, "Sinusoidal signals with random amplitude: least squares estimators and their statistical analysis", *IEEE Trans. on Signal Processing*, vol. 43, pp. 2733-2744, November 1995.

[2] D. R. Brillinger, "Fitting cosines: some procedures and some physical examples", in I. MacNeil and G. Umphrey (eds.), *Applied Probability, Stochastic Processes, and Sampling Theory*, pp. 75-100, D. Reidel Publ. Co., 1987.

[3] R. F. Dwyer, "Fourth-order spectra of Gaussian amplitude modulated sinusoids," *Journal of the Acoust. Soc. of America*, vol. 90, pp. 918-926, August 1991.

[4] G. B. Giannakis and G. Zhou, "Parameter estimation of cyclostationary amplitude modulated time series with application to missing observations," *IEEE Trans. on Signal Processing*, vol. 42, pp. 2408-2419, September 1994.

[5] A. Swami, "Multiplicative noise models: parameter estimation using cumulants," *Signal Processing*, vol. 36, pp. 355-373, 1994.

[6] H. L. Van Trees, *Detection, Estimation and Modulation Theory: Part III, Radar-Sonar Signal Processing and Gaussian Signals in Noise*, Ch. 11, New York: Wiley, 1971.

[7] G. Zhou and G. B. Giannakis, "Harmonics in Gaussian multiplicative and additive noise: Cramér-Rao bounds," *IEEE Trans. on Signal Processing*, vol. 43, pp. 1217-1231, 1995.

[8] G. Zhou and G. B. Giannakis, "Harmonics in multiplicative and additive noise: performance analysis of cyclic estimators," *IEEE Trans. on Signal Processing*, vol. 43, pp. 1445-1460, June 1995.

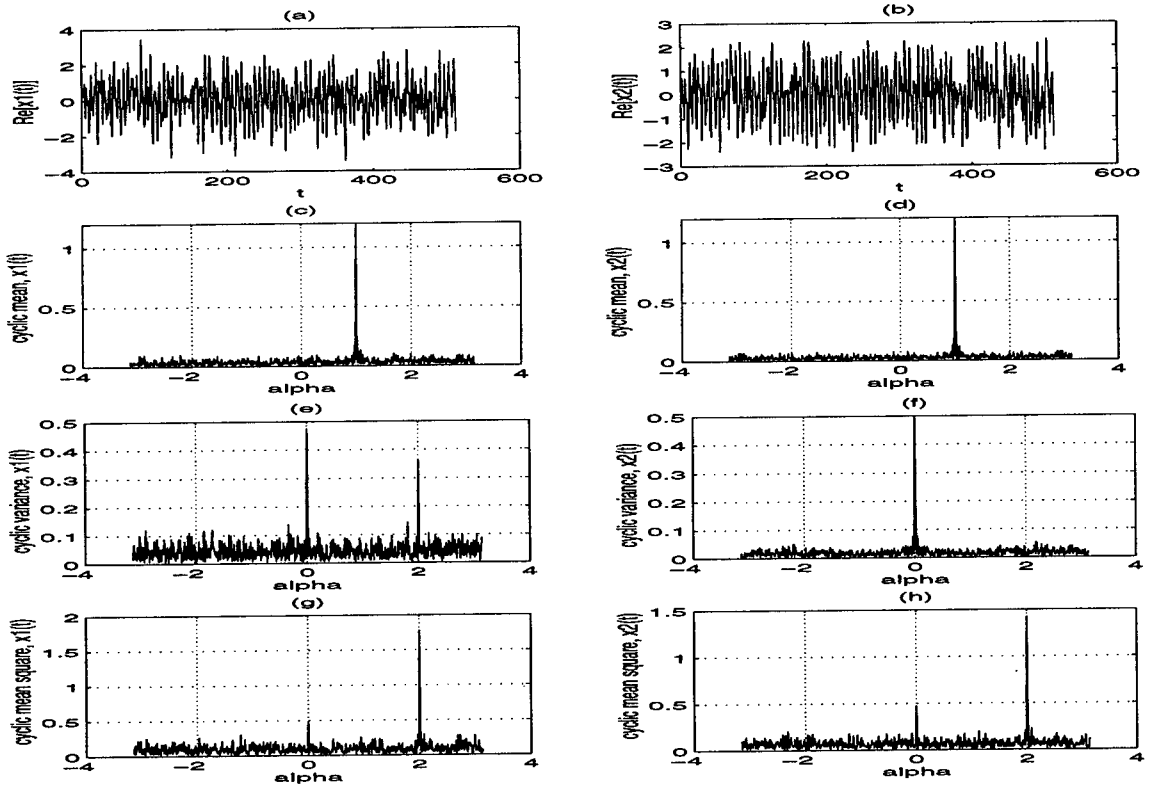


Figure 1. A non-zero mean random amplitude harmonic and a constant amplitude harmonic

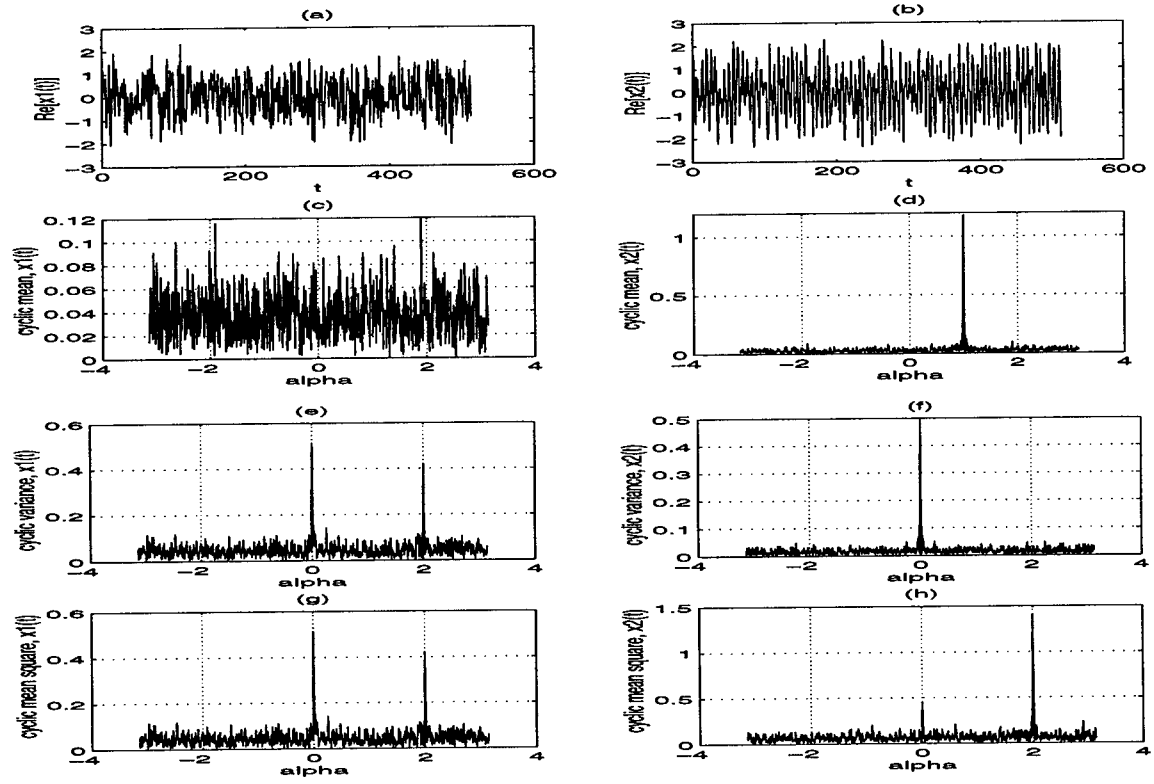


Figure 2. A zero-mean random amplitude harmonic and a constant amplitude harmonic

Classification of Non-Stationary Random Signals using Multiple Hypotheses Testing

Geoff Roberts and Boualem Boashash
Signal Processing Research Centre QUT
GPO Box 2434 Brisbane Q. 4001 Australia
ga.roberts@qut.edu.au

Abstract

In this paper we introduce a new time-frequency based method for classifying non-stationary random signals. The method involves dividing the signal into overlapping or non-overlapping segments considered to be subpopulations of the entire population. From each sub-population we calculate a test statistic which can be used to construct a single hypothesis test. To control the global type-I error it is necessary to consider the hypotheses from all subpopulations simultaneously. We use the generalised sequentially rejective Bonferroni *multiple hypothesis test which provides an efficient method to simultaneously test multiple hypotheses while maintaining the global type-I error. Finally, we show the results of classifying time-dependent AR(1) processes which have identical expected instantaneous power and power spectral densities but different time-frequency representations.*

1. Introduction

The problem of signal classification can be divided into three consecutive sub-problems: detection of the presence of a signal; segmentation to determine the time interval of the signal; and classification of the signal into one of a finite number of classes. In this work we will focus on classifying an observation signal into one of two classes, i.e., we assume that the signal is present and its time interval is known.

The original contribution of this paper involves the extension of a frequency domain classifier for stationary signals [7] to a time-frequency classifier for non-stationary signals. The motivation for this extension is straightforward: the classical technique is only optimal (in the sense of minimising the probability of misclassifying an observation of one kind for a fixed misclassification rate of the other kind) if the signal is stationary. This leads us to consider a technique that does not require the signal to be stationary. In particular, we introduce a time-varying quadratic discriminant function using the spectrogram. We apply the *generalised sequentially rejective Bonferroni* test to the multiple hypotheses that can be constructed at different points in time from this discriminant function.

Other classification techniques have been suggested recently using time-frequency distributions (TFD). In [8] the authors extended the log-spectral distance to the time-frequency case and in [3] the authors proposed a technique based on the cross Wigner-Ville distribution. In the sequel we will discuss how our method deviates from the existing solutions.

2. Time-frequency discrimination

In this paper we developed the theory for the simplest case of classifying the signal into one of two classes. It is straightforward to extend the results for a larger number of classes. To classify a signal into one of two classes we formulate the test

$$H: \ \mathbf{X} = \mathbf{S}_1 + \mathbf{U}$$
$$K: \ \mathbf{X} = \mathbf{S}_2 + \mathbf{U}$$

where $\mathbf{S}_1$ and $\mathbf{S}_2$ are zero mean non-stationary Gaussian signals and $\mathbf{U}$ is zero mean white Gaussian noise. A discrete time-frequency distribution of a random vector $\mathbf{X} = [X_1, \ldots, X_N]'$, is defined as [2]

$$S_{\mathbf{X}}(n,k) = \sum_{m=-(N-1)/2}^{(N-1)/2} R_{XX}(n,m) e^{-j2\pi mk/N} \quad (1)$$

for $n \in [0, N-1]$, where $R_{XX}(n,m)$ is the time-dependent covariance of the signal. In this case we choose $R_{XX}(n,m)$ such that the resulting TFD is the spectrogram, however, in general, this discriminant function is applicable for any TFD. For the case of classifying a signal into one of two classes,

0-8186-7576-4/96 $5.00

we define the time-dependent discriminant:

$$d(\mathbf{x},n) = \sum_{k=0}^{N-1} S_{\mathbf{x}}(n,k)\left(S_2^{-1}(n,k) - S_1^{-1}(n,k)\right) \quad (2)$$

where: $S_{\mathbf{x}}(n,k)$ is an estimate of the TFD from $\mathbf{x} = [x_1, x_2, \ldots, x_N]'$, a realisation of $\mathbf{X}$; $S_q(n,k), q = 1,2$, are estimates of the TFDs representing the two different classes and are assumed to be non-zero; and $k = 0, \ldots, N-1$, is discrete frequency (assuming $\mathbf{x}$ is analytic). This time-dependent discriminant function can be interpreted as an extension of the power spectrum quadratic discriminant function defined for stationary random processes in [7].

In general, existing classification algorithms form a single test statistic from a discriminant function and this is used to perform a single hypothesis test to determine if the observation belongs to class 1 or class 2. In our case the discriminant function given by Eq (2) returns a value at each time. Each value is used to construct a hypothesis, which are then combined and treated simultaneously. This approach differs from previous time-frequency based methods [1, 3, 8] where the solutions all involve integration over time to form a single hypothesis which can lead, in practical situations, to misclassification.

Smoothing. If there are zero terms in the TFDs of the population they will dominate Eq (2). To reduce this problem, and to lower the variance of the estimates, the discriminant function can be evaluated using a smoothed TFD

$$\hat{S}_{\mathbf{x}}(n,k) = \sum_{m,l} W(n+m, k+l) S_{\mathbf{x}}(n+m, k+l) \quad (3)$$

where $W(n,k)$ is an appropriate window [2].

3. Multiple hypotheses

Multiple comparison procedures provide a technique for simultaneously treating a collection of separate tests derived from sub-populations, while maintaining a global level of significance. If the level of significance for each individual test is set at α, then the global level of significance may be much higher [4]. In the time-frequency setting the signal is divided up into non-overlapping or overlapping segments. Each segment is a sub-population for which a test statistic can be derived and a hypothesis test can be constructed.

In the following section we discuss the generalised sequentially rejective Bonferroni test which controls the global level of significance.

3.1. Generalised sequentially rejective Bonferroni test (GSRBT)

The GSRBT was successfully applied to a signal processing problem in [9]. Eq (2) is defined for all time samples, however we are using the spectrogram, so we only evaluate $d(\mathbf{x},n)$ at the centre of the window. We use the statistic $D_i = d(\mathbf{X}, (2i-1)M/2)$ where M is the size of the sub-population or the spectrogram window length (no overlap). If the signal is from class q then $M^{-1}D_i$ is normal and estimates of the mean and variance are given by [7]

$$\begin{aligned}\hat{m}_{qi} &= \frac{1}{M}\sum_{k=0}^{M-1}(S_2^{-1}((i-1/2)M,k) \\ &- S_1^{-1}((i-1/2)M,k))S_{\mathbf{x}}((i-1/2)M,k) \quad (4)\end{aligned}$$

and

$$\begin{aligned}\hat{\sigma}^2_{qi} &= \frac{2}{M^2}\sum_{k=0}^{M-1}(S_2^{-1}((i-1/2)M,k) \\ &- S_1^{-1}((i-1/2)M),k)^2 S_{\mathbf{x}}((i-1/2)M,k)^2 \quad (5)\end{aligned}$$

If the smoothed TFD from Eq (3) is used then Eq's (4) and (5) will need to be adjusted according to the chosen window. Each local test can be constructed as testing $H_i : D_i \sim N(m_{1i}, \sigma^2_{1i})$ against the alternative $K_i : D_i \sim N(m_{2i}, \sigma^2_{2i})$, for $i = 1, \ldots, P$, and P is the number of test statistics. However, as previously discussed, we need to test all P hypotheses simultaneously. To do this we use the GSRBT as follows:

1. Calculate the p values, i.e., the probability that D_i exceeds its observed value under H_i. The p values are calculated as $\mathcal{P}_i = 1 - F((d_i - \hat{m}_{1i})/\hat{\sigma}_{1i})$, where we assume $F(y)$ is the normal cumulative distribution function since D_i is asymptotically normal.

2. It is possible to customise the p values to take into account *a priori* information pertinent to an application. This is achieved by using a set of positive real constants $c_1, \ldots, c_P$, which have values directly proportional to the importance of the individual hypothesis. The constants can be set to attain a more powerful test. If the constants are all equal then this procedure reduces to the sequentially rejective Bonferroni test which the GSRBT is a generalisation of [5]. The new p values are defined as $\mathcal{S}_i = \mathcal{P}_i/c_i$.

3. Order the p values in ascending order, $\mathcal{S}_{(1)} \leq \mathcal{S}_{(2)} \leq \ldots \leq \mathcal{S}_{(P)}$ and let $c_{(i)}$ and $H_{(i)}$ be the corresponding constants and hypotheses respectively. Also, let $\alpha_i = \alpha / \sum_{j=i}^{P} c_{(j)}$.

4. The GSRBT, depicted in Figure 1, is performed as follows: If $\mathcal{S}_{(1)} > \alpha_1$ then retain $H_{(1)}, \ldots, H_{(P)}$ and stop; otherwise, reject $H_{(1)}$ and test the next hypothesis. This procedure is repeated until either all the hypotheses are rejected or a set of hypotheses is retained.

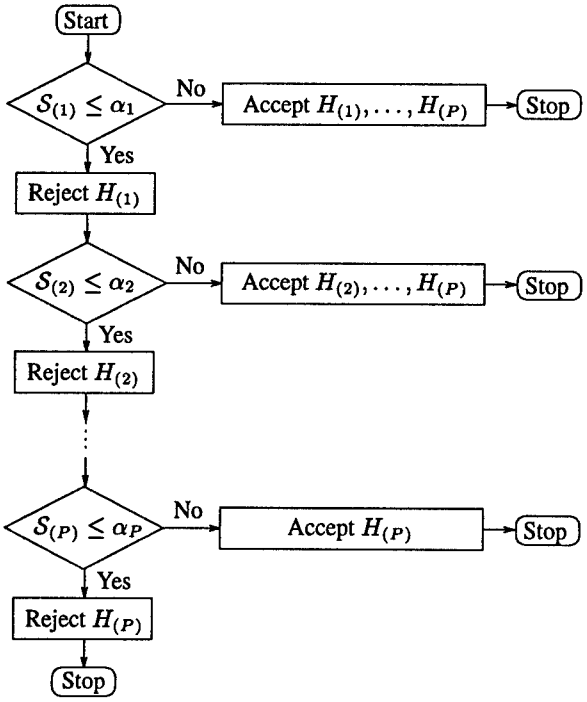


Figure 1. Generalised sequentially rejective Bonferroni test

5. Finally, a global decision is made based on the set $\mathcal{H} = \{H_{(i)}\}$ of retained hypotheses. This decision will depend on the application.

Now we will summarise the result given in [5] which proves the GSRBT maintains the the global significance level α. Consider a single hypothesis where $\Pr(\mathcal{P}_i > \alpha_i)$ under the null hypothesis is equal to $1 - \alpha_i$. It is this value α_i which gives us confidence in our test. Similarly the objective of a multiple test procedure is to maintain the global level of significance over all the hypotheses. Let I be the set of indices of true null hypotheses, then the equivalent expression for forming a confidence interval for the GSRBT is [5]

$$\Pr\left(S_i > \frac{\alpha}{\sum_{j \in I} c_j}, \quad \forall i \in I\right) \geq 1 - \alpha \qquad (6)$$

This equation is shown to be true in [5] and therefore the global level of significance is maintained.

4. Simulations

In this section we show results for the classification of two classes of first order time-varying autoregressive (TAR) signals. The classes are separable only in the time-frequency space. The TAR(1) process is defined as:

$$X_n = -a(n)X_{n-1} + U_n \qquad (7)$$

where U_n is zero mean Gaussian with time dependent variance, $\sigma_u(n)$. The AR parameter $a(n)$ gives a single pole rotating on the unit circle, i.e., $a(n) = -0.99e^{j2\pi f_1(n)}$ where

$$f_1(n) = \begin{cases} \frac{0.3}{N/2-1}n + 0.1 & 0 \leq n \leq N/2 - 1 \\ -\frac{0.3}{N/2-1}n + 0.7 & N/2 \leq n \leq N - 1 \end{cases} \qquad (8)$$

is the position of the pole on the unit circle for the first signal and

$$f_2(n) = \begin{cases} -\frac{0.3}{N/2-1}n + 0.4 & 0 \leq n \leq N/2 - 1 \\ \frac{0.3}{N/2-1}n - 0.3 & N/2 \leq n \leq N - 1 \end{cases} \qquad (9)$$

is the pole position for the second signal. These signals were chosen because they have the same expected power at each time instant and the same frequency content over $[0, N-1]$. The SNR for the following experiments is calculated using $\text{SNR} = 10\log_{10}\left(\sum_{n=0}^{N-1} |x_n|^2/(\sigma_{wr}^2 + \sigma_{wi}^2)\right)$, where σ_{wr}^2 and σ_{wi}^2 are the variances of the real and imaginary parts of additive white Gaussian noise.

To assess the performance of our method an Operating Characteristic (OC) curve was constructed for each class. We used 15 realisations from each class to estimate $S_1(n,k)$ and $S_2(n,k)$. The constants c_i were set equal for these experiments. The OCs are shown in Figure 2. The matched filter and frequency spectrum method [7] naturally do not perform well for this class of signals. The template used for the matched filtering was an ensemble average calculated with 15 realisations of signals from each class. Figure 3 compares the classification performance of the multiple hypotheses method against a non-parametric time-frequency method that discriminates between classes using the distance between the log of the signal TFD and the log of the class TFDs [8].

5. Discussion

There is a number of optimisations which can be included for a particular application. The window length and the overlap used to estimate the TFDs in Eq's (2) and (3) can be optimised to reflect the degree of non-stationarity in the classes. The GSRBT can also be customised to an application. As mentioned in Section 3.1 the weights, $c_1, \ldots, c_P$, can be used to increase the power of the test and, in addition the accepted hypothesis can be combined in any arbitrary way to make a global decision. For example if two or more hypotheses are mutually exclusive, this information will influence the global decision.

In Section 3.1 we assumed that the test statistics were normal, for narrowband signals this is not valid. The normal distribution is used to calculate the p values and therefore, is crucial to the performance of the test. In [7] it is shown that for narrowband signals the discriminant function in Eq (2) is

a summation of approximately chi-square random variables. This result can be used to improve the performance of the algorithm [6].

The disadvantages with this method are twofold: firstly we assume local stationarity to estimate the TFDs; and secondly, we assume that the signals are Gaussian. The method presented in [8], is non-parametric and so will be more appropriate if the Gaussian assumption is not valid.

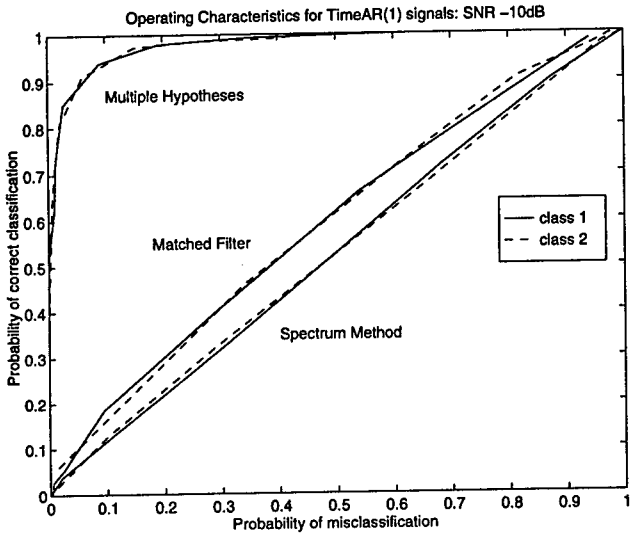


Figure 2. OC for classification of the two signals: SNR $= -10$dB. Comparison of matched filter, spectrum, log TFD, and multiple hypotheses methods.

6. Conclusion

We have presented a new method for the classification of non-stationary Gaussian signals by combining time-frequency analysis with multiple hypothesis testing. A time-frequency distribution is used to separate classes of signals that are inseparable in either the time or the frequency domain alone. The use of a multiple hypotheses test, the generalised sequentially rejective Bonferroni test (GSRBT), allowed the simultaneous treatment of the set of test statistics that arise from the time-dependent discriminant function. The GSRBT can be customised to a particular application to increase the power of the test.

The performance of this method was evaluated empirically by classifying two classes of zero mean non-stationary Gaussian signals. It performed favourably when compared to the classical methods and another non-parametric time-frequency method. This gain in performance is dependent on the Gaussian and local stationarity assumptions.

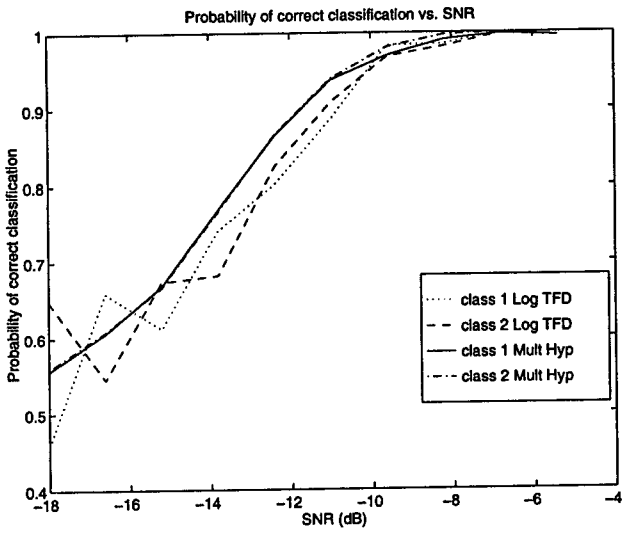


Figure 3. Probability of correct classification Vs. SNR. Comparison of log TFD and multiple hypotheses methods.

References

[1] S. Abeysekera and B. Boashash. Methods of signal classification using the images produced by the Wigner-Ville distributions. *Pattern Recognition Letters*, 12:717–729, Nov 1991.

[2] B. Boashash. Time-frequency signal analyisis. In S. Haykin, editor, *Advances in Spectrum Analysis*, chapter 9. Prentice Hall, New Jersey, 1991.

[3] B. Boashash and P. O'Shea. A methodology for detection and classification of some underwater acoustics signals using time-frequency analysis techniques. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(11):1829–1841, nov 1990.

[4] Y. Hochberg and A. Tamhane. *Multiple Comparison Procedures*, page 363. John Wiley & Sons, United States of America, 1987.

[5] S. Holm. A simple sequentially rejective multiple test procedure. *Scand J Statist*, 6:65–70, Nov 1979.

[6] G. Roberts, A. Zoubir, and B. Boashash. Time-frequency discriminant analysis for non-stationary Gaussian signals. In *International Synposium on Signal Processing and its Applications, ISSPA'96*, Gold Coast, Australia, 1996, Submitted.

[7] R. Shumway. Discriminant analysis for time series. In P. Krishnaiah and L. Kanal, editors, *Handbook of Statistics 2*, chapter 1. North-Holland, New York, USA, 1982.

[8] I. Vincent, C. Doncarli, and E. L. Carpentier. Non stationary signal classification using time-frequency distributions. In *Time-Frequency and Time-Scale Analysis*, pages 233–236, Philadelphia, USA, Oct 1994.

[9] A. Zoubir and J. Böhme. Bootstrap multiple tests applied to sensor location. *IEEE Transactions on Signal Processing*, 43(6):1386–1396, June 1995.

A Larger Class of Wavepacket Eigenfunction Systems Which Contains Cauchy Wavelets and Coherent States

Fuminori SAKAGUCHI
Faculty of Engineering, Fukui University, Fukui-shi 910 Japan
e-mail:saka@digcom.fuee.fukui-u.ac.jp

Abstract *A larger class of the systems of almost localized wavepackets is proposed. This class contains the Cauchy wavalet system and the coherent state system as special cases. The wavepackets systems in this class are the eigenfunction systems of a kind of linear operators, and they are invariant under the time-shift. Moreover, we show that they are 'pseudo-orthogonal' over-complete systems in L^2. An application of them to the diagonalization problem of the auto-correlation functions of stochastic processes via over-complete wavepacket systems are also proposed.*

1 Introduction

Time scale analysis and the time-frequency analysis are closely related to two types of wavepacket systems, the system of wavelets[1-6] with continuous parameters and the other is the system of coherent states[2,7-9], respectively. Both are over-complete systems of almost localized wavepackets. It has been known that these systems belong to 'generalized coherent states'[10]. It has been shown that the two systems, with special wavepackets, can be represented as the eigenfunction systems of two respective linear operators [12,11,14]. Both eigenfunction systems have 'time-shift invariant' property what may be called, where the shift of a parameter causus only the shift of the time with the shape of the wavepacket unchanged. In this paper, we will extend this research into a larger class of the eigenfunction systems which have these properties.

2 Wavepacket Systems and Quantum Mechanics

We will begin by summarizing the mathematical relations between these two kinds of wavepacket systems and the quantum mechanics. For $h(t) \in L^2(\mathbf{R})$ whose Fourier transform $H(\omega)$ satisfies

$$C_h \triangleq \int_{-\infty}^{\infty} \frac{|H(\omega)|^2}{\omega}\, d\omega \ < \infty, \tag{1}$$

define

$$h^{(a,b)}(t) \triangleq |a|^{-\frac{1}{2}} h(\tfrac{t-b}{a}) \ . \tag{2}$$

Then the set $\{h^{(a,b)}(t)|\ a \in \mathbf{R}, b \in \mathbf{R}\}$ is a 'pseudo-orthogonal' (NB: not 'orthogonal') over-complete wavelet system in $L^2(\mathbf{R})$, where for an arbitraly $f(t) \in L^2(\mathbf{R})$, the relation

$$\tfrac{1}{C_h} \int\int_{R^2} \tfrac{da\, db}{a^2} \langle h^{(a,b)}, f\rangle\ h^{(a,b)}(t) = f(t) \tag{3}$$

holds[2-5]. In physical context, the above wavelet system can be regarded as the system of the generalized coherent states[10] associated with the affine group[15,16], and the parallelism between the wavelets and the (usual) coherent states associated with Weyl-Heisenberg group has been pointed out [2,17]. We will summarize this parallelism. Let $g(t)$ be an element of $L^2(\mathbf{R})$ such that

$$C_g' \triangleq 2\pi \int_{-\infty}^{\infty} |g(t)|^2 dt < \infty. \tag{4}$$

Then, with the definition of the (usual) coherent state

$$g^{(q,p)}(t) \triangleq e^{ip(t-q)}\ g(t-q), \tag{5}$$

the set $\{g^{(q,p)}(t)|\ q \in \mathbf{R}, q \in \mathbf{R}\}$ is an over-complete system in $L^2(\mathbf{R})$, and, for an arbitrary $f(t) \in L^2(\mathbf{R})$, the 'pseudo-orthogonal' relation

$$\tfrac{1}{C_g'} \int\int_{R^2} dq\ dp\ \langle g^{(q,p)}, f\rangle\ g^{(q,p)}(t) = f(t) \tag{6}$$

holds, which is parallel to (3). As a special case, when

$$g(t) = g_0(t) \triangleq \pi^{-\frac{1}{4}} e^{-\frac{t^2}{2}}, \tag{7}$$

the above-defined coherent states are just corresponding to the wavefunctions of the (usual) coherent states in quantum mechanics [8,9] in the following sense; Let Q and P be the position-coordinate operator and the momentum operator, respectively which satisfy the commutation relation $[Q,P] = iI$ (I: Identity op.), and define the operator

$$a \triangleq \tfrac{1}{\sqrt{2}}(Q + iP). \tag{8}$$

0-8186-7576-4/96 $5.00

Then the above coherent states are the eigenfunctions of the operator a in the following sense:

$$g_0^{(q,p)}(t) = {}_Q\langle t|\alpha\rangle_a \tag{9}$$

$$a|\alpha\rangle_a = \alpha|\alpha\rangle_a \quad \text{with} \quad \alpha \triangleq \frac{1}{\sqrt{2}}(q+ip). \tag{10}$$

(NB: This operator a is corresponding to the operator $\mathcal{F}$ such that

$$(\mathcal{F}f)(t) = \frac{1}{\sqrt{2}}(tf(t) + \frac{df(t)}{dt})$$

in the expression used in signal processing.)

As is known well, this coherent state in quantum mechanics satisfies the relation

$$\frac{1}{\pi}\int_C d^2\alpha \ \ |\alpha\rangle_a \ {}_a\langle\alpha| = I \tag{11}$$

which is mathematlcally equivalant to a special case of the relation (7), and the shift of the (complex) eigenvalue α can be made by the unitary transformation

$$\begin{aligned} &e^{\gamma a^\dagger - \gamma^* a}|\alpha\rangle_a = e^{iIm(\gamma\alpha^*)}|\alpha+\gamma\rangle_a \\ &(where \quad \gamma \triangleq \frac{1}{\sqrt{2}}(q'+ip') \ ). \end{aligned} \tag{12}$$

On the other hand, in the wavelet case, there exists 'the analogue in wavelet version' of the above eigenvalue-shift relation in terms of the unitary representation of the affine group[16]. If, with an appropriate functions θ and λ, an operator A satisfies the ralation

$$e^{-iqP}e^{isB}|\alpha\rangle_A = e^{i\theta(\alpha;s,q)}|\lambda(\alpha;s,q)\rangle_A \tag{13}$$

(where $B = \frac{1}{2}\{Q,P\}$), then

$$e^{-isB}e^{iqP}Ae^{-iqP}e^{isB} = \lambda(A;s,q) \tag{14}$$

where $\lambda(A;s,q)$ denotes the operator which is obtained by substituting the operator A into the function $\lambda(\alpha;s,q)$ instead of α. A kind of the non-trivial operators which have this property are

$$A_k \triangleq Q - ikP^{-1} \qquad (k = 1,2,3,...), \tag{15}$$

and when $A = A_k$, $\lambda(\alpha;s,q) = e^{-s}(\alpha+q)$ [11]. The system of the eigenfunction of the operator A_k is corresopnding to the Cauchy wavelet system with

$$h(t) = h_k(t) \triangleq \frac{G_k}{(t-i)^{k+1}} \tag{16}$$

(where G_k is a constant). In other words,

$$h_k^{(a,b)}(t) = {}_Q\langle t|\alpha\rangle_{A_k} \tag{17}$$

$$\text{with} \quad A_k|\alpha\rangle_{A_k} = \alpha|\alpha\rangle_{A_k} \quad (\alpha = b+ai). \tag{18}$$

(NB: This operator A_k is corresponding to the operator $\mathcal{G}_k$ such that

$$(\mathcal{G}_k f)(t) = tf(t) + k\int_{-\infty}^{t} f(s)ds$$

in the expression used in signal processing.)

The Fourier transform of this eigenfunction is identical to the wavefunction of the affine coherent state proposed by Paul[12](See also [17,18]). Moreover, 'the analogue in wavelet version' of the relation (3) is obtained by 'translating the relation (3) into the quantum-mechanical language' with $a = e^{-s}$ and $b = e^{-s}q$, as

$$\frac{1}{C_k}\int_C \frac{d^2\alpha}{(Im\ \alpha)^2} \ \ |\alpha\rangle_{A_k} \ {}_{A_k}\langle\alpha| = I \tag{19}$$

The operator A_k satisfies the commutation relation

$$[A_k, A_k^\dagger] = 2kP^{-2} = -\frac{1}{2k}(A_k - A_k^\dagger)^2, \tag{20}$$

which is more complicated than $[a, a^\dagger] = I$.

3 A Larger Class of Eigenfunction Systems

The wavelet system and the coherent states system are the eigenfunction systems of the operators a and A_k. These two operators belong to the class

$$\{c(Q+iy(P))|\ c : \text{real},\ y(P) : \text{func.of } P\}. \tag{21}$$

With $c = \frac{1}{\sqrt{2}}$ *and* $y(P) = P$, the operator a related to the coherent states is obtained, while the operator A_k related to the wavelets is obtained with $c = 1$ *and* $y(P) = -kP^{-1}$. Here the constant c is not essential because the scalar product does not change the eigenfunction system. So we will investigate the eigenfunction system of the operators which belongs to

$$\{Q+iy(P)\ ;\ y(P) : \text{func.of } P\}, \tag{22}$$

without loss of generality. Define

$$A^{(y)} \triangleq Q + iy(P), \tag{23}$$

and denote the eigenfunction (in the position representation) of this operator by

$$\psi_\alpha^{(y)}(t) = {}_Q\langle t|\alpha\rangle_{A^{(y)}} \tag{24}$$

$$\text{with} \quad A^{(y)}|\alpha\rangle_{A^{(y)}} = \alpha|\alpha\rangle_{A^{(y)}}. \tag{25}$$

(NB: This operator $A^{(y)}$ is corresponding to the operator $\mathcal{G}^{(y)}$ such that

$$(\mathcal{G}^{(y)}f)(t) = tf(t) + i\Big(y(-i\tfrac{d}{dt})\ f\Big)(t)$$

in the expression used in signal processing.)

Then, formally, the eigenfunction $\psi_\alpha^{(y)}(t)$ is the solution of the equation

$$t\psi_\alpha^{(y)}(t) + iy(-i\tfrac{d}{dt})\psi_\alpha^{(y)}(t) = \alpha\psi_\alpha^{(y)}(t). \tag{26}$$

(Note that the inverse of the differential operator is the integral operator.) Hence, it is shown that the Fourier transform of $\psi_\alpha^{(y)}(t)$

$$\Psi_\alpha^{(y)}(p) \triangleq \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_\alpha^{(y)}(t) e^{-ipt} dt \quad (= {}_P\langle p|\alpha\rangle_{A(y)}) \tag{27}$$

satisfies the relation

$$i\tfrac{d}{dp}\Psi_\alpha^{(y)}(p) + iy(p)\Psi_\alpha^{(y)}(p) = \alpha\Psi_\alpha^{(y)}(p). \tag{28}$$

From this,

$$\tfrac{d}{dp}(log\ \Psi_\alpha^{(y)}(p)) = -y(P) - i\alpha. \tag{29}$$

Let $\{p_n; n = 0, 1, 2, 3..., M\}$ be the set of zeros of $\Psi_\alpha^{(y)}(p)$ (including $\pm\infty$ formally when $\lim_{p\to\pm\infty} \Psi_\alpha^{(y)}(p) = 0$). Then we obtain the solutions

$$\Psi_\alpha^{(y)}(p) = \Psi_{n,\alpha}^{(y)}(p) \tag{30}$$

$$\triangleq \begin{cases} C_{n,\alpha}^{(y)} e^{-Y(p)+ap} e^{-ibp} & (p_n < p < p_{n+1}) \\ 0 & (\text{otherwise}) \end{cases}$$

with

$$Y(p) \triangleq \int y(p)dp \tag{31}$$

$$b \triangleq Re\ \alpha, \qquad a \triangleq Im\ \alpha. \tag{32}$$

Here, $C_{n,\alpha}^{(y)}$ is determined so that the eigenfunction may be normalized as

$$\int_{p_n}^{p_{n+1}} |\Psi_{n,\alpha}^{(y)}(p)|^2 = 1. \tag{33}$$

Because the factor e^{-ibp} in (30) has no influence on this normalization condition, we can choose the constant $C_{n,\alpha}^{(y)}$ which does not depend on b but only on $g(\cdot)$ *and* a, as

$$C_{n,\alpha}^{(y)} = C_{n,ai}^{(y)} \tag{34}$$

Note that there is not always a normalized solution with support $[p_n, p_{n+1}]$ for any α, because the function may not square-integrable in some interval. However, as is easily shown, whether the normalized solution exists or not depends only on $y(\cdot)$ and the imaginary part a of the eigenvalue α. For example, in the wavelet case with $y(P) = -kP^{-1}$, the function $\Psi_{0,\alpha}^{(y)}(p)$ with support $(-\infty, 0]$ can be normalized for $a > 0$, while the function $\Psi_{1,\alpha}^{(y)}(p)$ with support $[0, \infty)$ can be normalized for $a < 0$. In the coherent state case with $y(P) = P$, the function $\Psi_{0,\alpha}^{(y)}(p)$ with support $(-\infty, \infty)$ can be always normalizad. It is easily shown that the function $\Psi_{n,\alpha}^{(y)}(p)$ $(1 < n < M-1)$ with a compact support $[p_n, p_{n+1}]$ can be normalized always unless $Y(P)$ contains any singularity on this support. Similarly, it is easily shown that there are a_- and a_+ (which may be $\pm\infty$) such that the function $\Psi_{0,\alpha}^{(y)}(p)$ with support $(-\infty, p_1]$ can be normalized for $a \geq a_+$ and the function $\Psi_{M,\alpha}^{(y)}(p)$ with support $[p_M, \infty)$ can be normalized for $a \leq a_-$ unless $Y(P)$ contains any singularity on these supports. When the normalization possible, then from (30),(33) and (34),

$$C_{n,\alpha}^{(y)} = e^{i\theta(a)} \Big(\int_{p_n}^{p_{n+1}} e^{-2Y(p)+2ap} dp \Big)^{-\frac{1}{2}} \tag{35}$$

with a real-valued function $\theta(a)$.

The solution (30) with the normalization coefficient (35) implies that its inverse Fourier tansform sataifies

$$\psi_{n,b+ai}^{(y)}(t) = \psi_{n,ai}^{(y)}(t-b). \tag{36}$$

This relation shows that the shift of the real part of the eigenvalue does not change the shape of the wavapacket but causes only the time shift. This property is very profitable for signal processing.

We can also show easily that the wavepacket $\psi_{n,ai}^{(y)}(t)$ has finite variance both in the time domain and the frequency domain, which implies the wavepackets are almost localized in both domain.

Another important property of these types of eigenfunction systems is the pseudo-orthogonality mentioned above when the inverse Fourier transform of the function $exp(2Y(s/2))$ or the inverse Laplace transform of the function $exp(a_+s + 2Y(-i\nu/2))$ or that of $exp(-a_-s + 2Y(-s/2))$ exists. Define them if exist, as

$$u(a) \triangleq \tfrac{1}{2\pi} \int_{-\infty}^{\infty} e^{2Y(-\frac{i\nu}{2})} e^{i\nu a} d\nu \tag{37}$$

$$u_+(a) \triangleq \tfrac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{a_+ s} e^{2Y(-\frac{s}{2})} e^{sa} ds \quad (c > a_+) \tag{38}$$

$$u_-(a) \triangleq \tfrac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{-a_- s} e^{2Y(\frac{s}{2})} e^{sa} ds \quad (c > -a_-). \tag{39}$$

Using these functons and the coefficients in (34),(35), define

$$w_n(a) \triangleq \tfrac{u(a)}{2\pi |C_{n,a}^{(y)}|^2} = \tfrac{1}{2\pi} u(a) \int_{p_n}^{p_{n+1}} e^{-2Y(p)+2ap} dp \tag{40}$$

$$w_n^+(a) \triangleq \tfrac{u_+(a-a_+)}{2\pi |C_{0,a}^{(y)}|^2} = \tfrac{1}{2\pi} u_+(a-a_+) \int_{-\infty}^{p_1} e^{-2Y(p)+2ap} dp \tag{41}$$

$$w_n^-(a) \triangleq \frac{u_-(a_- - a)}{2\pi |C_{M,a}^{(y)}|^2} = \frac{1}{2\pi} u_-(a_- - a) \int_{p_M}^{\infty} e^{-2Y(p)+2ap} dp. \quad (42)$$

Then we can show that the pseudo-orthogonal completeness relation

$$\sum_{n=0}^{M} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dadb \cdot v_n(a) \psi_{n,\alpha}^{(y)}(t) \psi_{n,\alpha}^{(y)*}(t') = \delta(t-t'). \quad (43)$$

holds where $v_n(a) = w_n(a)$ for $1 \leq n \leq M$, $v_0(a) = w_0^-(a)$ and $v_M(a) = w_M^+(a)$ (The outline of the proof is given in the latter paper of [14]).

4 An Application

The eigenfunction systems proposed above can be applied to the pseudo-diagonalizations of the autocorrelation functions of stochastic processes. In order that we regard a stochastic process as a superposition of uncorrelated random wavepackets, we must know how to diagonalize the auto-correlation function by the over-complete waveletpacket system. In the case where the over-complete wavepacket system is a wavelet system, a systematic method for this problem has been proposed[13,14]. This method is based on non-commutative operator algebra, and it utilizes the fact that the wavelet system is the eigenfunction system of the operator A_k. We can extend this method directly to the cases of more general wavepacket systems proposed above.

Let $f(\alpha, \alpha^*)$ be a function of a complex variable α which is expanded by the operator defined in (23) as

$$f(\alpha, \alpha^*) = \sum_{m,n} C_{m,n} \alpha^m \alpha^{*n}. \quad (44)$$

Then, define the operator in 'normal order' and the operator in 'antinormal order' (in extended version) by

$$\mathcal{N}^{(y)}\{f(\alpha, \alpha^*)\} \triangleq \sum C_{m,n} A^{(y)\dagger n} A^{(y)m} \quad (45)$$

and

$$\mathcal{A}^{(y)}\{f(\alpha, \alpha^*)\} \triangleq \sum C_{m,n} A^{(y)m} A^{(y)\dagger n}, \quad (46)$$

respectively. $A^{(y)}$ and its adjoint $A^{(y)\dagger}$ do not commute but satisfy the commutation relation

$$[A^{(y)}, A^{(y)\dagger}] = 2\frac{dy(x)}{dx}|_{x=P} \quad (47)$$

When a stochastic process $\{x(t)\}$ with mean 0 and finite variance is given, then define the auto-correlation function of $\{x(t)\}$ as

$$R(t_1, t_2) \triangleq E[x(t_1) x^*(t_2)] \ . \quad (48)$$

Then, in a similar manner to the method used in [13,14], we can show that the following relation holds.

$$R(s,t) = \sum_n \int \int w_n(a) dadb \cdot \gamma_n(a,b) \psi_{n,b+ai}^{(y)}(s) \psi_{n,a+ai}^{(y)*}(t) \ , \quad (49)$$

$$\gamma_n(a,b) \triangleq \gamma_n'(b+ai) \quad (50)$$

$$\gamma_n'(\alpha) \triangleq \mathcal{A}^{(y)-1}\{\mathcal{N}^{(y)}\{\beta_n'(\alpha)\}\}. \quad (51)$$

$$\beta_n'(\alpha) \triangleq \beta_n(Im\ \alpha, Re\ \alpha) \ . \quad (52)$$

$$\beta_n(a,b) \triangleq \int \int \psi_{n,b+ai}^{(y)*}(s) R(s,t) \psi_{n,b+ai}^{(y)}(t) dsdt \ . \quad (53)$$

Using these relation, the auto-correlation function can be transformed in the pseudo-orthogonal form (49) by the wavepackets of the eigenfunction system.

5 Conclusion

A larger class of the time-shift-invariant eigenfunction systems of almost localized wavepackets has been proposed, which contains the Cauchy wavelets and the coherent states. The 'pseudo-orthogonality' of these function systems has been investigated also. An application of them to the diagonalization of the auto-correlations of random signals via wavepackets are also proposed.

REFERENCES [1] Morlet,J. et al., *Geophysics*, **47**, 203- (1982). [2] Daubechies,I. *IEEE Trans. Information Theory*, **36**, 961- (1990). [3] Grossman,A. et al., in *Wavelets* (Combes,J.M. et al. eds.), Springer, 2- (1989). [4] Daubechies,I. *Ten Lectures on Wavelets*, SIAM (1992). [5] Meyer,Y., *Wavelets: Algorithms & Applications*(Ryan,R.D. tr.), SIAM . [6] Mallat,S., *IEEE Trans. Pattern Analysis & Machine Intelligence*, **11**, 674- (1989). [7] Glauber,R.J., *Phys. Rev.* , **131**, 2766- (1963). [8] Klauder,J.R. & Sudarshan E.C.G., *Fundamentals of Quantum Optics*, W. A. Benjamin, Inc. (1968). [9] Louisell,W.H., *Quantum Statistical Properties of Radiation*, John Wiley & Sons, Chap.3 (1973). [10] Perelomov,A.M., *Soviet Physics Uspekhi*, **20**, 703- (1977). [11] Sakaguchi,F., *Proc. ICCS/ISITA '93, Siagapore*, 107- (1993);. [12] Daubechies,I. & Paul,T., *Proc. 8th Int. Congr. Math. Phys., Marseilles*, 675- (1986); Paul,T. & Seip,K., in *Wavelets and Their Applications*(Ruskai eds.), Jones and Bartlett Publ., 303- (1992). [13] Sakaguchi,F., *The Transactions of IEICE, Series A*, **J77-A**, 1065- (1994) (in Japanese). [14] Sakaguchi,F., *Proc. ISITA 1994, Sydney*, vol.1, 309- (1994); *id., ibid.*, 303-. [15] Gel'fand,I.M. & Neumark,M.A. *C.R. Acad. Sci. USSR*, **55**, 567- (1947). [16] Aslaksen,E.W. & Klauder,J.R. *J. Math. Phys.*, **9**, 206- (1968); *J. Math. Phys.*, **10**, 2267- (1969). [17] Torrésani, B., J. Math. Phys., **32**, 1273- (1991). [18] Falomir,H. et al., J. Math. Phys., 35, 1939- (1994).

CHANNEL AND INTENSITY ESTIMATION FOR A CLASS OF POINT PROCESSES

Ananthram Swami & Brian Sadler
Army Research Lab
AMSRL-IS-TA, 2800 Powder Mill Road, Adelphi, MD 20783-1197, USA

ABSTRACT

We show that the blind LTI channel estimation problem, when the input sequence is independent, but has time-varying statistics, mimics that for the i.i.d. case under appropriate persistence of excitation conditions. Hence, consistent parametric and non-parametric estimators based on a single realization are readily obtained. We establish an ergodicity theorem for the time-averages of non-stationary continuous time processes; we use this to establish blind identifiability of the LTI channel of a filtered inhomogeneous point process, with multiplicative marks. These results extend to a class of time-varying channels as well. The theoretical results are corroborated by simulations.

1 LTI SYSTEMS WITH NON-STATIONARY WHITE INPUTS

Let $x(t)$ be a temporally independent discrete-time (DT) sequence whose statistics are time-varying, i.e., its cumulants (assuming that they exist) can be written in the form,

$$C_{kx}(t;\tau_1,...,\tau_{k-1}) := \text{cum}\ (x(t), x(t+\tau_1), \cdots, x(t+\tau_{k-1})) = \gamma_{kx}(t)\ \delta(\tau_1)\cdots\delta(\tau_{k-1}) \quad (1)$$

A simple example of such a process is the scaled process, $x(t) = v(t)s(t)$, where $v(t)$ is an iid sequence, and $s(t)$ is non-random. This model is often used to approximate seismic reflectivity sequences, where the variance of the process $x(t)$ is known to decay exponentially with time.

Let $h(t)$ be the impulse response of a linear time-invariant (LTI) system, and let

$$y(t) = \sum_{\rho} h(\rho)x(t-\rho); \quad z(t) = y(t) + w(t)\ , \quad (2)$$

where $w(t)$ is assumed to be stationary and independent of the signal $y(t)$. If input $x(t)$ satisfies (1), then with $\tau_0 \equiv 0$,

$$C_{ky}(t;\tau_1,...\tau_{k-1}) = \sum_{\rho} \gamma_{kx}(\rho) \prod_{i=0}^{k-1} h(t+\tau_i-\rho)\ . \quad (3)$$

If $x(t)$ is i.i.d., $\gamma_{kx}(\rho)$ is independent of ρ and (3) reduces to the well-known Bartlett-Brillinger-Rosenblatt formula [7]. The $\gamma_{kx}(\rho) \equiv \gamma_{kx}$ case has been well-studied [7].

Our objectives are: given only the noisy output $z(t)$ in (2), we may want to estimate the channel $h(t)$, or the input $x(t)$, or some statistics of the input. The continuous-time version of this problem has been studied for the $k = 2$ case in [8], under certain restrictive assumptions on $h(t)$. If $\gamma_{kx}(\rho)$ is periodic (e.g., $x(t) = u(t)s(t)$, with $u(t)$ stationary, and $s(t)$ periodic), then C_{ky} is also periodic in t, and one can use cyclic statistics to estimate the channel.

Assume as in [3] that the joint cumulants of $y(t)$ and $w(t)$ are absolutely summable, and that the appropriate limits exist (the assumptions hold under the sufficient conditions of bounded $\gamma_{kx}(t)$'s, and exponentially bounded $h(t)$'s; in the case of the scaled process $x(t) = u(t)s(t)$ considered earlier, $s(t)$ should be bounded.). Let

$$H(0) := \sum_t h(t); \qquad \Gamma_{kx} := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \gamma_{kx}(t)\ .$$

Under our modeling assumptions, both $H(0)$ and Γ_{kx} are well-defined, and are finite valued. Assume for convenience that $H(0)\Gamma_{1x} = 0$; let $\underline{\tau}_k := (\tau_1, \cdots, \tau_{k-1})$, and

$$< f(t) > \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} f(t)\ ; \quad M_{hx}(\underline{\tau}_k) := \sum_t \prod_{i=0}^{k-1} h(t+\tau_i)\ ,$$

which is the deterministic k-th order correlation of the impulse response (IR) $h(t)$. From (3), we obtain

$$< C_{ky}(t;\underline{\tau}_k) > \ = \ \Gamma_{kx} M_{kh}(\underline{\tau}_k)\ .$$

From [3], we know that the sample estimate

$$\begin{aligned} M_{ky,T}(\underline{\tau}_k) \quad &:= \quad \frac{1}{T} \sum_{t=1}^{T} \prod_{i=0}^{k-1} y(t+\tau_i) \\ &\longrightarrow_{ms} \quad \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left\{ \prod_{i=0}^{k-1} y(t+\tau_i) \right\} \\ &= \quad < M_{ky}(t;\underline{\tau}_k) > \ := \ M_{ky}(\underline{\tau}_k)\ , \end{aligned}$$

which is the time-averaged k-th order ensemble moment of the random process $y(t)$,

For $k = 2, 3$ we obtain for the linear model in (2)

$$\begin{aligned} M_{3y}(\tau_1,\tau_2) &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} M_{3x}(t;\tau_1,\tau_2) \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{\rho} \gamma_{3x}(t-\rho)h(\rho)h(\rho+\tau_1)h(\rho+\tau_2) \\ &= M_{3h}(\tau_1,\tau_2)\Gamma_{3x} \quad (4) \\ M_{2y}(\tau) &= M_{2h}(\tau)\Gamma_{2x} \quad (5) \end{aligned}$$

If the input is persistently exciting in the sense that $\Gamma_{2x} > 0$, it follows from (5) that the sample estimate of $M_{2y}(\tau)$ yields $M_{2h}(\tau)$, the deterministic correlation of IR $h(t)$, from which one can obtain a spectrally equivalent estimate of $h(t)$. Similarly if $\Gamma_{3x} \neq 0$, the sample estimate of $M_{3y}(\tau_1,\tau_2)$ yields $M_{3h}(\tau_1,\tau_2)$, the deterministic bicorrelation of IR $h(t)$, from which one can estimate $h(t)$ without making any phase assumptions; see [13] for some caveats. Several (non-) parametric methods are available to obtain the IR estimates. If

0-8186-7576-4/96 $5.00

the additive noise $w(t)$ in (2) is stationary, has zero-mean and zero bispectrum, then

$$M_{3x,\mathcal{T}}(\tau,\rho) \longrightarrow M_{3h}(\tau,\rho)\Gamma_{3x} \ .$$

Once $H(z)$ has been estimated, we can construct a zero-forcing equalizer (zfe) to estimate the channel input $x(t)$. In the SIMO case, where we have measurements, $y_i(t) = \sum_\rho h_i(\rho)x(t-\rho)$, $i = 1,...,L$, we can estimate $H_i(z) = B_i(z)/A_i(z)$ $i = 1,...,L$. To estimate $x(t)$, we find FIR filters $G_i(z)$ such that $\sum_{i=1}^{L} B_i(z)G_i(z) = 1$. The existence of such G_i's is guaranteed by the Bezoutian theorem provided the $B_i(z)$'s are co-prime [6]. This overcomes problems in inverting the individual $H_i(z)$'s (e.g., zeros on the unit circle, sharp band-pass filters, etc.). The input estimate is provided by $\sum_{i=1}^{L} G_i(z)A_i(z)y_i(t)$. Of course, zfe's are not useful in the noisy case, where one must use Wiener-type filters. The additive non-Gaussian case can be handled under poor SNR conditions by first estimating the spectrum of the noise [9, 10]. We note that some of these ideas have recently been used in the context of fractional sampling.

The fourth-order case is a bit more complicated. It also illustrates that although cumulants appear to be the natural tools for dealing with stationary processes and LTI systems, the convenience is lost when one deals with non-stationary processes (similar difficulties are apparent in the treatment of fourth-order cyclic cumulants, and with multiplicative models, which are intrinsically non-linear).

A natural way to estimate the time-averaged cumulant is to combine the time-averaged moments in the usual manner,

$$\begin{aligned}\hat{C}_{4y}(\tau_1,\tau_2,\tau_3) &= < M_{4y}(t;\tau_1,\tau_2,\tau_3) > \\ &\quad -[3] < M_{2y}(t;\tau_1) > < M_{2y}(t;\tau_2-\tau_3) > \\ &= < C_{4y}(t;\tau_1,\tau_2,\tau_3) > \\ &\quad +[3] \ < M_{2y}(t;\tau_1)M_{2y}(t+\tau_3;\tau_2-\tau_3) > \\ &\quad -[3] < M_{2y}(t;\tau_1) > < M_{2y}(t;\tau_2-\tau_3) >\end{aligned}$$

where the [3] denotes the three terms obtained by permuting the τ_i's. Since $< M_{2y}(t;\tau_2-\tau_3) > = < M_{2y}(t+\tau_3;\tau_2-\tau_3) >$, we obtain

$$\begin{aligned}\hat{C}_{4y}(\tau_1,\tau_2,\tau_3) &= < C_{4y}(t;\tau_1,\tau_2,\tau_3) > \\ &\quad +[3] \ dcov \ (M_{2y}(t;\tau_1), M_{2y}(t+\tau_3;\tau_2-\tau_3)) \qquad (6)\end{aligned}$$

where $dcov$ denotes the deterministic covariance,

$$dcov \ (f(t),g(t)) := < f(t)g(t) > \ - \ < f(t) > < g(t) > \ .$$

Eqn (6) expresses the estimate $\hat{C}_{4y}$ as the sum of the true quantity $< C_{4y}(t;\tau_1,\tau_2,\tau_3) >$ and 3 error terms; these error terms are small if the deterministic covariance of the second moment function is small; the error term can even be zero for specific τ_i's. Eqn (6) gives a precise quantification of the error and what 'slow variations' should be, if the stationary assumption is to be invoked in estimating fourth-order cumulants. From (3) we note that this is a function of both the filter $h(t)$, and of the variations $\gamma_{kx}(t)$.

For example, let $x(t) = u(t)s(t)$ with $u(t)$ iid, and $s(t) = 1 + \alpha\cos(2\pi f_o t)$. Let $H(f)$ be band-limited to $f_c \pm B$; then the 'error' terms will be non-zero only if $mf_o \in [f_c \pm 2B]$, $m = 1,2$. The analysis is readily extended to any periodic $s(t)$; a simple conclusion is that if $H(f)$ is low-pass or high-pass compared with f_o, the fundamental frequency, then the error terms vanish.

Simulations

The LTI system was chosen to be an AR(2) model with parameters $[1, 0, 0.75]$; an i.i.d. Laplace sequence, $u(n)$, with parameter $\lambda = 1$ was generated; the input to the LTI system was the non-stationary i.d. sequence, $x(n) = s(n)u(n)$, where $s(n)$ is a deterministic amplitude scaling sequence. In the first case, $s(n)$ decreased linearly from 2 at $n = 1$ to 1 at $n = N$, where $N = 1024$ is the number of samples. In the second case, $s(n) = 1 + \cos(t/10)$ is periodic with a period of 20π samples. The three panels of Figure 1 show the estimated autocorrelation sequence for the stationary case ($s(n) \equiv 1$) and for the two non-stationary cases; the zero-lag term was normalized to account for scaling differences ($\Gamma_{kx} \not\equiv 1$); the panels show the true values, the mean estimate and the error bars estimated from a set of 100 Monte Carlo trials. Figure 2 displays the estimates of $C_3(\tau,0)$ for the same set of data. It is clear that the estimates are unbiased - the curves corresponding to the true value and the mean estimate are virtually indistinguishable in the two figures. Correlation- and cumulant-based normal equations were used to estimate the AR parameters for the three cases. Table 1 shows the mean and standard deviations. In accordance with the theory, good parameter estimates are obtained in all three cases.

2 FILTERED INHOMOGENEOUS POISSON PROCESSES

An interesting extension of the preceding ideas is in estimation problems connected with the class of marked filtered inhomogeneous Poisson processes (IPP's), which are continuous time processes. The extension is natural since IPP's are limiting case of non-stationary Bernoulli processes.

Let τ_i denote the occurrence times of an IPP with intensity rate $\lambda(t)$; let $\underline{u}(i)$ denote the mark process, which is assumed to be iid and independent of the IPP; let

$$y(t) = \sum_{n=1}^{N(t)} h(t,\tau_n;\underline{u}_n) \ , \quad t_0 \le t \le T, \qquad (7)$$

where h is the IR of a causal linear time-varying (LTV) system, and $N(t)$ denotes the number of events over $[t_0,t]$. We make the standard assumption that the mark process is iid and independent of $N(t)$.

The second characteristic functional of $y(t)$ in (7) is, (following [12, Sec 5.7])

$$\begin{aligned}\psi(v) &:= \ln\phi(v) := \ln E\left\{\exp\left[j\int_{t_o}^{T} y(\sigma)v(d\sigma)\right]\right\} \\ &= \int_{t_o}^{T} \lambda(\tau)E_u\left[\exp\left\{j\int_{t_o}^{T} h(\sigma,\tau;\underline{u})dv(\sigma)\right\} - 1\right] d\tau\end{aligned}$$

where the expectation is with respect to (wrt) the marks, and $v(\sigma)$ is a suitably chosen function. For example, with

$$v(\sigma) := \begin{cases} 0, & t_0 \le \sigma < t_1 \\ \sum_{\ell=1}^{i}\alpha_\ell & t_i \le \sigma < t_{i+1} \\ & i = 1,..,k; t_{k+1} = T \end{cases} ,$$

we obtain,

$$\begin{aligned}\psi(v) &= \ln E\left\{\exp\left[j\sum_{i=1}^{k}\alpha_i y(t_i)\right]\right\} \\ &= \int_{t_0}^{T}\lambda(\tau)E_u\left\{\exp\left[j\sum_{i=1}^{k}\alpha_i h(t_i,\tau;\underline{u})\right] - 1\right\} d\tau\end{aligned}$$

Differentiating wrt the α_i's, and evaluating the derivative at the origin, we obtain,

$$\begin{aligned}&\mathrm{cum}(y(t_1),\cdots,y(t_k)) \\ &\quad = \int_{t_0}^{T}\lambda(\tau)E_u\left[\prod_{i=1}^{k} h(t_i,\tau;\underline{u})\right] d\tau \qquad (8)\end{aligned}$$

$$= E\{u^k\} \int_{t_0}^{T} \lambda(\tau) \left[\prod_{i=1}^{k} h(t_i - \tau) \right] d\tau \qquad (9)$$

The last equality follows from the assumption that $h(t, \tau_n; \underline{u}_n) = h(t - \tau_n)u_n$, i.e., the marks are multiplicative and the system is time-invariant; this model has been used to study low frequency noise [12, p 217].

Notice the similarity between (3) and (9). Here,

$$\gamma_{kx}(t) \equiv \lambda(t) E\{u^k\}.$$

The unmarked process is obtained when $u(t_i) \equiv 1$. It is interesting to note that the cumulants of all orders of the unmarked IPP are all non-zero, and are all identical to one another; this also demonstrates that the Poisson process is strongly non-Gaussian, and does not easily satisfy Gaussian central limit theorems [2]. On the other hand, this also facilitates performance analysis.

Let the time-averaged intensity be denoted by ($t_o = 0$)

$$\Lambda := \lim_{T \to \infty} \frac{1}{T} \Lambda(T) := \lim_{T \to \infty} \frac{1}{T} \int_0^T dt\ \lambda(t)\ .$$

For the IPP, the earlier assumption $H(0)\Gamma_{1x}(0) = 0$ translates to $Eu = 0$ or $H(0) = 0$ in practice (the case of a finite support $\lambda(t)$, i.e., $\Lambda = 0$, will be discussed elsewhere). The condition $H(0) = 0$ is readily guaranteed, for example, by cascading a DC notch filter.

In the previous section we used the ergodicity results of [3] which were derived for discrete-time processes. Here, we state the continuous-time counterparts (proofs being omitted due to lack of space).

Theorem 1. Let the processes $\{y_i(t)\}_{i=1}^{M}$ have absolutely integrable joint cumulants,

$$\int \sup_t |\tau_j \mathrm{cum}(y_{i_0}(t), \cdots, y_{i_{K-1}}(t + \tau_{K-1}))| d\underline{\tau} < \infty,$$

where $d\underline{\tau} := d\tau_1 \cdots d\tau_{K-1}$, $j = 1, ..., K-1$, and the i_ℓ's take on possibly repeated values in $[1, ..., M]$. Let $\tau_0 := 0$, and assume that the limit

$$M_K(\underline{\tau}_K) := \lim_{T \to \infty} \frac{1}{T} \int_0^T dt\ E \left\{ \prod_{i=0}^{K-1} y_i(t + \tau_i) \right\}$$

exists and is finite. Let

$$\hat{M}_{K,T}(\underline{\tau}_K) := \frac{1}{T} \int_0^T \prod_{i=0}^{K-1} y_i(t + \tau_i)\ dt$$

denote the sample estimate. Then,

$$\mathrm{cum}(\hat{M}_{K_0,T}(\underline{\tau}_{K_0}), ..., \hat{M}_{K_m,T}(\underline{\tau}_{K_m})) = O(T^{-m})\ .$$

Further, the estimates are asymptotically normal. □

As in the case of DT processes, the $k = 2, 3$ cases are illustrative (recall the assumption $Eu = 0$ or $H(0) = 0 \longrightarrow Ey(t) = 0$):

$$\hat{M}_{2y}(\tau) \longrightarrow < M_{2y}(t; \tau) >$$
$$= \lim_{T \to \infty} \frac{1}{T} \int_0^T dt\ C_{2y}(t; \tau)$$
$$= \Lambda E\{u^2\} \int h(\sigma) h(\sigma + \tau)\ d\sigma \qquad (10)$$

$$\hat{M}_{3y}(\tau, \rho) \longrightarrow < M_{3y}(t; \tau, \rho) >$$
$$= \Lambda E\{u^3\} \int h(\sigma) h(\sigma + \tau) h(\sigma + \rho)\ d\sigma \qquad (11)$$

From the estimated M_{2y} and/or M_{3y}, one can estimate $h(t)$ (parametric and/or non-parametric); in practice, one may sample the filter output $y(t)$, and use DT algorithms. As in the DT case, the fourth-order case is complicated.

Thus, using the ergodicity theorems for CT processes, we have established that the channel $h(t)$ can be estimated blindly, i.e., without knowledge of $\lambda(t)$, provided the persistence of excitation condition, $\Lambda > 0$, holds. As in the DT case, additive noise whose joint pdf's are symmetrically distributed (e.g., Gaussian, Laplace) can be handled by using M_{3y}. The non-Gaussian noise case can be handled as mentioned in an earlier section.

Once the channel $h(t)$ has been estimated, we can use EM-type approaches [1] or constrained estimators [11] to overcome problems with the MLE. Alternatively, we can equalize the channel and estimate $x(t)$; as noted earlier, this is easier in the SIMO case. Since the Poisson process is CT, very fine sampling is required.

In the noisy case, we must detect the points; here again, the approach of [9, 10] allows us to recover the spectrum of the noise. We can now use either LS or ML [12] to estimate the parameterized intensity function.

The non-parametric $\lambda(t)$ problem is generally ill-posed, since consistent estimates cannot be obtained from a single realization. If $\lambda(t)$ is periodic (e.g., auditory physiology, optical range-rate finding, shot noise in phase tracking loops, radar clutter in scatter communications, etc), then, one can use cyclic estimates. The period can be estimated efficiently via the FT - a simple generalization of the result in [14]. Using the results of Theorem 1 it is easy to show that the *post-stimulus time* (PST) histogram, see, e.g., [5], is a consistent estimate of $\lambda(t)$: the mean is $\lambda(t)$ and the variance is $\lambda(t)/K$, where K is the number of periods used to construct the estimate.

Simulations

We can generate an HPP easily since the inter-arrival times are independent, stationary and exponentially distributed. We can convert this to an IPP by non-linear warping of the time-axis. We can also generate inter-arrival times of an IPP by generating r.v.'s with the pdf $p(u) = \lambda(u)/\Lambda(T)$. The last two approaches involve computation of the inverse function, $\Lambda^{-1}(v)$. An alternative is the Lewis-Shedler thinning algorithm [4] which first generates points of an 'easy' IPP (i.e., an IPP whose $\Lambda(u)$ function is easily inverted), with intensity function $\lambda_e(t) \geq \lambda(t)$, $\forall t \in [t_o, T]$; points are then deleted by generating an auxiliary set of uniform r.v.'s. We used this approach and generated the HPP points with constant intensity $\lambda \geq \max_t \lambda(t)$. Simulation parameters were: analog filter $h(t) = \exp(-3t)\cos(5t)$, output sampling interval of 0.01s, 500 Laplace-distributed marks, 1024 output samples, and 100 trials. An AR(2) model, which is appropriate for this $h(t)$, was fitted to the output time series; the mean and standard deviations, estimated over a Monte Carlo run of 100 trials, are shown in Table 2 for the three input processes: a stationary process, an HPP and an IPP. The TV intensity function for the IPP was $\lambda(t) = 1 + \Delta(t; 0.2, 0.3) + \Delta(t; 0.2, 0.7)$, where $\Delta(t; \tau, t_o)$ denotes a triangular pulse of duration τ centered at $t = t_o$. The stationary process consisted of 500 equi-spaced samples with Gaussian amplitudes. Good parameter estimates were obtained in all three cases, as promised by the theoretical development.

3 DISCUSSION

We used the theory of mixed-time averages in [3] to establish the identifiability of DT linear systems driven by non-stationary white processes. We extended the theory to continuous time processes, and used that to establish blind identifiability of CT linear systems driven by a marked inhomogeneous Poisson process. We stress that these results

were established under weak persistence of excitation conditions. As in the case of cyclic statistics, deterministic signals in noise, and multiplicative noise models [13], we note that moment statistics are easier to manipulate than cumulant statistics. The theoretical results were corroborated by simulations.

These results can be extended to a class of self-exciting point processes, called non-stationary renewal processes (nsrp), where the intensity function is of the form

$$\mu(t; w_1, ..., w_{N(t)}) = \lambda(t) r(t - w_{N(t)}) ,$$

where $r(t)$ is the *recovery function* which is assumed to be monotonically non-decreasing with values in $[0, 1]$. Here, w_i's are the occurrence times and $N(t)$ is the number of events in $[t_o, t)$, so that $w_{N(t)}$ is the time of occurrence of the last event prior to t. For this model, we are interested in estimating both $\lambda(t)$ and $r(t)$ given the filtered output $y(t)$. In the case of marked processes the analysis is particularly easy if the marks are multiplicative and i.i.d. (typical assumptions); the non-zero mean case is easier to handle. If the intensity function is periodic, one can exploit cyclic statistics as well. Some of these issues and the extension of Theorem 1 to nsrp's will be presented in the full paper.

IPP's may be useful in modeling impulsive correlated noise (bursts). Here, a non-conventional renewal function (decreasing rather than increasing) is required to model the fluctuations; e.g., $\mu(t|N_t) = \lambda(t)[1 + f(t - N_t)]$. The pdf of the marks dictate the amplitude distribution of the noise. As in the case of DT-LTV models, one could expand $\lambda(t)$ on a set of known basis functions (e.g., sines and cosines), and then estimate the coefficients.

The extension to the TV channel is straightforward, provided the time-variations can be expressed in terms of known basis functions, i.e., the unknown projection coefficients themselves are not time-varying. It is interesting to note that the theory of time-averaged moments is also useful in the input-output analysis of non-linear systems.

REFERENCES

[1] N. Antoniados & A.O. Hero, "Time-delay estimation for filtered Poisson processes using an EM-type algorithm", *IEEE Trans SP*, 2112-23, Aug 1994.

[2] D.R. Brillinger, "Comparative aspects of the study of ordinary time series and of point processes", in *Developments in Statistics*, vol 1, 33-133, Academic Press, 1978.

[3] A.D. Dandawate and G.B. Giannakis, "Asymptotic theory of mixed-time averages", *IEEE Trans Info Theory*, 45, 216-32, Jan 1995.

[4] L. Devroye, *Non-uniform random variate generation*, Springer-Verlag. 246-258, 1986.

[5] D.H. Johnson and A. Swami, "The Transmission of Signals by Auditory-Nerve Fiber Discharge Patterns", *J. Acoust. Soc. Amer.*, 74(2), 493-501, Aug. 1983.

[6] T. Kailath, *Linear systems*, Prentice-Hall, 1980.

[7] J.M. Mendel, "Tutorial on higher order statistics ...", *Proc. IEEE*, 79(3), 278-305, Mar 1991.

[8] A. Papoulis, "Identification of systems driven by non-stationary noise", *IEEE Trans. Info. Theory*, 240-4, March 1978.

[9] B. Sadler, G.B. Giannakis and K.-S. Lii, "Estimation and detection in non-Gaussian noise using higher-order statistics", *IEEE Trans. SP*, 42(10), 2729-41, Oct 1994.

[10] B. Sadler, "Detection in correlted impulsive noise using fourth-order cumulants", *Proc. CISS-95*, 600-5, John Hopkins University Press, Baltimore, MD, Mar 1995.

[11] R.E. Sequeira & J.A. Gubner, "Intensity estimation from shot noise data", *IEEE Trans. SP*, 1527-31, June 95.

[12] D.L. Snyder & M. I. Miller, *Random Point Processes in Time and Space*, Springer-Verlag, 1991.

[13] A. Swami, "Pitfalls in polyspectra", *Proc. ICASSP-93*, Minneapolis, MN, Vol IV, 97-100, April 1993.

[14] D. Vere-Jones, "On the estimation of frequency in point-process data", *Applied Probability Trust*, 383-394, 1982.

Table 1. AR(2) system identification with stationary and nonstationary two-sided exponential driving sequence. Second and fourth-order statistics results are compared.

True	*2nd*-order	*4th*-order
stationary case		
0	−.0003 (.0301)	−.0154 (.1790)
.75	.7397 (.0205)	.7498 (.1490)
nonstationary case 1		
0	−.0075 (.0423)	−.0126 (.2787)
.75	.7406 (.0302)	.7525 (.1598)
nonstationary case 2		
0	−.0036 (.0390)	−.0172 (.0797)
.75	.7404 (.0298)	.7338 (.0695)

Table 2. AR Estimates for filtered IPP

	a(1)	a(2)
Sty	-1.8786 (0.1035)	0.8863 (0.0967)
HPP	-1.8719 (0.1068)	0.8801 (0.0996)
IPP	-1.8599 (0.1447)	0.8690 (0.1348)

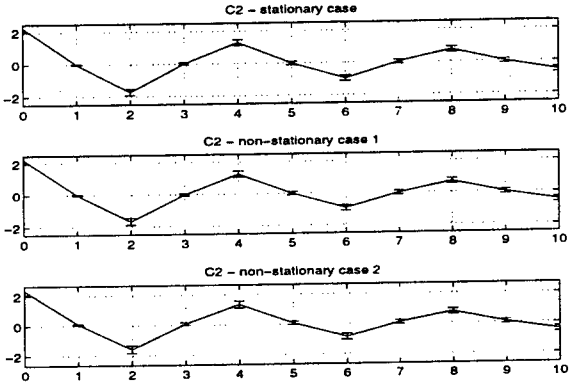


Figure 1. C2 Estimates

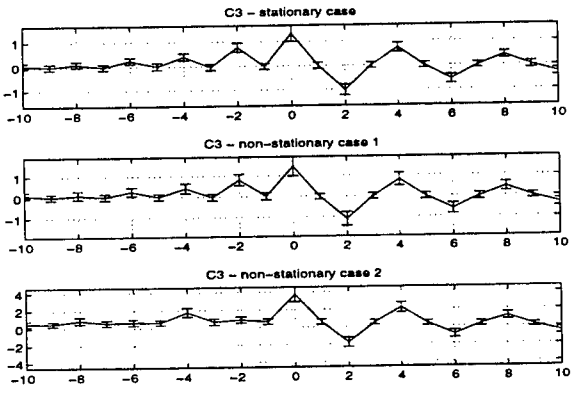


Figure 2. C3 Estimates

Statistical Analysis of Polynomial Phase Signal Parameter Estimates based on Structured Auto-Regressive Modeling *

Jakob Ängeby
Chalmers University of Technology
412 96 GÖTEBORG, SWEDEN

Abstract

A statistical analysis of the polynomial phase signal parameter estimates achieved when using the structured auto-regressive approach is presented. The estimates are consistent for high SNR *or large number of samples, N. An expression for the covariance of the estimates is given. Numerical examples confirm that the theoretical covariance apply well to empirical data for a wide range of* SNR *and N. The performance of the estimator depends on the filter length, n, and the sampling strategy which may be non-uniform. The optimal choice of n for evenly sampled cisoids is given as a function of N. The variance is inversely proportional to* SNR^2 *for small* SNR, *and to* SNR *for medium and high* SNR.

1 Introduction

In a variety of applications, such as radar, sonar, geophysics and communication it is of great interest to estimate the parameters of a non-linear phase function of complex valued signals. Such signals can be modeled as

$$s(t) = \sum_{l=1}^{m} b_l(t) \exp\{j\, a_l(t)\} \qquad (1)$$

where $a_l(t)$ and $b_l(t)$ are real valued continuous time functions that model the phase and amplitude, respectively. The signal $s(t)$ is sampled at time instants $\{t_k\}_{k=0}^{N-1}$, and the measurements of the signal are corrupted by noise, which in many scenarios can be modeled as additive, white and Gaussian

$$y(t_k) = s(t_k) + e(t_k) \qquad (2)$$

*The work was supported by *Ericsson Microwave Systems AB* and *Ericsson Infocom Consultants AB*

where $E[e(t_k)] = 0$, $E[e(t_k)e(t_l)] = 0$ and $E[e(k)e^*(t_l)] = \sigma^2\delta_{k,l}\ \forall\ k,l$. A special case of non-linear phase is the polynomial phase function

$$a_l(t_k) = a_0 + a_1 t_k + a_2 t_k^2 + \ldots + a_q t_k^q \qquad (3)$$

The polynomial phase function has been proven useful in important applications. In Doppler radar, for example, radar returns from maneuvering targets give rise to a non-linear phase that can be modeled by a polynomial. Estimates of the polynomial coefficients can then be used to determine the target's kinematic parameters (velocity, acceleration etc). Considerable attention has been payed to the estimation of the parameters of non-linear phase signals, see for example [1, 3, 4, 5, 6] and the references therein.

In [1] an approach based on a structured auto-regressive model was proposed and proven successful using simulations. The proposed method estimates the phase and amplitude parameters of a quite general class of signals, including polynomial phase signals, and has some interesting properties. For example, the structured AR model is a model based time-frequency representation (TFR), and the data is *not* constrained to be evenly sampled in time, as is the case in [3, 4, 5, 6]. On the contrary, it was empirically shown that the variance of the estimates can be significantly reduced by using a time-varying sampling period. In [1] a numerical study of the performance of the structured AR approach showed that the bias is negligible and the relative efficiency (the variance of the estimates derived by Cramér-Rao lower bound) typically attains a value of 1.5-2.0. No theoretical analysis was, however, presented and a number of important questions were left unanswered. Here a theoretical analysis is presented for the case of a mono-component polynomial phase signal with con-

0-8186-7576-4/96 $5.00

stant amplitude:

$$s(t_k) = b_0 e^{j\,a(t_k)} \qquad (4)$$

where b_0 is a positive constant and $a(t_k)$ was defined in (3).

2 Structured AR modeling

Consider a linear projection of $\{y(t_{k-p}\}_{p=1}^{n}$ to $y(t_k)$

$$\hat{y}(t_k) = -\sum_{p=1}^{n} c_p y(t_{k-p})$$

where $(c_1, \ldots, c_n)^T \triangleq \theta$ denote the AR parameters. It is well known how to calculate the AR parameters that minimize the variance of the prediction errors when the signal is WSS and correlation ergodic. Then, the AR parameters carry information on the signal $\{s(t_k)\}_{k=0}^{N-1}$ and can be used as a tool to get estimates of the signal parameters. The information carried in the AR parameters resembles an average over time, and is obviously of little value if the signal is time-varying. If, however, an *ensemble* of M realizations was available; then the "instantaneous" θ at time instant t_k, say $\theta(t_k)$, could be calculated from the ensemble, and $\theta(t_k)$ would carry information on the signal at time t_k. Properties such as the instantaneous frequency and spectral density could then easily be obtained. The *analytical* AR filter of order n that minimizes the *expected* prediction error variance at time instant t_k can be derived as a function of the signal parameters. Let ϑ_0 denote the true signal parameters. Consider the hypothetical case when an infinite number of realizations are available. The AR parameters that minimize the prediction error variance at time t_k, denoted by $\varepsilon(t_k;\vartheta)$, satisfy the well known projection theorem, i.e. the projection error shall be orthogonal to the data used,

$$\mathrm{E}[\varepsilon(t_k;\vartheta)y^*(t_{k-p})] = 0,\ p = 1, \ldots, n$$

where E[.] denotes expectation taken over the ensemble. The solution is

$$\theta(t_k;\vartheta) = -\bar{\mathbf{R}}^{-1}(t_k;\vartheta)\ \bar{\mathbf{r}}(t_k;\vartheta) \qquad (5)$$

which defines the mapping from ϑ to the AR parameters. In (5) $\bar{\mathbf{R}}(t_k;\vartheta)$ and $\bar{\mathbf{r}}(t_k;\vartheta)$ denote the "structured" covariance matrix and vector, respectively, whose elements consist of the analytical covariance function

$$\bar{r}(t_k;\vartheta,u,v) \triangleq E[y(t_{k-u})y^*(t_{k-v})] \qquad (6)$$

Consider a "structured" analytical AR filter, denoted by $\theta(t_k;\vartheta)$, that is a function of a *model* signal, parameterized by ϑ . In order to derive the structured AR filter, the following notation is convenient:

$$\begin{aligned} x(t_k;\vartheta) &\triangleq \sum_{l=1}^{p} b_l(t_k;\vartheta)e^{j\,a_l(t_k;\vartheta)}, \\ & \quad b_1(t_0;\vartheta) \equiv 1 \qquad (7) \\ \mathbf{x}(t_{k-1};\vartheta) &\triangleq (x(t_{k-1};\vartheta), \ldots, x(t_{k-n};\vartheta))^T \end{aligned}$$

The reason for the normalization of $b_1(t_0;\vartheta)$ to 1 is explained below.
It is straightforward to show that $\bar{\mathbf{r}}(t_k;\vartheta)$ and $\bar{\mathbf{R}}(t_k;\vartheta)$ can be written as

$$\begin{aligned} \bar{\mathbf{r}}(t_k;\vartheta) &= \mathbf{x}^*(t_{k-1};\vartheta)x(t_k;\vartheta) \\ \bar{\mathbf{R}}(t_k;\vartheta) &= \mathbf{x}^*(t_{k-1};\vartheta)\mathbf{x}^T(t_{k-1};\vartheta) + \sigma^2\mathrm{I} \end{aligned}$$

where I denotes an $(n \mid n)$ identity matrix and '$*$' denotes complex conjugate. Formula (5) involves an inversion of the $(n \mid n)$ matrix $\bar{\mathbf{R}}(t_k;\vartheta)$ which implies heavy computations. This can be avoided by a straightforward use of the matrix inversion lemma; $(A+BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}$. Using $A = \sigma^2\mathrm{I}_n$, $B = \mathbf{x}^*(t_{k-1};\vartheta)$, $C = 1$ and $D = \mathbf{x}^T(t_{k-1};\vartheta)$ and assuming that $\sigma^2 \neq 0$ gives the final formula for the structured AR-filter:

$$\begin{aligned} \theta(t_k;\vartheta) &= \beta(t_k;\vartheta)\ \mathbf{x}^*(t_{k-1};\vartheta)x(t_k;\vartheta) \\ \beta(t_k;\vartheta) &= -\frac{\mathrm{SNR}_0}{1+\mathrm{SNR}_0\ \mathbf{x}^H(t_{k-1};\vartheta)\mathbf{x}(t_{k-1};\vartheta)} \\ \mathrm{SNR}_0 &\triangleq \frac{b_1^2(t_0;\vartheta)}{\sigma^2} \end{aligned}$$

At this point the normalization in (7) becomes clear. Consider $\beta(t_k;\vartheta)$ above. Suppose $x(t_k;\vartheta)$ was not normalized and write $x(t_k;\vartheta) = \tilde{b}\ \tilde{x}(t_k;\vartheta)$, where $\tilde{b}$ is some constant. If the normalization was not introduced then the mapping from the duplet $(\tilde{b}^2, \sigma^2)$ to $\beta(t_k;\vartheta)$ would not be 1:1, which would cause the Hessian used in the non-linear search for the signal parameters to become singular.
Note that $\bar{r}(t_k;\vartheta,u,v)$ and hence $\theta(t_k;\vartheta)$ can be calculated for non-uniformly sampled data. For the case of a mono-component signal with time invariant amplitude $\beta(t_k;\vartheta)$ simplifies to

$$\beta(t_k;\vartheta) \equiv \beta_0 \triangleq -\frac{\mathrm{SNR}}{1+n\mathrm{SNR}} \qquad (8)$$

where

$$\mathrm{SNR} \triangleq \frac{b_0^2}{\sigma^2}$$

The structured AR filter $\theta(t_k;\vartheta)$ is used to predict $y(t_k)$ and the prediction is given by

$$\hat{y}(t_k;\vartheta) = -\theta^T(t_k;\vartheta)\,\mathbf{y}(t_{k-1})$$
$$\mathbf{y}(t_{k-1}) \triangleq (y(t_{k-1}),\ldots,y(t_{k-n}))^T$$

which implies

$$\begin{aligned}\varepsilon(t_k;\vartheta) &= y(t_k) - \hat{y}(t_k;\vartheta)\\ &= y(t_k) + \theta^T(t_k;\vartheta)\,\mathbf{y}(t_{k-1})\end{aligned}$$

The signal parameter estimates are found by applying $\theta(t_k;\vartheta)$ to a single realization and minimizing the prediction error variance with respect to ϑ :

$$\hat{\vartheta} = \arg\min_{\vartheta} V(\vartheta) \qquad (9)$$

$$V(\vartheta) = \tfrac{1}{N}\textstyle\sum_{k=n}^{N-1} \mid \varepsilon(t_k;\vartheta)\mid^2 \qquad (10)$$

The minimization of (10) is easily implemented using e.g. a Gauss-Newton search. The derivatives of $\theta(t_k;\vartheta)$ with respect to ϑ are straightforward to compute. The search can be implemented off- or on-line (tracking).

3 Statistical performance

In the following it is assumed that the signal is described by (3)-(4). All results are given without proof due to lack of space. The proofs can be found in [2].

Result 1 *The structured AR signal parameter estimates are consistent for large number of samples or large SNR, i.e.* $\hat{\vartheta} \to \vartheta_0$ *as* $N \to \infty$ *or* $\mathrm{SNR} \to \infty$.

For asymptotic SNR, the result holds under the assumption that the data set consists of at least q samples with non zero sampling interval. For asymptotic N, the result holds under the assumption that t_k increases without bound as as $N \to \infty$.

Result 2 *Assume the noise is white and Gaussian and that* $\hat{\vartheta}$ *is close to* ϑ_0 *. Then the signal parameter estimates are Gaussian with covariance matrix*

$$\mathrm{cov}(\hat{\vartheta}) = E[V''(\vartheta)]^{-1}\,\mathrm{cov}(V'(\vartheta))\,E[V''(\vartheta)]^{-1}$$

where

$$\begin{aligned}\mathrm{cov}(V'(\vartheta)) &= \tfrac{1}{2}\sigma^4{\beta_0}^2\ldots\\ &\ldots \sum_{k=n}^{N-1}\sum_{l=\max(n,k-n)}^{\min(N-1,k+n)} \nabla\mathbf{a}_k^T\Lambda_{k,l}\nabla\mathbf{a}_l\end{aligned}$$

$$\begin{aligned}\Lambda_{k,l} &= (1-\beta_0)(\mathrm{SNR}\,\mathbf{1}\mathbf{1}^T+\mathrm{I})\delta_{k,l} - \ldots\\ &\quad {\beta_0}^2(\mid l-k\mid \mathrm{SNR}+1)\,\mathbf{1}\mathbf{1}^T - \ldots\\ &\quad {\beta_0}^2\mid l-k\mid \mathrm{I}_{l-k} - \ldots\\ &\quad \beta_0(\mathbf{1}_{k-l}\mathbf{e}_{l-k}^T + \mathbf{e}_{k-l}\mathbf{1}_{l-k}^T) - \ldots\\ &\quad {\beta_0}^2\mathbf{1}_{k-l}\mathbf{1}_{l-k}^T\end{aligned}$$

$$E[V_N''(\vartheta_0)] = \sigma^2{\beta_0}^2\sum_{k=n+1}^{N}\nabla\mathbf{a}_k^T(\mathrm{SNR}\mathbf{1}\mathbf{1}^T+\mathrm{I})\nabla\mathbf{a}_k$$

$$\nabla\mathbf{a}_k = \begin{pmatrix} t_k - t_{k-1}, & \ldots, & t_k^q - t_{k-1}^q\\ \vdots & & \vdots\\ t_k - t_{k-n}, & \ldots, & t_k^q - t_{k-n}^q\end{pmatrix}$$

and β_0 *was defined in (8),* I *denotes the identity matrix of order* n, I_{l-k} *denotes an* $(n \mid n)$ *matrix with zeros everywhere except on the* $(l-k)$*:th diagonal which is filled with ones,* $\mathbf{1}$ *denotes an* $(n \mid 1)$ *vector filled with ones,* $\mathbf{1}_{l-k}$ *denotes an* $(n \mid 1)$ *vector with ones on all entries except on entries* $1,\ldots,(l-k)$ *if* $(l-k) > 0$ *or* $n+(l-k),\ldots,n$ *if* $(l-k) < 0$ *which are filled with zeros, and* e_{l-k} *denotes an* $(n \mid 1)$ *vector with zeros on all entries except the* $(l-k)$*:th which equals 1.*

Result 2 holds provided $\hat{\vartheta}$ is close to ϑ_0 which, according to Result 1, will happen either for large enough SNR *or* N. The statistical performance of the signal parameter estimates depend on the structured AR filter length. Result 2 can be used to derive the optimal filter length, n, that minimizes the variance of the signal parameter estimates. The following result holds for the case of a uniformly sampled linear phase signal:

Result 3 *For linear phase signals that are uniformly sampled and of medium to high SNR, the optimal choice of* n *is* $\frac{N}{3}$.

4 Numerical examples

In the figures below, dashed curves correspond to empirical results, solid curves to theoretical (Result 2) and dash-dotted curves correspond to the Cramér-Rao lower bound. The empirical variance was calculated using Monte Carlo simulations based on 50 runs for each set of variable values. The signal was chosen to be a quadratic FM signal; $a(t_k) = \pi + 30\pi t_k - 80\pi t_k^2 + 70\pi t_k^3$. The figures illustrate the performance of $\hat{a}_2$. The corresponding figures for $\hat{a}_1$ and $\hat{a}_3$ show the same behavior and are therefore excessive.
Figures 1-2 illustrate how the variance of the structured AR estimates depend on the filter

length n (SNR $\equiv$5dB), and on SNR ($n \equiv 15$), respectively; $N = 100$, data uniformly sampled with $t_k - t_{k-1} = 0.01$; $\{t_k\}_{k=0}^{N-1} = (0.01, \ldots, 1)$. As seen from Fig 1 the variance of the phase parameters rapidly decreases with increasing n until an optimal value is reached. The theoretical variance closely follows the empirical and, most importantly, successfully predicts the optimal choice of n, which for this case is $n_{\text{opt}} = 14$. This is of great practical importance since it implies that the theoretical variance expression can be used for optimal filter design. Result 3 only applies to linear phase signals ($q = 1$) and can therefore not be applied to this example of a quadratic FM ($q = 3$). Expressions for the optimal choice of n for $q > 1$ is under current investigation. An empirical investigation indicates, however, that n_{opt} is close to inversely proportional to q.

Figure 2 illustrates how the variance depends

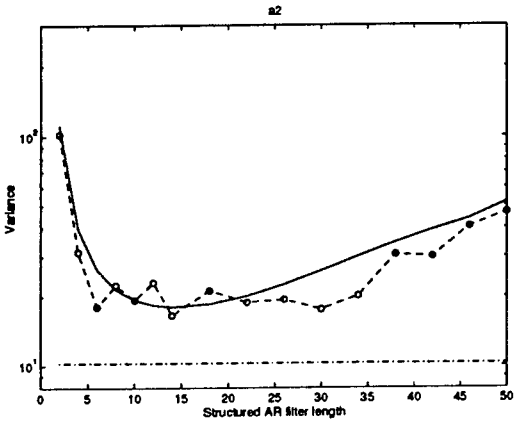


Figure 1. Variance vs n

on SNR. The theoretical variance is inversely proportional to SNR^2 for low SNR, and inversely proportional to SNR for medium and high values of SNR. The theoretical variance coincides well with the empirical down to a threshold, say SNR_T, below which it no longer applies. For SNR:s below SNR_T the series expansion used to derive the theoretical variance is not valid.

The theoretical covariance expression also applies very well to small data sets (N $\sim 10 - 20$, SNR=5dB). Illustrations can be found in [2] but are not presented here due to lack of space.

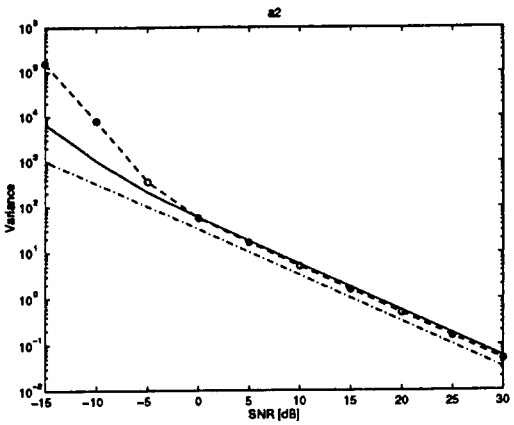


Figure 2. Variance vs SNR

5 Concluding remarks

The theoretical covariance has been compared to empirical results for a wide scenario of N, SNR, filter lengths, and (non-uniform) sampling strategies. It has been verified that the theoretical expression accurately predicts the empirical variance for SNR and N down to a threshold. The threshold is, however, low and Result 2 can be applied to most scenarios of practical interest.

References

[1] J. Ängeby. Structured autoregressive instantaneous phase and frequency estimation. In *ICASSP*, volume 3, pages 1768–1771. IEEE, May 1995.

[2] J. Ängeby. Statistical analysis of polynomial phase signal parameter estimates based on structured auto-regressive modeling. Technical Report CTH-TE-43, Chalmers University of Tech., Göteborg, Sweden, 1996.

[3] P. M. Djuric and M. Kay. Parameter estimation of chirp signals. *IEEE Trans. Sig. Proc.*, 38:2118–2126, Dec 1990.

[4] B. Friedlander and J. M. Francos. Estimation of amplitude and phase parameters of multicomponent signals. *IEEE Trans. Sig. Proc.*, 43(4):917–926, Apr 1995.

[5] S. Peleg and B. Porat. The crame'r rao lower bound for signals with constant amplitude and polynomial phase. *IEEE Trans. ASSP*, 39:749–752, March 1991.

[6] S. Peleg and B. Porat. Estimation and classification of polynomial phase signals. *IEEE Trans. Inform. Theory*, 37:422–430, March 1991.

Comparative Performance Analysis of Two Algorithms for Instantaneous Frequency Estimation*

Guy Reina and Boaz Porat
Department of Electrical Engineering
Technion-Israel Institute of Technology
Haifa 32000, Israel

Abstract

This paper compares two algorithms for estimating the instantaneous frequency of complex signals: the high-order ambiguity function (HAF) and the polynomial Wigner-Ville distribution (PWVD). Comparison is made by asymptotic first-order error analysis, which is verified by simulation. It is shown that when the signal phase is a polynomial function of time, the HAF always outperforms the PWVD. Two other advantages of the HAF over the PWVD are: (a) it has lower computational complexity; (b) unlike the PWVD its use is not limited to the instantaneous frequency at the middle of the observation interval.

1. Introduction

Let $s(t)$ be the complex signal

$$s(t) = a(t)\exp\{j\phi(t)\}, \quad 0 \le t \le T. \tag{1}$$

The instantaneous frequency (IF) of this signal is defined as

$$\omega(t) = \frac{d\phi(t)}{dt}, \quad 0 \le t \le T. \tag{2}$$

Estimation of the IF over the interval [0,T] from noise-corrupted measurements of $s(t)$ is important in areas such as radar, communication and sonar.

This paper examines two methods for estimating the IF. The methods are:

1. The high-order ambiguity function (HAF). This method models the phase function of the signal as an Mth-order polynomial

$$s(t) = C\exp\{j\phi(t)\}, \quad \phi(t) = \sum_{m=0}^{M} \alpha_m t^m. \tag{3}$$

The coefficients of the polynomial are estimated successively, starting at the highest order. The IF of the signal is obtained by differentiating the estimated phase polynomial. The definition of the HAF and the details of the algorithm are given in [1].

2. The Polynomial Wigner-Ville Distribution (PWVD), introduced by Boashash in [2] is defined as

$$PWVD(t,\omega) = \int_0^{\bar{T}} K^{(M)}(t,\tau) e^{-j\omega\tau} d\tau, \tag{4}$$

where

$$K^{(M)}(t,\tau) = \prod_{i=0}^{M/2} \left[s(t + c_i\tau)\right]^{b_i} \left[s^*(t + c_{-i}\tau)\right]^{b_{-i}}, \tag{5}$$

and $\bar{T} = \min\{t, T-t\} / \max\{c_i\}$. The IF is estimated as

$$\hat{\omega}(t) = \arg\max_{\omega}\{PWVD(t,\omega)\} \tag{6}$$

The parameters $\{b_i, b_{-i}, c_i, c_{-i}\}$ are chosen so as to make $\hat{\omega}(t)$ unbiased.

This paper compares the two methods from accuracy and complexity points of view. Accuracy is evaluated by means of analytic derivation of the asymptotic errors, which are then verified by simulations. The analytic derivation of the asymptotic errors of the HAF has been done before [3], but that of the PWVD is new. The analysis is limited to the case of constant amplitude polynomial phase signals.

The paper gives the details of the analysis of the two algorithms, followed by a selected set of simulation results.

2. Derivation of the error formulas

In this section, we present the analytic derivation of the asymptotic error formula for the PWVD. The details of the error formula for the HAF, can be found in [3]. The results obtained are asymptotic in respect to the data

* B. Porat acknowledges the support of the Technion's VPR Fund for the Promotion of Research.

0-8186-7576-4/96 $5.00 © 1996 IEEE

point number tending to infinity. The derivations do not impose any restriction on the SNR.

2. 1. Preliminaries

The signal model is

$$y(n) = z(n) + w(n), \tag{7}$$

where $z(n)$ is a unit-modulus polynomial phase discrete-time signal, defined by

$$z(n) = \exp\{j\phi(n)\}, \quad \phi(n) = \sum_{m=0}^{M} \alpha_m n^m, \tag{8}$$

and $w(n)$ is a complex circular white Gaussian noise with variance σ^2. We can rewrite (7) as

$$y(n) = z(n)\left[1 + w(n)/z(n)\right] = z(n)\left[1 + v(n)\right], \tag{9}$$

where $v(n)$ is probabilistically equivalent to $w(n)$, that is, white and Gaussian, with moments given by

$$\begin{aligned} &E\left\{\left[v(n)\right]^k\right\} = 0, \quad k = 1,2,\ldots \\ &E\left\{\left[v(n)\right]^k\left[v^*(n)\right]^\ell\right\} = 0, \quad k \neq \ell \\ &E\left\{\left[v(n)\right]^k\left[v^*(n)\right]^k\right\} = k!\sigma^{2k}, \; k = 0,1,2,\ldots \end{aligned} \tag{10}$$

The following two formulas will be needed later:

$$E\left\{\left[1+v(n)\right]^k\right\} = \sum_{i=0}^{k}\binom{k}{i}E\left\{\left[v(n)\right]^i\right\} = 1 \tag{11}$$

$$E\left\{\left[1+v(n)\right]^k\left[1+v^*(n)\right]^\ell\right\} = \sum_{i=0}^{\min\{k,\ell\}}\binom{k}{i}\binom{\ell}{i} i!\sigma^{2i} \tag{12}$$

The noise, being Gaussian, has bounded moments. Let $\{X_n\}$ be a sequence of (real or complex) random variables, and $\{a_n\}$ a sequence of positive real numbers. We will use the notation $X_n = O_m(1)$ to mean that all the moments of X_n are bounded uniformly in n, that is: for every positive integer there exists a positive constant $B(k)$ such that $E|X_n|^k \le B(k)$ for all n. The notation $X_n = O_m(a_n)$ will mean that $X_n/a_n = O_m(1)$.

Both the HAF and the PWVD algorithms search for the maximum point of a discrete Fourier transform (DFT). We therefore need a formula expressing the perturbation of the maximum point of the DFT as a function of the measurement error. Let $\omega_0 \in [-\pi, \pi]$ and

$$y(n) = e^{j\omega_0 n} + a(n), \tag{13}$$

where $\{a(n)\}$ is some additive interference. Let $Y(\omega)$ denote the DFT of $y(n)$, also let $A^{(k)}(\omega)$ denote the kth derivative of the DFT of $a(n)$ with respect to ω (with k = 0 denoting the DFT itself). Introduce the following assumptions:

- **Assumption A:** $a(n) = O_m(1)$;
- **Assumption B:** $A^{(k)}(\omega_0) = O_m\left(N^{k+0.5}\right)$.

Under assumptions A and B one can show [3] that

$$\hat{\omega} - \omega_0 = 12N^{-3}\Re\left\{A^{(1)}(\omega_0) - j0.5N\left[A^{(0)}(\omega_0)\right]^*\right\} + O_m\left(N^{-2}\right), \tag{14}$$

where $\hat{\omega}$ is the point of local maximum of $|Y(\omega)|^2$.

2. 2. The error formula for the PWVD

As mentioned before, the PWVD algorithm include maximization of the DFT of the signal (9) after the application of the transformation (5). The discrete version of the transformation kernel is

$$K_y^{(q)}(n,m) = \prod_{k=0}^{q/2}\left[y(n+c_k m)\right]^{b_k}\left[y^*(n+c_{-k}m)\right]^{-b_{-k}}. \tag{15}$$

Passing the input signal (9) through the above kernel, yields the signal

$$k_y^{(q)}(n,m) = \exp\{j2\pi m\, IF(n)\}\left[1 + a_n(m)\right], \tag{16}$$

where $a_n(m)$ is defined by

$$a_n(m) = \prod_{k=0}^{q/2}\left[1+v(n+c_k m)\right]^{b_k}\left[\left(1+v(n+c_{-k}m)\right)^*\right]^{b_{-k}} - 1, \tag{17}$$

and $IF(n)$ is the instantaneous frequency at time n.

The mean of $a_n(m)$ is

$$E\left[a_n(m)\right] = E\left[\prod_{k=0}^{q/2}\left[1+v(n+c_k m)\right]^{b_k}\left[\left(1+v(n+c_{-k}m)\right)^*\right]^{b_{-k}} - 1\right]. \tag{18}$$

For the case $m \neq 0$, the two terms in the product are independent, and the mean can be expressed as

$$E\left[a_n(m)\right] = \prod_{k=0}^{q/2}E\left[1+v(n+c_k m)\right]^{b_k}E\left[\left(1+v(n+c_{-k}m)\right)^*\right]^{b_{-k}} - 1. \tag{19}$$

By (11)

$$E\left[1+v(n+c_k m)\right]^{b_k} = E\left[\left(1+v(n+c_{-k}m)\right)^*\right]^{b_{-k}} = 1, \tag{20}$$

and $E\left[a_n(m)\right] = 1 - 1 = 0$.

Next we need to prove that, $a_n(m)$ meets the requirements of assumption A. Each moment of $a_n(m)$ is a finite sum of moments of $v(n)$. Since $v(n)$ is $O_m(1)$, any finite sum of its moments is bounded. This leads to the conclusion that $a_n(m)$ is also $O_m(1)$ and $a_n(m)$ meets the requirements of assumption A.

We will now prove that $a_n(m)$ also meets the requirements of assumption B. Denote

$$A_n(\omega)=\sum_{m=1}^{N}a_n(m)e^{-j\omega m}, \tag{21}$$

then

$$A_n^{(r)}(\omega_0)=(-j)^r\sum_{m=1}^{N}m^r a_n(m)e^{-j\omega_0 m}. \tag{22}$$

We will define $\tilde{a}_n(m)$ to be

$$\tilde{a}_n(m)=a_n(m)e^{-j\omega_0 m}. \tag{23}$$

Notice that $a_n(m)$ and $\tilde{a}_n(m)$ are probablistically equivalent. For the case $m\neq\ell$ we have

$$E\left[a_n(m)a_n(\ell)\right]=0 \tag{24}$$

and

$$\begin{aligned}E\left[\left|A_n^{(r)}(\omega_0)\right|^2\right]&=E\left[\left(\sum_{\ell=1}^{N}\ell^r\tilde{a}_n(\ell)\right)\left(\sum_{m=1}^{N}m^r\tilde{a}_n^{*}(m)\right)\right]\\&=\sum_{\ell=1}^{N}\sum_{m=1}^{N}\ell^r m^r E\left[\tilde{a}_n(\ell)\tilde{a}_n^{*}(m)\right]\\&=\sum_{\ell=1}^{N}\ell^{2r}E\left[\left|\tilde{a}_n(\ell)\right|^2\right]\end{aligned} \tag{25}$$

The sequence $\left\{\tilde{a}_n(\ell),1\le\ell\le N\right\}$ is zero mean i.i.d. Hence

$$E\left[\left|A_n^{(r)}(\omega_0)\right|^2\right]=\frac{E\left[\left|\tilde{a}_n(n)\right|^2\right]\ (N+1)^{2r+1}}{2r+1}+O\left((N+1)^{2r}\right). \tag{26}$$

Due to the fact that all moments of $\tilde{a}_n(m)$ are bounded $A_n^{(r)}(\omega_0)=O_m\left(N^{r+0.5}\right)$ and $\tilde{a}_n(m)$ meets assumption B. Now that we have proved that $\tilde{a}_n(m)$ meets all the requirements of (14), we can now apply (14). This yields the following error formula

$$\begin{aligned}\delta\omega&=12N^{-3}\Re\left\{A^{(1)}(\omega_0)-j0.5N\left[A^{(0)}(\omega_0)\right]^{*}\right\}+O_m\left(N^{-2}\right)\\&\approx 12N^{-3}\Im\left\{\sum_{n=1}^{N}(n-0.5N)a(n)\right\}\end{aligned} \tag{27}$$

The variance of the error is

$$\begin{aligned}E\left[\delta\omega^2\right]=E\Bigg[&144N^{-6}\Im\left\{\sum_{m=1}^{N}(m-o.5N)a_n(n)\right\}\\&\Im\left\{\sum_{m=1}^{N}(m-o.5N)a_n^{*}(m)\right\}\Bigg]\end{aligned} \tag{28}$$

We will use the equality

$$E\left[\Im\{x\}\Im\{y\}\right]=0.5\Re\left[E\left[xy^{*}\right]-E\left[xy\right]\right], \tag{29}$$

to get

$$\begin{aligned}E\left[\delta\omega^2\right]=72N^{-6}&\sum_{\ell=1}^{N}\sum_{m=1}^{N}(\ell-0.5N)(m-0.5N)\\&E\left[a_n(\ell)a_n^{*}(m)-a(\ell)a(m)\right]\end{aligned} \tag{30}$$

Since $a_n(m)$ is zero mean i.i.d, the cross-terms vanish and (30) becomes

$$\begin{aligned}E\left[\delta\omega^2\right]=72N^{-6}&\sum_{m=1}^{N}(m-0.5N)^2\\&E\left[a_n(m)a_n^{*}(m)-a_n(m)a_n(m)\right]\end{aligned} \tag{31}$$

In order to get a closed form error formula we need to develop an estimate for $E\left[a_n(m)a_n^{*}(m)\right]$ and $E\left[\left(a_n(m)\right)^2\right]$.

$$\begin{aligned}E\left[a_n(m)a_n^{*}(m)\right]=E\Bigg[&\left(\prod_{k=0}^{q/2}\left[1+v(n+c_k m)\right]^{b_k}\left[\left(1+v(n+c_{-k}m)\right)^{*}\right]^{b_{-k}}-1\right)\\&\left(\prod_{k=0}^{q/2}\left[\left(1+v(n+c_k m)\right)^{*}\right]^{b_k}\left[1+v(n+c_{-k}m)\right]^{b_{-k}}-1\right)\Bigg]\end{aligned} \tag{32}$$

Using (11) and (12), (32) simplifies to

$$\begin{aligned}E\left[a_n(m)a_n^{*}(m)\right]=&\prod_{k=0}^{q/2}\left(\sum_{i=0}^{b_k}\binom{b_k}{i}^2 i!\sigma^{2i}\right)\\&\left(\sum_{i=0}^{b_{-k}}\binom{b_{-k}}{i}^2 i!\sigma^{2i}\right)-1.\end{aligned} \tag{33}$$

Following a similar route, we can show that

$$E\left[a_n^{\,2}(n)\right]=0. \tag{34}$$

Inserting (34) and (33) in (30) leads to

$$E\left[\delta\omega^2\right]=72N^{-6}E_{aa'}\sum_{n=1}^{N}(n-0.5N)^2, \tag{35}$$

where

$$E_{aa'}=\prod_{k=0}^{q/2}\left(\sum_{i=0}^{b_k}\binom{b_k}{i}^2 i!\sigma^{2i}\right)\left(\sum_{i=0}^{b_{-k}}\binom{b_{-k}}{i}^2 i!\sigma^{2i}\right)-1. \tag{36}$$

For the selection $b_k=b_{-k}$, $E_{aa'}$ is

$$E_{aa'}=\prod_{k=0}^{q/2}\left(\sum_{i=0}^{b_k}\binom{b_k}{i}^2 i!\sigma^{2i}\right)^2-1. \tag{37}$$

It is easy to prove that

$$\sum_{n=1}^{N}(n+0.5N)^2=\frac{1}{12}N^3+O(N^2) \qquad (38)$$

Using (38), we can derive the following estimate of (35),

$$E\left[\delta\omega^2\right]\approx\frac{72}{12}N^{-3}E_{aa^*}=6N^{-3}E_{aa^*}. \qquad (39)$$

3. Comparison of the two algorithms

In this section, we compare the variance of the IF estimate of both algorithms. The comparison is done using the error formulas derived above. We will also compare the complexity of the two algorithms.

In Figure 1, we compare the variance of the IF estimates at the middle of the observation interval. We can see that the variance of the HAF estimate is consistently lower than that of the PWVD estimate.

In Figure 2, we compare the performance of the algorithms at points other than the midpoint of the interval. We can see that while the HAF algorithm performs quite well for a large part of the observation interval, the PWVD estimates degrade quite rapidly as we move away from the midpoint of the interval.

In Table 1, we compare the derived error formula for the PWVD to simulation results. We can see that the analytic estimates agree quite well with the simulations.

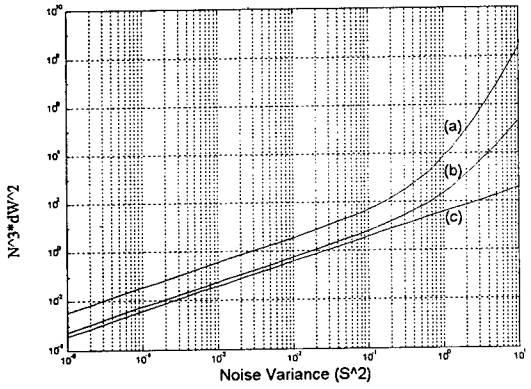


Figure 1- The variance of the IF, estimated in the middle of the sampling interval, as a function of input noise variance. Phase polynomial order = 4. (a) PWVD (b) HAF (c) CRB.

The computational complexity of the two algorithms is about the same as far as IF estimation is concerned. However, the PWVD requires a preliminary step of data resampling (interpolation), which increases its computational complexity. In addition, the HAF algorithm enables, after the polynomial parameters are computed, to estimate the IF at any desired point in the interval. The PWVD, on the other hand, requires repetition of the entire procedure for each new time point.

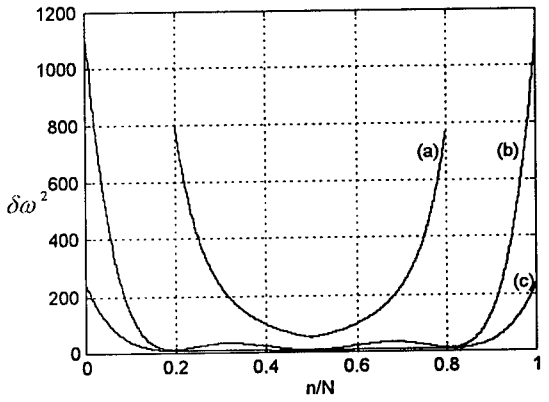


Figure 2- The error variance in the estimation of the IF as a function of n/N . σ^2=0.1; (a) PWVD (b) HAF (c) CRB.

Table 1- Comparison between the estimate of the error variance using (39) and simulation results, for various values of noise variance and for various window sizes.

N	$\sigma^2=0.1$	$\sigma^2=0.01$	$\sigma^2=0.001$
512	8.8612	0.6268	0.0531
1024	8.7624	0.6400	0.0514
2048	8.6891	0.5771	0.0589
Estimated	8.6391	0.6226	0.0602

References

[1] B. Porat, *Digital Processing of Random Signals*, Prentice-Hall, 1994.

[2] B. Boashash and P. O'Shea, "Polynomial Wigner-Ville Distributions and Their Relationship to Time-Varying Higher Order Spectra," *IEEE Trans. Signal Processing*, 42, pp. 216-220, 1994.

[3] B. Porat and B. Friedlander, "Asymptotic Statistical Analysis of the High-Order Ambiguity Function for Parameter Estimation of the Polynomial-Phase Signals," To appear in *IEEE Trans. Information Theory*.

Parameter Estimation of Multicomponent Polynomial-Phase Signals by Intersection of Signal Subspaces

Sergio Barbarossa

INFOCOM Dept. University of Rome "La Sapienza",Via Eudossiana 18, 00184 Rome, ITALY

e-mail: sergio@infocom.ing.uniroma1.it, Tel: (+39)6-44585-497, Fax: (+39)6-4873300

Abstract

The aim of this work is the parameter estimation of polynomial-phase signals (PPS) embedded in white noise. The proposed estimation algorithm is a generalization of the method based on the Polynomial-Phase Transform (PPT), able to solve an ambiguity problem appearing when applying the PPT to PPSs having the same highest order phase coefficients. The proposed approach is based on the intersection of two (or more) signal subspaces sharing only the useful components but not the undesired spurious harmonics. Different signal subspace are obtained by exploiting a multilag definition of the PPT.

1. Introduction

The aim of this work is the parameter estimation of multicomponent polynomial-phase signals (MC-PPS) embedded in additive white noise (AWN). The proposed approach is a generalization of the approaches proposed in [6], [7] (see also [9], Chapter 12). The case of MC-PPS was already analysed in [8] and [10] using the Polynomial-Phase Transform, introduced in [7], later called High order Ambiguity Function (HAF) [9]. The HAF allows the parameter estimation of multicomponent polynomial-phase signals, thus providing a clear advantage with respect to alternative techniques based on the computation of the instantaneous phase, followed by polynomial fitting [4], not able to deal with multicomponent signals. Of course, being nonlinear, the HAF suffers from the presence of cross-terms when applied to multicomponent signals. In general, as the number of samples increases, the effect of the cross terms diminishes. However when the signal components share the same highest order phase coefficients, the HAF exhibits spurious peaks that make the detection and parameter estimation ambiguous [1]. In such a case even the increase of the number of samples does not provide any help to remove the ambiguity. Since this situation is common to a number of applications where the polynomial-phase modelling can be of interest, like Synthetic Aperture Radar (SAR) signal processing [11] or communications over channels affected by multipath propagation [3], it is important to provide an accurate analysis of the ambiguity problem together with a possible solution. In this work we propose a solution based on an algebraic approach that exploits the redundancy of the Multi-Lag HAF (ML-HAF), introduced in [2]. The degrees of freedom related to the choice of the lags present in the ML-HAF can be exploited to solve the ambiguity problem. In this work we will show how to take advantage of these degrees of freedom using an algebraic approach based on the projection of the observed signal onto a signal subspace obtained as the intersection of signal subspaces estimated using different sets of lags.

The paper is organized as follows. In section 2 we will review the multilag HAF and the associated ambiguty problem. In section 3, we will describe the Signal-Subspace Intersection (SSI) method. Section 4 shows some performance obtained by simulation.

2. The multilag instantaneous high order moment and the ambiguity problem

Given a discrete-time signal $s(t)$, with $t = 0, \ldots, T-1$, its M-th order multi-lag High order Instantaneous Moment

0-8186-7576-4/96 $5.00

(ml-HIM) $s_M(t;\tau_1,\ldots,\tau_{M-1})$ of $s(t)$ is defined by the following iterations [2]:

$$s_1(t)=s(t), s_2(t;\tau_1)=s_1(t+\tau_1)s_1^*(t-\tau_1),\ldots$$
$$s_M(t;\tau_1,\ldots,\tau_{M-1})=s_{M-1}(t+\tau_{M-1};\tau_1,\ldots,\tau_{M-2})$$
$$\cdot s_{M-1}^*(t-\tau_{M-1};\tau_1,\ldots,\tau_{M-2}). \quad (1)$$

We use the acronym ml-HIM as a generalization of the HIM, introduced in [12]. The PPT [7] or HAF [9] is defined as the Fourier Transform of the ml-HIM, with respect to t, in the particular case in which the lags τ_k are all equal to each other. When applied to multicomponent polynomial-phase signals, the ml-HIM contains both *auto-terms* (the useful terms) and *cross-terms*. In particular, given an input signal:

$$s(t)=\sum_{k_1=1}^{K} A_{k_1} e^{j2\pi\sum_{m=0}^{M} a_{k_1,m}t^m/m!} \quad (2)$$

the ml-HIM satisfies the following properties (the proofs are given in [3]):

Prop. 1: The auto-terms are complex sinusoids whose frequencies are: $f_k=2^{M-1}\prod_{i=1}^{M-1}\tau_i a_{k,M}$, for $k=1,\cdots,K$.

Prop. 2: If the PPS components share the highest order polynomial-phase coefficients ($a_{k,l}=a_l$, for $l>m$ and $m>1$), the ml-HIM contains spurious harmonics, besides the useful harmonics given by the auto-terms. In particular, if the PPS components share all the phase coefficients from the second up to the M-th order, the ml-HIM contains *only* harmonics, whose frequencies are:

$$f_k=2^{M-1}\,M!\prod_{i=1}^{M-1}\tau_i\,a_M+\sum_{i=1}^{2^{M-2}}(a_{k_i,1}-a_{k_{2^{M-2}+i},1}). \quad (3)$$

Prop. 3: If the PPS components have the same highest order phase coefficients ($a_{k,l}=a_l$, for $k>m$ and $m>1$), the *only* sinusoids present in the ml-HIM, whose frequency is *proportional* to the product of all the lags, have a frequency $f_k=2^{M-1}\prod_{i=1}^{M-1}\tau_i a_{k,M}$.

From these properties, it is possible to envisage the parameter estimation technique. According to Prop. 1, if the input signal is a polynomial-phase signal of degree M, its M-th order ml-HIM is a sinusoid whose frequency is proportional to the highest order phase coefficient of the signal; the estimation of the highest order phase coefficeint can then be recast as a conventional frequency estimation problem which can be solved using FFT-based methods, as in [7], or using signal-subspace projection methods, as anticipated in [6]. Once the highest order coefficient has been estimated, the degree of the signal polynomial phase can be lowered by multiplying the input signal by $exp[-j2\pi\hat{a}_M t^M/M!]$, where $\hat{a}_M$ is the estimated coefficient. If the estimation is correct, the degree decreases and the process can be iterated to estimate the lower order coefficients. Indeed, Prop. 2 reveals the existence of spurious sinusoids, when the input signal is composed by PPSs having the same highest order phase coefficients. However, using Prop. 3 is possible to solve the ambiguity problem. In [1] the ambiguity was eliminated by multiplying the Fourier transforms (properly scaled) of the ml-HIMs corresponding to different sets of lags. In this work, we propose an algebraic approach, potentially able to provide better resolutions than the non-parametric FFT-based approach.

3. The Signal-Subspace Intersection Method

The freedom in the choice of the lags used for computing the ML-HIM can be properly exploited to remove the ambiguity. Let us compute the ML-HIMs $s_M(t;\tau_1^{(l)},\ldots,\tau_{M-1}^{(l)})$ corresponding to L different sets of lags $\tau_k^{(l)}$'s, for $l=1,2,\ldots,L$. We will assume that each pair of sets satisfies the condition: $\mathbf{c1}:\prod_{k=1}^{M-1}\tau_k^{(i)}=I/J\prod_{k=1}^{M-1}\tau_k^{(j)}$ with $i\neq j$ and where I and J are integer numbers. According to Prop.3, if the input signal contains K PPSs of degree M, the ml-HIMs contain sinusoidal auto-terms whose frequencies are related by the following relationship: $f_k^{(i)}=I\,f_k^{(j)}/J$, having indicated by $f_k^{(i)}$ the frequency of the k-th auto-term, $k=1,\ldots,K$, corresponding to the i-th set of lags, $i=1,\ldots,L$. To compare different ml-HIMs, is in general necessary to resample them. More specifically, to compare the generic i-th with the j-th ml-HIMs, the i-th ml-HIM has to be downsampled by a factor J, whereas the j-th ml-HIM has to be downsampled by a factor I. After downsampling, according to Prop.3, the two ml-HIMs share some sinusoids in common: *the sinusoids corresponding to the auto-terms*. To extract the information about the common sinusoids, we can then use an algebraic approach based on the estimation of the signal subspaces associated to different ml-HIMs and on their intersection. The algorithm, denoted Signal-Subspace In-

tersection (SSI) algorithm is the following (the algorithm is described in the case of two sets of lags, for simplicity, but the generalization to L sets is straightforward):

1. Compute the ml-HIMs $\mathbf{s}_M^{(i)}$, i=1, 2, for two different sets of lags satisfying condition **c1**, with given values of I and J;
2. If $I \neq J$, resample the ml-HIMs;
3. Estimate the covariance matrices $\mathbf{C}^{(i)}$ corresponding to the two ml-HIMs;
4. Compute the Singular Value Decomposition (SVD) of each covariance matrix: $\mathbf{C}^{(i)} = \mathbf{U}^{(i)}\mathbf{D}^{(i)}\mathbf{V}^{(i)^H}$;
5. Select the common order $\tilde{d}$ as the greatest value between the orders estimated from the singular values of each covariance matrix;
6. Estimate the signal subspaces $\mathcal{S}^{(i)}$ as the spaces spanned by the columns of the matrices $\mathbf{S}^{(i)}$ defined as
$$\mathbf{S}^{(i)} = [\mathbf{u}(1)^{(i)}, \ldots, \mathbf{u}(\tilde{d})^{(i)}] \qquad (4)$$
where $\mathbf{u}(k)^{(i)}$ denotes the k-th column of $\mathbf{U}^{(i)}$, for $i = 1, 2$;
7. Compute the intersection between the subspaces $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ (see [5]):
   (a) Compute $\mathbf{A} = \mathbf{S}^{(1)^H} \cdot \mathbf{S}^{(2)}$;
   (b) Compute the SVD of $\mathbf{A}$: $\mathbf{A} = \mathbf{Y}\,\mathbf{\Sigma}\,\mathbf{Z}^H$
   (c) Select order: $\mathbf{\Sigma}$ is a diagonal matrix whose entries are the cosines of the principal angles between the two subspaces $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ [5]. If we order the cosines in a decreasing order, the dimension d of the intersection subspace can be estimated as the index such that the following inequalities hold: $cos(\theta_1) = \ldots = cos(\theta_d) = 1 > cos(\theta_{d+1})$;
   (d) Estimate the intersection space as the space spanned by the matrix $\mathbf{E}$ defined as follows: $\mathbf{E} = \mathbf{S}^{(1)} \cdot [\mathbf{y}(1) \ldots \mathbf{y}(d)]$ where $\mathbf{y}(k)$ stands for the k-th column of the matrix $\mathbf{Y}$);
8. Estimate the pseudo-spectrum as the square norm of the projection of the steering vector $\mathbf{e}(\omega)$ onto the intersection subspace: $p(\omega) = \mathbf{e}^H(\omega)\ \mathbf{E}\ \mathbf{e}(\omega)$ where $\mathbf{e}(\omega) = (1, e^{j\omega}, \ldots, e^{j(N_f-1)\omega})$ and N_f is the number of samples on the frequency axis.

4 Performance

As a comparison term, Fig.s 1 and 2 show the HAF and the pseudo-spectrum, obtained using the SSI algorithm, of a signal composed by the sum of two cubic-phase signals having the same amplitude and phase parameters: $a_{1,1} = 0.125, a_{1,2} = 0.25/N, a_{1,3} = 0.25/N^2, a_{2,1} = 0.5, a_{2,2} = 0.5/N, a_{2,3} = a_{1,3}$; the number of samples is 1440; two sets of lags have been used: $\tau_1^{(1)} = \tau_2^{(1)} = 240$ and $\tau_1^{(2)} = 240, \tau_2^{(2)} = 120$ (so that $I = 2$ and $J = 1$). In Fig.1 we can clearly see three peaks, two of which (the lateral ones) are spurious peaks. Conversely, the pseudo-spectrum exhibits only one peak. In the presence of noise and dealing with finite length sequences, the estimation of the order d at step 7c) in the SSI algorithm can be done by using a threshold because the cosines of the principal angles are random variables themselves. From simulation (using 1000 Montecarlo runs), we have observed that, in the case of the signal analyzed in Fig.1, embedded in white noise (SNR=10 dB), the average values of $cos(\theta_1)$ and $cos(\theta_2)$ are 0.9991 and 0.7862, respectively, and the corresponding standard deviations are 5.7e-4 and 5.8e-2. Therefore the two random variables are well separable, for SNR=10 dB. The performance of the method in the presence of noise have been evaluated by computer simulations. In particular, Fig.3 shows the standard deviation of the estimate of the third order coefficient a_3 vs. the input SNR. The input signal is the same as the one analyzed in Fig.s 1 and 2, plus white noise. From Fig.3, we can see that the variance of the HAF-based method does not decrease as the SNR increases, due to the ambiguity problem.

5. Conclusion

In this paper we have proposed an estimation algorithm able to remove the ambiguity related to the HAF when applied to multi-component signals having the same highest order phase coefficients. The price paid by the proposed algorithm, besides the higher computational cost, is that the estimation variance is greater than that achievable with the HAF, if the ambiguity could be properly removed. The higher error is due to errors in the estimate of the covariance

matrices, errors which propagate in the estimation of the signal subspaces and, as a consequence, on the intersection subspace. Further analyses are necessary to optimize the SSI algorithm, in terms of size of the covariance matrix, number of intersections, all parameters that greatly affect the final performance. It is important to outline that the intersection idea could be extended to different signal processing problems, whenever is possible to set up different experiments where only the useful signals are in common.

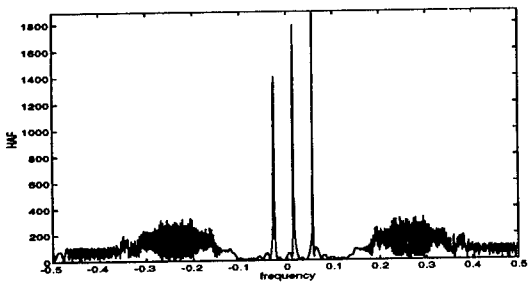


Figure 1: *HAF of the sum of two cubic phase signals.*

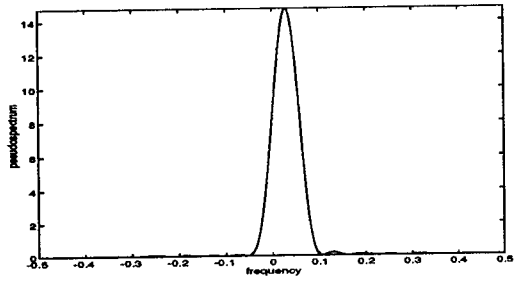


Figure 2: *SSI pseudo-spectrum of the same signal as in Fig.1.*

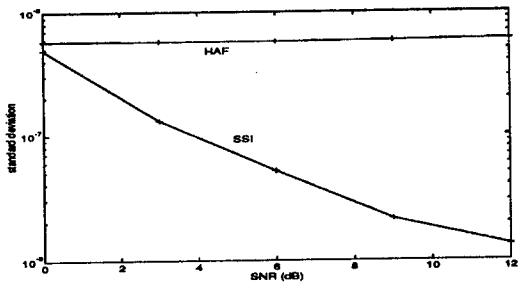


Figure 3: *Standard deviation of the estimate of third order phase coefficient using HAF and SSI algorithms.*

References

[1] S.Barbarossa, A.Porchia, A.Scaglione, "Multiplicative Multi-Lag Higher Order Ambiguity Function", *Proc. IEEE Conf. Acoustics, Speech, Signal Proc.*, Atlanta, GA, May 7-11, 1996.

[2] S. Barbarossa, G. Schiappa, "Analysis of Multicomponent Signals by Multilinear Time-Frequency Representations", *Proc. IEEE Conf. Acoustics, Speech, Signal Proc.*, Adelaide (Australia), April 1994.

[3] S.Barbarossa, A.Scaglione, G.B.Giannakis, "Multilag high-order ambiguity function for multicomponent polynomial-phase modelling", submitted to *IEEE Trans. on Signal Processing*, 1996.

[4] P. M. Djurić and S. M. Kay, "Parameter estimation of chirp signals," *IEEE Trans. on ASSP*, vol. 38, pp. 2118-2126, 1990.

[5] G.Golub, C.F. Van Loan, "Matrix Computations", The Johns Hopkins Univ. Press, Baltimore, 1983 (Section 12.4).

[6] R. Kumaresan and S. Verma, "On estimating the parameters of chirp signals using rank reduction techniques", *Proc. 21st Asilomar Conf. on Signals, Systems, and Computers*, pp. 555-558, Pacific Grove, CA, 1987.

[7] S. Peleg and B. Porat, "Estimation and classification of signals with polynomial phase," *IEEE Trans. on Information Theory*, vol. 37, pp. 422-430, 1991.

[8] S. Peleg, *Estimation and classification of signals with polynomial phase*, Ph.D. Thesis, Univ. of California, Davis, CA, 1993.

[9] B. Porat, *Digital Processing of Random Signals, Theory & Methods*, Prentice Hall, Englewood Cliffs, NJ, 1994.

[10] B.Porat, B.Friedlander, "Accuracy Analysis of Estimation Algorithms for Parameters of Multiple Polynomial-Phase Signals", *Proc. of Intl. Conf. on ASSP*, Detroit, May 1995.

[11] A.Porchia, S.Barbarossa, A.Scaglione, G.B.Giannakis, "Autofocusing techniques for SAR imaging based on the multilag high order ambiguity function", *Proc. IEEE Conf. Acoustics, Speech, Signal Proc.*, Atlanta, GA, May 7-11, 1996.

[12] G.Zhou, G.B. Giannakis, A.Swami, "On Polynomial Phase Signals with Time-Varying Amplitudes" to appear on *IEEE Trans. on Signal Processing* (Oct.94).

Nonlinear Time-Frequency Distributions with Multiplication-Free Kernels

A. Scaglione, S. Barbarossa, A. Porchia, G. Scarano
Infocom Dpt., Univ. of Rome "La Sapienza"
Via Eudossiana 18, 00184 Roma (ITALY)
sergio@ infocom.ing.uniroma1.it, gaetano@ infocom.ing.uniroma1.it

Abstract

In this paper we introduce and analyze the so called Complex Sign WVD (CS-WVD), defined as the Wigner-Ville Distribution (WVD) where one of the two signals is substituted by its complex sign. The substitution provides a consistent simplification for the implementation on dedicated hardware. In particular, the number of multiplications is drastically reduced. In spite of the hard nonlinearity used in the CS-WVD, the new transform is still able to deal with multi-component chirp signals. In the paper we provide a statistical analysis of the introduced transformation, in the case of polynomial-phase signals embebbed in additive white Gaussian noise. The theoretical analysis is compared to simulation results and to the Cramér-Rao lower bounds.

1 Introduction

Time-frequency distributions (TFD) such as the Wigner-Ville Distribution (WVD) are particularly suited for the analysis of Linear Frequency Modulation (LFM) signals [4], [1]. The WVD is known for its good localization properties, but it suffers from high cross-terms. A more general family of distributions, namely the Cohen's class [3], has been introduced for designing TFDs showing a good compromise between resolution and cross-terms [3]. Indeed, the members of Cohen's class can all be expressed as smoothed versions of the WVD. In this work we will concentrate on the WVD, but the proposed approach can be directly extended to the general case. One of the inconvenients related to the WVD is its higher computational cost with respect, for example, to the Short- Time Fourier Transform (STFT). In this paper we will introduce the Complex Sign-WVD (CS-WVD) aimed to drastically reduce the number of complex multiplications necessary to compute the WVD. The use of the complex sign introduces a performance loss, but gives rise to a transformation which is much easier to implement, especially on dedicated hardware, because it simplifies the scaling problem and does not require any multiplication for computing the transformation kernel (the product of the signals). The aim of this paper is the statistical analysis of a method for estimating the parameters of polynomial-phase signals, based on the CS-WVD. In spite of the hard nonlinearity introduced in the transformation, the proposed method is still able to deal with multicomponent signals, if the number of samples is sufficiently high. Indeed we will prove that the CS-WVD of LFM signals tends to coincide with the WVD of the same signals, as the number of samples increases. The paper is organized as follows. In Section 2 we will define and give some examples of the CS-WVD. We will also show the asymptotic properties of the CS-WVD. In Section 3 we will give a statistical analysis of the CS-WVD in the presence of additive white Gaussian noise (AWGN). Finally, in Section 4, we will analyze a parameter estimation method, for polynomial-phase signals, based on the CS-WVD.

2 Complex-Sign Wigner-Ville Distribution

In this section, we will introduce the CS-WVD and analyze its asymptotic properties, as the number of samples tends to infinity.

0-8186-7576-4/96 $5.00

2.1 Definition and examples

The Wigner-Ville Distribution of an infinite length signal $s(t)$ is defined as [3]:

$$WVD_s(t,f) := \lim_{T\to\infty}\frac{1}{T}\int_{-T}^{T} s(t+\tau)\overline{s}(t-\tau)e^{-j4\pi f\tau}d\tau \quad (1)$$

We define the Complex Sign-WVD (CS-WVD) of a signal $s(t)$ as:

$$CSW_s(t,f) := \lim_{T\to\infty}\frac{1}{T}\int_{-T}^{T} s(t+\tau)csign(\overline{s}(t-\tau))e^{-j4\pi f\tau}d\tau \quad (2)$$

where the complex sign of a complex variable z is defined as:

$$csign(z) \stackrel{def.}{=} \frac{\sqrt{2}}{2}sign(\Re(z)) + j\frac{\sqrt{2}}{2}sign(\Im(z)) \quad (3)$$

and the overbar denotes conjugation. The use of complex-sign makes the computation of the kernel multiplication free.

Indeed, we can prove the following theorem:

Theorem: Given a LFM signal $s(t) = Ae^{j\phi(t)} = Ae^{j2\pi(a_0+a_1t+a_2t^2)}$, its CS-WVD tends to be proportional to its WVD, as the number of samples tends to infinity:

$$\lim_{T\to\infty} CSW_s(t,f) = \frac{4}{\sqrt{2}A\pi}WVD_s(t,f) \quad (4)$$

Proof: Using the Fourier series expansion of the complex-sign of $Ae^{j\phi}$:

$$csign[Ae^{j\phi}] = \frac{1}{\sqrt{2}}\sum_{n=1}^{\infty}\frac{4}{n\pi}e^{jn\phi} \quad (5)$$

we can single out the first term, thus obtaining:

$$CSW_s(t,f) = \frac{2\sqrt{2}}{A\pi}WVD_s(t,f) + \quad (6)$$

$$+\lim_{T\to\infty}\frac{1}{T}\frac{\sqrt{2}}{2}\sum_{n=2}^{\infty}\frac{4}{n\pi}\int_{-T}^{T}Ae^{j(\phi(t+\tau)-n\phi(t-\tau)-4\pi f\tau)}d\tau$$

We can prove that the second term in (6) is null. In fact, for LFM signals all the integral arguments are quadratic-phase functions, whose second order coefficient is certainly different from zero, for $n > 1$. Since

$$\int_{-\infty}^{\infty} e^{-j\frac{\pi}{2}v^2}dv = 1 - j, \quad (7)$$

all the integrals give a finite result. Therefore, the limit in the second term of (6) is equal to zero.

[q.e.d.]

Examples of application of the CS-WVD to linear frequency modulation (LFM) signals are shown in Fig.1, for a monocomponent signal, and in Fig.2 for two LFM components. We can observe that, in spite of the hard non-linearity, the CS-WVD still allows the detection of chirp signals, even in the multicomponent case. This capability improves as the number of samples increases. Figs. 1 and 2 have been obtained using a number of samples $N = 128$.

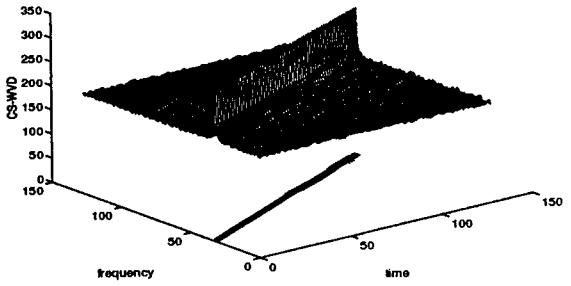


Figure 1: *CS-WVD of an LFM signal with* $a_1 = 0.25$ ***and*** $a_2 = 0.25/N$.

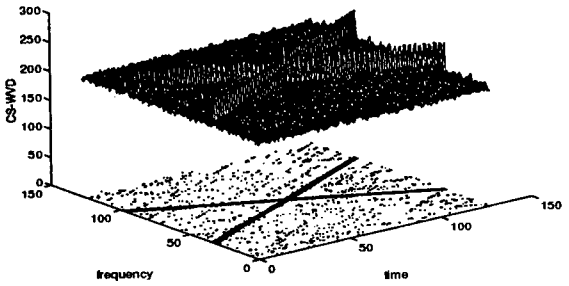


Figure 2: *CS-WVD of the sum of two LFM signals having* $a_{1,1} = 0.25$, $a_{2,1} = 0.75$, $a_{1,2} = 0.25/N$, ***and*** $a_{2,2} = -0.25/N$.

3 Statistical analysis

A statistical analysis of the CS-WVD of chirp signals, in the presence of Additive White Gaussian Noise (AWGN), has been carried out computing expected value and variance. In particular, referring to its discrete-time version and to the case of a limited number of samples, the CS-WVD takes this form:

$$CSW_x(n,f) = \sum_{k=1}^{N} x(n+k)csign(\overline{x}(n-k))e^{-j4\pi fk} \quad (8)$$

where $x(n) = s(n) + w(n)$, $s(n)$ is the useful signal and $w(n)$ is AWGN. The moments of the CS-WVD can be computed using the cumulant series expansion introduced in [5]. Under the hypothesis of Gaussian random variables, the expansion is greatly simplified because all the cumulants of order higher than two are equal to zero. In particular, the expected value of the CS-WVD is:

$$E\{CSW_x(n,f)\} = \sum_{k=1}^{N} s(n+k)\overline{R}(n-k)e^{-j4\pi fk} \quad (9)$$

and the second order moment is:

$$E\{|CSW_x(n,f)|^2\} = \quad (10)$$

$$\sum_k \sum_l \{s(n+k)\overline{s}(n+l)\overline{R}(n-k)R(n-l) + \sigma^2\delta(l-k) +$$
$$+\sigma\delta(l+k)s(n+k)S(n-k)R(n-l)$$
$$+\sigma^2\delta(l+k)s(n+l)\overline{R}(n-k)S(n-l)$$
$$+\sigma^2\delta(l+k)S(n-k)S(n-l)\}e^{-j4\pi f(k-l)}$$

where $Q(x)$, $R(n)$ and $S(n)$ are defined as follows:

$$S(n) = \frac{1}{2\sqrt{\pi}}(e^{-\frac{s_r^2(n)}{2\sigma^2}} + e^{-\frac{s_i^2(n)}{2\sigma^2}}) \quad (11)$$

$$R(n) = \frac{\sqrt{2}}{2}(1 - 2Q(\frac{s_r(n)}{\sigma})) + j\frac{\sqrt{2}}{2}(1 - 2Q(\frac{s_r(n)}{\sigma})) \quad (12)$$

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{-\infty} e^{-t^2/2}dt. \quad (13)$$

4 Parameter estimation

Given a LFM signal embedded in AWGN, we will now propose a method for estimating the phase parameters based on the CS-WVD. We initially define the transformation:

$$P_x^{(2)}(g,h) = \sum_n CSW_x(n, g+hn) \quad (14)$$

$$= \sum_n \sum_k x(n+k)csign(\overline{x}(n-k))e^{-j4\pi(g+hn)}$$

which provides a mapping from the input signal onto a plane whose axes are the signal mean frequency and sweep rate. For each chirp we observe a peak in the plane (g,h). Therefore detection and parameter estimation are carried out together: if a peak exceeds a suitable threshold, we decide for the presence of a chirp whose parameters are the peak's coordinates. The overall mapping was introduced in [1]. The method is asymptotically efficient and provides a good rejection capability in the presence of multicomponent signals. Its main disadvantage is the computational cost. Once again, this cost can be reduced by resorting to the complex-sign. This possibility was already proposed in [2], with a transformation called Hibrid-Nonlinear Integrated Generalized Ambiguity Function (HNL-IGAF). In [2] the performance were provided by simulation results; in this work we present a theoretical statistical analysis of the HNL-IGAF, based on the perturbation method, thus providing an analytical expression for the variances of both frequency and sweep-rate estimates, valid under the hypothesis of high SNR.

Given $x(n) = s(n) + w(n)$, where $w(n)$ is AWGN, denoting by δg and δh the estimation errors of frequency g_0 and sweep-rate h_0 respectively, for high SNR we have [6]:

$$\delta g \simeq \frac{Bv - Cu}{(DC - B^2)} \quad (15)$$

$$\delta h \simeq \frac{Bu - Dv}{(DC - B^2)} \quad (16)$$

where $B \stackrel{def.}{=} \frac{\partial^2 P_s^{(2)}(g,h)}{\partial g \partial h}|_{\substack{g_0 \\ h_0}}$, $C \stackrel{def.}{=} \frac{\partial^2 P_s^{(2)}(g,h)}{\partial h^2}|_{\substack{g_0 \\ h_0}}$, $D \stackrel{def.}{=} \frac{\partial^2 P_s^{(2)}(g,h)}{\partial g^2}|_{\substack{g_0 \\ h_0}}$ and

$$v = -4\pi \sum_n \sum_k knx(n+k)g(\overline{x}(n-k))e^{-j4\pi(g_0+h_0n)}$$

$$u = -4\pi \sum_n \sum_k kx(n+k)g(\overline{x}(n-k))e^{-j4\pi(g_0+h_0n)}.$$

The variance of the estimates of the two highest order coefficients are shown in Figs.3 and 4, from which we can observe the good asymptotic agreement between theoretical and simulation results for SNR values above a certain threshold ($\simeq 10dB$). We can also observe that the agreement between theory and simulation increases as the number of samples increases. As expected a saturation effect exists. The main price paid for using the complex

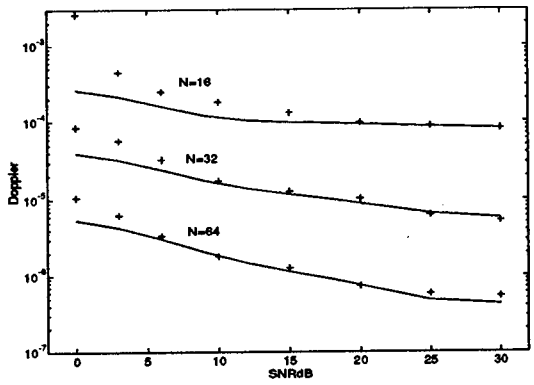


Figure 3: ***Variance of the estimations vs. SNR of the frequency g_0 obtained by using HNL-IGAF: theoretical analysis (solid line); simulation (+).***

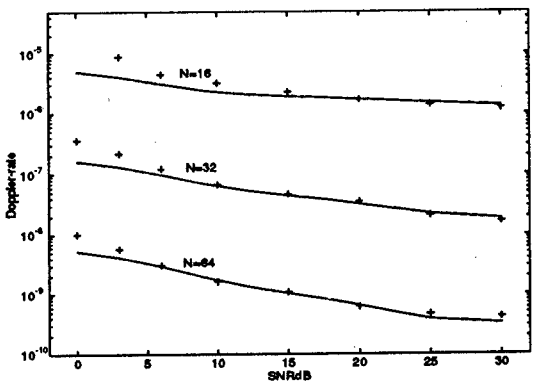


Figure 4: ***Variance of the estimations vs. SNR of the sweep-rate h_0, obtained by using HNL-IGAF: theoretical analysis (solid line); simulation (+).***

sign is a kind of saturation effect at high SNR. Indeed the variance of the estimation decreases more slowly than the inverse of SNR, as predicted by the Cramér-Rao Lower Bound (CRLB). As regards the dependence on the number of samples, the estimate of the m-th order polynomial phase coefficient is characterized by a CRLB that decreases as the inverse of N^{2m+1}. In our case this behavior if approximated only if N is sufficiently high.

The estimator based on the HNL-IGAF can be proved to be consistent. In fact, considering an infinite length LFM signal, we can prove (see [6] for the analytical details) that the expected value of the HNL-IGAF of an LFM embedded in AWGN tends to a Dirac pulse, centered on the signals parameters, and its variance tends to zero, as the number of samples tends to infinity.

5. Conclusions

In this work we have proposed and analyzed a nonlinear method for analyzing linear frequency modulation signals that presents some advantages for implementation on dedicated hardware, because it strongly reduces the number of complex multiplications necessary to compute the Wigner-Ville Distribution. The main price paid for the simplification is a performance loss. The method can be extended to the more general Cohen's class of time-frequency distributions as well as to the high order ambiguity functions.

References

[1] S.Barbarossa, "Analysis of Multicomponent LFM Signals by a Combined Wigner-Hough Transform", *IEEE Trans. on Signal Processing*, Vol.43, No.6, pp. 1511-1515, June 1995.

[2] S.Barbarossa G.Scarano "Analysis of Polinomial-Phase Signals by a fast Hybrid Nonlinear Generalized Ambiguity Function",*Proc. of IEEE Workshop on Higher Order Statistics*, Parador de Aguiblava, June 1995.

[3] L.Cohen, "Time-Frequency Distributions - A Review", *IEEE Proceedings*, Vol.77,pp. 941-981, July 1989.

[4] S.Kay, G.F.Boudreaux-Bartels, "On the Optimalitity of Wigner Distribution", in *proc. IEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, March 1985,pp.27.2.1-4.

[5] G.Scarano "Cumulant Series Expantion of Hibrid Nonlinear Moments of Complex Random Variables", *IEEE Trans. on Signal Processing*, Vol.39, No.4, April 1991.

[6] A. Scaglione, S. Barbarossa, A. Porchia, G. Scarano "Nonlinear Time-Frequency Distributions with Multiplication-Free Kernels",in preparation.

Adaptive Spectrogram for Time-Frequency Signal Analysis

Xiaobing Sun and Zheng Bao
The Laboratory for Radar Signal Processing, Xidian University
Xi'an, 710071, P. R. China
Tel: +86-29-8222381 Fax: +86-29-5263159

Abstract

A novel method for the determination of the window parameters of adaptive spectrogram is given in this paper. It is based on the detecting the maximum value and the width of the peak of the Radon transform of the modulus of the ambiguity function of the signals[1]. The proposed method effectively reduces the cross-terms and the noise of linear frequency modulated signals compared with Wigner distribution and the classical fix window spectrogram. A possible further extension of the method is also given to fit it to larger classes of signals.

I. Introduction

Time-frequency distributions (TFDs) is a powerful tool for detection and analysis of time-varying signals. They have found use in various fields such as radar, sonar, speech, biomedicine and geophysics. The classical methods such as Fourier transform can't provide a evident representation of the relation between the time and frequency content of the signals. Various time-frequency analysis methods have been proposed. The TFDs of Cohen's class is widely studied[1]. The most prominent methods among which are the Wigner distribution(WD) and the spectrogram (the squared magnitude of short-time Fourier transform, STFT). Although these two methods looks very different in their behaviors for analysis time-varying signals, they can be interpreted in the same point of view, i.e. we may consider the Wigner distribution as a special STFT which use the signal itself as the window function[2].

The cross-terms among multi-component signal is a severe limitation on the use of WD. It is a result when using a component of the signal as the window which is applied to the other components of the signal. This effect can easily be avoided by spectrogram which using only one window in any time[2]. But as pointed out in [3], the choice of the window dramatically affects the appearance of, or the signal concentration in, the spectrogram (or the STFT). We must face the tradeoff between time and frequency resolution of a preselected window.

Several authors have proposed to use the Gaussian function with variable length and obliquity which best matched to signal as the window of STFT[3]-[5]. It was shown that this method greatly improved the time-frequency concentration of time-varying signals. The proposed methods for determining the parameters of the window are either computationally expensive or need iterative. In this paper, we propose a method for determining the parameters of the window by making use of the Radon transform of the modulus of the ambiguity function(AF) of the signals. A systemic method for determining the parameters is given. The experimental results show this method greatly improved the time-frequency concentration of signals even in low signal to noise ratio(SNR) situation. A further extension of the method is also given in the fourth section of this paper.

II. Adaptive Window Parameters Determining Procedure

The classical spectrogram using a fixed low-pass

[1] This work is supported by the Institute of Electronics Researching of China ,grant no. J94.01.01-9461122 and DJ94.17.10-9571122.

0-8186-7576-4/96 $5.00 © 1996 IEEE

function as the window. It can be considered to be the order zero approximate for the time-varying signals since it use a constant frequency in the window to represent the time-varying frequency content of the signals. So, we must face the time-frequency tradeoff in the selection of the window length. Also, for some signals, the longer window does not means a better frequency resolution, since the frequency varying in the window may be large. If we allow the frequency content in the window varying linearly, i.e. using the linear frequency modulating(LFM) window instead of the constant frequency window as the basis to represent the signals, we can get a order one approximation of the time-varying signals. This certainly results a more precise representation of the frequency content of the signals.

The spectrogram of a signal x(t) is defined as:

$$Spec_x^g(t,\omega) = \left| \int x(\tau) g(t-\tau) e^{-j\omega\tau} d\tau \right|^2 \qquad (1)$$

where g(t) is the window of spectrogram (all integral are from $-\infty$ to ∞ unless otherwise stated). Here we choose the form of g(t) as:

$$g(t) = e^{-\alpha t^2 - j\beta t^2} \qquad (2)$$

where the parameter α controls the aspect ratio (or the length of the window), the parameter β controls the obliquity direction in the time-frequency plane (or the frequency modulating rate). These are the two parameters which we try to estimate adaptively for different signals. If α =0.5, β =0, then the WD of g(t) is a circle in the (t, ω) plane, where the unit of t is in second and ω in radian/sec/sec. Hence we consider these parameters defining a equal resolution window in (t, ω) plane.

As stated above, in any segment of the signals, we may better represent the signal using the window as shown in (2) than using the classical fixed low-pass window. But the condition to fulfill this goal is the correct estimation of the parameters α and β. In the following part of this section, we try to estimate these parameters adaptively from the AF of the signals. The AF of a signal x(t) is defined as:

$$AF_x(\nu,\tau) = \int x(t+\frac{\tau}{2}) x(t-\frac{\tau}{2}) e^{j\nu t} dt \qquad (3)$$

As shown in [6], the modulus of AF of any LFM signal is a line in the AF plane which traverses the origin. By calculating the Radon transform of the modulus of AF through the origin, the two dimensional function $AF_x(\nu,\tau)$ can be projected into a one dimensional function P(φ) [6] which is defined as:

$$P(\varphi) = \Re\left\{ \left| AF_x(\nu,\tau) \right| \right\} = \int \left| AF_x(r\sin\varphi, r\cos\varphi) \right| dr \qquad (4)$$

where $\Re$ represents the Radon transform, r is the radius and φ is the angle of the polar coordinate of the AF plane. The range of φ is $0 \le \varphi < \pi$.

We first assume the signal under analysis is of the same form as g(t) shown in (2), because by changing the parameters it can approximate a large classes of signals. It is not difficult to shown that the AF of the signal x(t) is :

$$AF_x(\nu,\tau) = \sqrt{\frac{\pi}{2\alpha}} e^{-\frac{4(\alpha^2+\beta^2)\tau^2 - 8\beta\nu\tau + \nu^2}{8\alpha}} \qquad (5)$$

Its Radon transform can be proved to be:

$$P(\varphi) = \frac{2\pi}{\sqrt{4(\alpha^2+\beta^2)\cos^2\varphi - 4\beta\sin\varphi\cos\varphi + \sin^2\varphi}} \qquad (6)$$

By differentiating $P(\varphi)$ respective to φ, it is also easy to find the value of φ_{max} which makes the maximum P(φ) and satisfies the relation:

$$\tan 2\varphi_{max} = \frac{4\beta}{1-4(\alpha^2+\beta^2)} \qquad (7)$$

For signals with evident time-varying frequency feature, the time duration of the signal must be long, therefor the α usually be small compared with β. So we may neglect α in (7). This results an estimation of β as:

$$\hat{\beta} \approx \frac{\tan\varphi_{max}}{2} \qquad (8)$$

So by finding the maximum direction of P(φ), we can estimate the parameter of β by (8).

We can farther use the -3dB width of the peak of P(φ) to estimate the parameter α. The physical mean of this estimation is based on the observation that the wider the peak of P(φ), the more different the signal from the LFM signal, so the shorter the window length should be. For a signal without frequency modulation (β =0 in (2)), it can be shown that: (for $\alpha << 0.5$, i.e. signal with a long duration)

$$\hat{\alpha} = \frac{\Delta\varphi_{3dB}}{4} \qquad (9)$$

where $\Delta\varphi_{3dB}$ represents the -3dB width of the peak of $P(\varphi)$. Although this estimation is achieved without frequency modulation and clearly different from the width of signal with frequency modulation, the experimental results show that it is still a better estimation.

The above analysis is on the uniform time-frequency plane, i. e. the unit of ν is in radn/sec and is in second. For some practical applications, the uniform coordinate may results φ_{max} to $\frac{\pi}{2}$, so a small error in estimation of φ_{max} will cause a grate difference in β. In this case, we should use the nonuniform coordinate system in (ν,τ) plane. Assuming the unit of ν is in k radn/sec and τ is in second, the β should be estimated as $\hat{\beta} \approx k\frac{\tan\varphi_{max}}{2}$.

III. Experimental Results

In this section we demonstrate the performance of the adaptive spectrogram using the parameters estimated by the methods which are stated in section II.

In the digital implement of the above procedure, the AF is calculated from discrete samples. In the calculating the Radon transform, the two dimension cubic interpolation is used to get the AF value not in the discrete grid of the rectangular coordinate system.

The first example examines the resolution advantage of adaptive spectrogram in analyzing two chirp signals with same frequency modulating rate. The envelopes of the signals are Gaussian functions. The adaptive spectrogram with both α and β estimated by (9) and (8) is shown in Fig.1(a). Fig. 1(b) shows the spectrogram with fixed α =0.5 and β estimated by (8). For comparison, Fig.1(c) and (d) show the WD and fixed window spectrogram(α =0.5, β =0) respectively. Please note the sampling rate of WD is two times fast than that of the spectrogram. It is evident that the adaptive spectrogram have nearly the same auto-component concentration as the WD but without its cross-terms. Compared with the classical fixed window spectrogram, the resolution of adaptive spectrogram is much higher.

In Fig.2, we show the same signal as in Fig.1 but now with additive noise. The signal to noise ratio (SNR) is 1dB. It is shown that the adaptive spectrogram still resolve the two signals clearly and almost without distortion, while the signal is totally embedded by noise in its WD and the signal is greatly distorted by classical spectrogram. The benefit of adaptive spectrogram comes from the fact that the window nearly plays the role of matched filter to the signals, so the noise effect is greatly reduced.

IV. Further Extension

One obvious limitation of adaptive spectrogram is that it most suits to chirp signals with nearly the same frequency modulating rate. A farther extension of this method is currently under investigation to make it suit to more large classes of signals. For examples, if two chirp signals with different frequency modulating rate is under analyze, this will results two different peaks in $P(\varphi)$. In this case, we may determine β by finding the weight center of $P(\varphi)$. This oblique direction can provide a good tradeoff between different requirements of signals.

It is also easy to extend the adaptive spectrogram to signals with vary frequency modulating rate. We can simply repeat use the adaptive spectrogram to each time segment of the signals.

V. Conclusions

In this paper, we proposed an adaptive spectrogram for analysis the time-varying signals. It is based on the Radon transform of the modulus of AF of signals to determine the window parameters. Adaptive spectrogram can yield excellent results over Wigner distribution and classical spectrogram, especially in the noise background. The conclusion is demonstrated by experimental results. Since the adaptive spectrogram can be considered as the order one approximation to the time-varying features of signals compared with order zero approximation of classical fixed window spectrogram and can be used to

any time segment of signals, it provides a more precise model to time-varying signals. A possible extension of the adaptive spectrogram for suit to larger classes of signals is also given.

Reference

[1]. L. Cohen "Time-frequency Distribution----A Review" Proc. IEEE. Vol.77, No.7, July, 1989, pp941- 981.

[2]. X. Sun and Z. Bao "On the Interpretation of Wigner Distribution by STFT" submitted to Int. Conf. on Signal Processing, Beijing, 1996.

[3]. D. L. Jones and T. W. Parks "A High Resolution Data-Adaptive Time-Frequency Representation" IEEE Trans on ASSP, Vol. 38, No.12, December 1989, pp2127-2135.

[4] D. Mihovilovic and R. N. Bracewell "Adaptive Chirplet Representation of Signals on Time-Frequency Plane" Electronics Letter, Vol.27, No.13, June, 1991, pp1159-1161.

[5] W. Kozek "Optimally Karhunen-Loeve-Like STFT Expansion of Nonstationary Processes" Int. Conf. on ASSP, 1993, IV428-431.

[6] B. Rstic and B. Boashash "Kernel Design for Time-Frequency Signal Analysis Using the Radon Transform" IEEE Trans on SP, Vol.41, No.5, May, 1993, pp1996-2008.

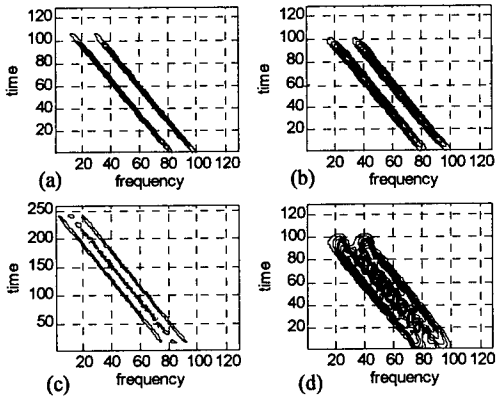


Figure1. (a) Adaptive spectrogram with both α and β estimated. (b). Adaptive spectrogram with α fixed. (c).Wigner distribution. (d). Fixed window spectrogram

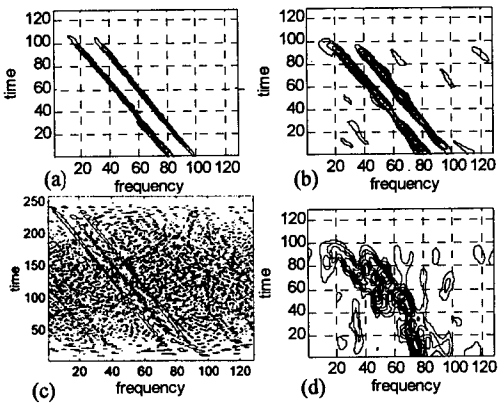


Figure 2: The same signal as in Fig. 1 but with addative noise. (a) Adaptive spectrogram with both α and β estimated. (b). Adaptive spectrogram with α fixed. (c).Wigner distribution. (d). Fixed window spectrogram

Plenary

Array Processing for Wireless Communications

B. Ottersten

Array Processing for Wireless Communications

Björn Ottersten*
Dept. of Signals, Sensors and Systems
Royal Institute of Technology
S–100 44 Stockholm, Sweden

Abstract

By equipping the base stations of a wireless network with antenna arrays, it is possible to more fully exploit the spatial dimension in a wireless communication system. Multiple antennas can provide a processing gain to increase the base station range and improve coverage. Also, by exploiting the spatial selectivity of an antenna array, interference may be reduced which in turn can be traded for increased capacity of the system. A wide range of wireless communication systems may benefit from spatial processing including high mobility cellular systems, low mobility short range systems, wireless local loop applications, satellite communications and wireless LAN. By employing an array of antennas, it is possible to multiplex channels in the spatial dimension just as in the frequency and time dimensions. This is often referred to as Spatial Division Multiple Access (SDMA). To increase system capacity, spatially selective reception as well as spatially selective transmission must be achieved. Herein, we present some different approaches and techniques for spatial/temporal processing. Critical aspects of SDMA for both high mobility cellular systems and low mobility or movable systems will be reviewed and the potential benefits examined.

1. Introduction

Wireless communications represent an important area of research, ultimately leading to the development of new and improved services and products. Substantial improvement in the capacity of these systems is a key issue as their use becomes more wide spread. The dramatic expansion of mobile communications over the last years has emphasized the importance of efficient use of frequency bandwidth. There is an increasing demand for capacity in wireless systems which traditionally directly translates into a demand for more bandwidth which is quite limited. Also, the infrastructure investment costs are often a limiting factor when deploying a new system aimed at wide area coverage. Increasing the range of current system is therefore also of great interest.

There are two critical factors in the design of wide area mobile communication systems, coverage and capacity. These factors have a direct impact on the cost and quality of the services since the spectral resources are limited and spectral efficiency is necessary. The spatial dimension is to a large extent unexplored in wireless systems. Traditional telecommunication schemes multiplex channels in frequency and/or time. However, the spatial dimension is in general used in a very rudimentary fashion by, for example, using some frequency channels in certain geographical areas (frequency planning) to limit interference. By incorporating antenna arrays and efficient spatial-temporal processing techniques into future systems, both the capacity and the range may be increased. With proper processing, it is possible to multiplex channels in the spatial dimension just as in the frequency and time dimensions. Spatially selective reception and transmission, can reduce interference in the system significantly allowing frequencies to be reused more often and thereby increasing capacity.

Each user has a unique spatial-temporal signature as seen by the base station. By identifying this signature for the user-to-base station communication link (up link), the signal of interest may be extracted from the noise while suppressing interference. Furthermore, with knowledge of the spatial-temporal signature describing the base station-to-user (down link) channel, transmission schemes may be devised which maximize the power of the signal of interest at the user while minimizing co channel interference and suppressing overall radiated power. This offers substantial capacity increases over current wireless system implementations.

In [2, 13, 28, 26, 31, 32], efficient use of the spatial dimension by employing antenna arrays at the base stations of wireless communication systems is explored. The up link problem has receive substantial attention [2, 26, 19, 20, 24] whereas the down link problem more recently has drawn interest [13, 28, 32]. Of course, the hardware requirements are

* Björn Ottersten is currently on sabbatical leave from the Royal Institute of Technology and is with ArrayComm Inc., San Jose, CA.

0-8186-7576-4/96 $5.00 © 1996 IEEE

more demanding when employing antenna arrays with multiple receivers and transmitters, but this permits a sparser infrastructure and will often be more cost effective. In general, increasing the range of cellular systems is of great interest initially, for example, when deploying the new PCS system in the United States. However, demand for increased system capacity is expected to follow shortly after adequate coverage is achieved in a successful system installation.

2. Modeling the Communications Channel

When attempting to exploit the spatial dimension, an additional independent parameter must be used when discussing channel models. The channel impulse response is now vector valued and will depend on the spatial distribution of the multipath propagation as well as the antenna array aperture and configuration. When designing a communications system, the channel model is critical. Given a description of the channel, efficient processing schemes may be devised and system performance can be analyzed.

There are low rank as well as high rank channel models and these concepts have impact on the spatial processing. The rank of the channel is also coupled to the concepts of narrow band and wide band signals which we use in the temporal domain. To exemplify this, consider a white noise sequence (which is wide band of course) arriving from broadside at a linear array. This is a spatial channel of rank one since the propagation for this case is described by a constant vector. However, as soon as the direction of the signal differs from broadside, the channel becomes high rank. The propagation of a perfect narrow band signal (sinusoid) is of course always described by a low rank channel. However, as the delay spread of the channel increases to the same order as the symbol time of a narrow band communications signal, the rank of the channel increases.

Another critical concept when discussing spatial processing is that of a parameterized array manifold. The array manifold is the collection of all array responses to a single point source over the parameter range (for example location) of interest. This is only a useful concept if the number of parameters and signals is small in relation to the number of antennas (for example the direction to the source for a fixed frequency) and the array response can be measured or modeled fairly accurately as a function of the parameters of interest. For example, near field scattering or mutual coupling at the array which is not calibrated or has a nice structure[1] is very difficult to model. The low rank signal model may still be quite useful even when it is not possible to parameterize the array response. In these cases, the response of the array or spatial-temporal *signature* characterizes a user. By noting this fact, the spatial dimension may be used to separate signals.

[1] An equi-spaced linear array with identical elements (uniform linear array) has a *nice* structure.

Below, we discuss some different channel models and also the use of a parameterized model. The concept of an array manifold can be modified to incorporate the special propagation environment often present in wireless communications. First, a simple low rank propagation model incorporating Rayleigh fading and directional information is described. This model is valid for narrow band signals and high base station antenna placement with little near field scattering. Second, a high rank channel model is described which is more suited for large time delay spreads and significant near field scattering at the array.

2.1. A Low Rank Channel Model

In [23, 31] a model of the flat fading due to local scattering is developed taking the spatial dimension into account. The array response is modeled as a stochastic vector which has a parameterized distribution. These parameters provide a useful description of the channel. The propagation between the mobile and the array is modeled as a superposition of a large number of rays originating from local scatterers in the vicinity of the mobile. We assume independent scattering, an angular distribution of the scatterers which is Gaussian (as seen from the array), and that the relative time delays for different propagation paths are small compared to the inverse of the bandwidth of the communication signal (small delay spread).

Assuming a uniform linear array with element spacing Δ in wavelengths, the signal received at the array may be modeled as

$$\mathbf{x}(t) = \mathbf{v}s(t) + \mathbf{n}(t) \quad (1)$$

$$\mathbf{v} \in N(0, \mathbf{R}(\theta, \sigma)) \quad (2)$$

$$\mathbf{R}(\theta, \sigma) \approx \mathbf{a}(\theta)\mathbf{a}^*(\theta) \odot \mathbf{B}(\theta, \sigma) \quad (3)$$

$$\mathbf{a}(\theta) = [1, e^{j2\pi\Delta \sin\theta}, \ldots, e^{j2\pi\Delta(m-1)\sin\theta}]^T \quad (4)$$

$$\{\mathbf{B}(\theta, \sigma)\}_{kl} = e^{-2[\pi\Delta(k-l)]^2\sigma^2\cos^2\theta} \quad (5)$$

where, $\mathbf{x}(t)$, is a complex valued $(m \times 1)$ vector, $s(t)$ is the complex envelop of the transmitted signal, $\mathbf{n}(t)$ is the additive noise, $\odot$ denotes element-wise multiplication, and $\mathbf{v}$ is the channel or spatial signature which is a complex, Gaussian random vector with a distribution function parameterized by the nominal direction to the mobile, θ, and the angular spread (standard deviation), σ, see Figure 1.

Equations (1-5) model the Rayleigh fading of the channel taking the spatial dimension into account. The vector $\mathbf{a}(\theta)$ is often termed the array response vector and represents the array output to a point source from direction θ. The angular spread, σ, is a critical parameter since this is a measure of the deviation from the point source or plane wave model. Frequency selective fading may be incorporated in this model by adding time delayed versions of the signal with different spatial characteristics. Also, interfering sources on the same frequency channel may easily be incorporated to the model.

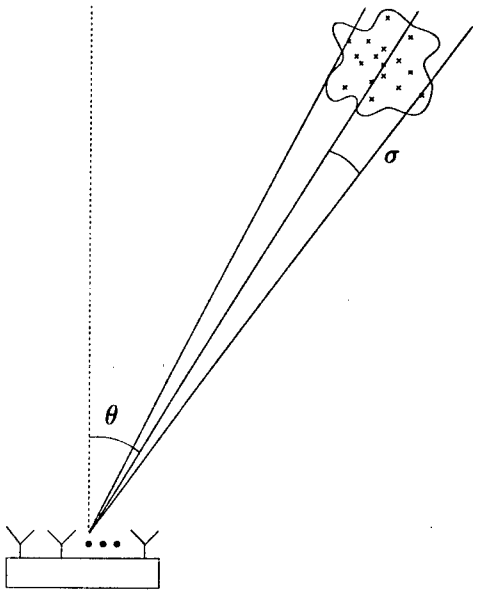


Figure 1. Geometry of the model characterizing the local scattering in the vicinity of the mobile.

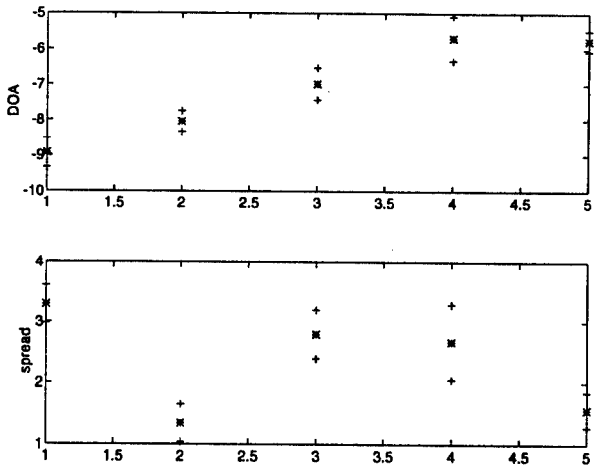


Figure 2. Estimated directions and angular spreads in degrees with standard deviations versus trial number.

2.1.1 Propagation Modeling and Data Experiments

The spatial channel model described above has been validated against experimental data collected by Ericsson Radio Systems. In the field experiments, a transmitter has been placed in urban areas with non line of sight approximately 1km from the receiving array which was elevated 30 meters above the ground [4]. The data has been processed to gain insight into propagation effects as well as into the behavior of some receiving algorithms. The standard deviation, σ, of the angular distribution is a critical parameter for SDMA systems, [32]. In [23, 31] the angular spread is found to be between two and six degrees in the experiments when the transmitter is placed 1km from the receiving array. In Figure 2, the estimated directions and angular spreads along with their associated standard deviations are displayed for a number of trials at one location.

The model above is only reasonable for small angular spreads. When the spread is large, which is the case in small cells (short range) or significant near field scattering at the array the spatial signature can not be parameterized by the direction.

2.2. A High Rank Channel Model

Below, a model is developed which is appropriate when there is a large delay spread among the multipaths and parameterization in terms of direction is not possible. This model is to some degree common to the so-called blind signal separation or channel estimation techniques. A paper by Tong et.al. [22] in 1991 sparked a great interest in the research community for blind channel estimation based on oversampled digital communications signals. A synchronously symbol sampled signal provides a sufficient statistic for detection, however, it does not allow the unique identification of the channel from second order statistics. Synchronization requires timing recovery and this is often achieved through oversampling in relation to the symbol rate of the signal. The oversampling may be achieved either in space or time and results in a cyclo-stationary process when viewed as a scalar process. However, if cast in an appropriate vector measurement model, the vector valued signal is stationary. Under certain identifiability conditions [16, 18], the channel may be consistently estimated from the second order statistics of the vector valued process.

Below, we will first view the oversampling as spatial, thereafter, oversampling in time will be introduced as well. By casting this model in an appropriate spatial-temporal vector form, the low rank nature of the signals is apparent.

Assume that a signal $s(t)$ is transmitted from a user, then the m element array output, $\mathbf{x}(t)$, is given by

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_m(t) \end{bmatrix} = h * s(t) + \mathbf{n}(t) = \mathbf{h}\mathbf{s}_L(t) + \mathbf{n}(t) \,, \tag{6}$$

where $\mathbf{n}(t)$ is the noise and

$$\mathbf{s}_L(t) = \begin{bmatrix} s(t) \\ s(t-1) \\ \vdots \\ s(t-(L-1)) \end{bmatrix} \quad (7)$$

$$\mathbf{h} = [\mathbf{h}(1), \mathbf{h}(2), \ldots, \mathbf{h}(L)] . \quad (8)$$

The impulse response of the channel, antenna elements, receiver and transmitter filters is modeled by $\mathbf{h}$ and will be termed simply the channel response. It will be modeled as finite with channel length L. If the channel is low rank, it may be modeled by a single complex vector ($L = 1$). However, if the propagation time of $s(t)$ across the aperture of the array is on the same order as the symbol time of the signal, a higher rank channel model must be used.

Remark

Note that in cases where a small number of dominant multipath signals with large time-delays are present, the low rank channel model is still useful. With appropriate spatial processing, the individual multipaths may be separated. Several options are available for combining the signal in the temporal domain [10], pre-detection, post-detection, soft-combining etc..

A high rank channel model is expected to be appropriate in environments with severe multipath and delay spread, for example, wireless local area networks with very high data rates.

2.3. Joint Spatial-Temporal Model

If $L < m$ the signal of interest is not full rank and subspace methods may be used to estimate the column space spanned by the channel $\mathbf{h}$ and determine the row space spanned by signal $\mathbf{s}_L(t)$, provided that $\mathbf{h}$ is rank L. In order to determine the actual signal, the finite alphabet property must be exploited, for example as in [21]. If $L \geq m$, the signal is no longer low rank in the spatial domain and joint spatial-temporal processing is required. This can be achieved by either oversampling or forming a sliding window or both.

Consider oversampling the received signal by a factor P and let the symbol time be 1. We have

$$\mathbf{x}(t + \frac{(i-1)}{P}) = \mathbf{h}^i \mathbf{s}_L(t) + \mathbf{n}^i(t) , \quad i = 1, \ldots, P . \quad (9)$$

The vector of all the oversampled antenna outputs is given by

$$\mathbf{x}^o(t) = \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{x}(t + \frac{1}{P}) \\ \vdots \\ \mathbf{x}(t + \frac{(P-1)}{P}) \end{bmatrix} = \mathbf{H}\mathbf{s}_L(t) + \mathbf{n}^o(t) , \quad (10)$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \vdots \\ \mathbf{h}^P \end{bmatrix} \quad \mathbf{n}^o(t) = \begin{bmatrix} \mathbf{n}(t) \\ \mathbf{n}(t + \frac{1}{P}) \\ \vdots \\ \mathbf{n}(t + \frac{(P-1)}{P}) \end{bmatrix} . \quad (11)$$

The channel matrix, $\mathbf{H}$, is $(mP \times L)$ and thus, if $mP < L$ the signal of interest will be confined to a low rank subspace of the joint mP-dimensional spatial and temporal measurement space.

Now, introduce a sliding window and form the vector

$$\mathbf{X}(t) = \begin{bmatrix} \mathbf{x}^o(t) \\ \mathbf{x}^o(t-1) \\ \vdots \\ \mathbf{x}^o(t-(M-1)) \end{bmatrix} = \mathcal{H}\mathbf{s}_{L+M-1}(t) + \mathbf{N}(t) , \quad (12)$$

where

$$\mathcal{H} = \begin{bmatrix} \mathbf{H} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{H} \end{bmatrix} \quad \mathbf{N}(t) = \begin{bmatrix} \mathbf{n}^o(t) \\ \mathbf{n}^o(t-1) \\ \vdots \\ \mathbf{n}^o(t-(M-1)) \end{bmatrix} . \quad (13)$$

The channel matrix, $\mathcal{H}$, is $(mPM \times (L+M-1))$ and has special structure due to the sliding window. In [6, 12, 17], it is noted that the channel matrix is linearly parameterized with respect to the channel coefficients lending itself to a two step subspace fitting approach to estimate the channel. Under appropriate identifiability conditions, [16, 18], the channel may be estimated up to a scaling and this in turn may be used to estimate the signals.

The model above is easily extended to multiple signals by letting $s(t)$ be a d-vector with the complex amplitudes of the d signals and $\mathbf{h}(k)$ is a $(m \times d)$ matrix of channel coefficients. Since the signal can be viewed as a low rank process in this measurement space, subspace based methods may be applied to separate the signals from the noise. Furthermore, since the span of the channel matrix may be identified, the influence of the channel may be remove from the signals. In the presence of multiple signals, temporal characteristics are require to estimate the individual signal sequences, Section 3.1.2.

3. Exploiting the Spatial Dimension

To achieve increased range in a wireless communication system, it may be argued that the mobile to base communication (up link) is the critical link. It is desirable that the mobiles operate at low powers and thus, for acquisition, the base stations must be able to detect weak signals of short duration in a noisy and possibly interfering environment. In

the down link (base to mobile communication), increased range may be achieved by for example increasing the transmit power.

To achieve increased system capacity by employing an array of antennas at the base stations, the frequency reuse distance may be decreased [4, 15] or the frequency channels may be reused with in a cell [32] (or a combination thereof). In both cases, the interference in the system induced by other users is of course increased. In the up link, this is manifested by the cross-talk problem. Mobiles operating on the same channel (frequency/time slot) with dramatically different signal amplitudes caused by, for example, fading are difficult to separate. It is difficult to adequately suppress the stronger signal when estimating the weaker signal resulting in cross-talk. In some sense the down link problem may be even more severe, especially in frequency division duplex (FDD) systems [32]. The fading caused by local scattering around the mobile (or the base station) is observable in the up link but unobservable in the down link due to the uncorrelatedness of the fading processes at the different frequencies. The up and down link channels are not reciprocal. The down link problem has received limited attention. In [15] a method is proposed which does not exploit directional information whereas in [32] a model based approach using this information is proposed.

3.1. Up Link Processing

When receiving communication signals at an antenna array, the proposed signal processing methods for distinguishing different messages, can be grouped in two main categories; those that exploit array response information and those that do not. Assuming the low rank channel model with small spread angle described above, it is possible to use direction estimation techniques which use array response information to separate signals. These methods will be referred to as using *directional information* and include techniques proposed in e.g. [32, 2, 19].

The other class of methods, makes few or no assumptions on the array response but rely on other properties for separating the signals.

3.1.1 Directional Information

Due to the local scattering, spatial signature represented by $\mathbf{v}$ does not belong to the array manifold, i.e.,

$$\mathbf{v} \neq \mathbf{a}(\theta), \quad \text{for any } \theta \,. \tag{14}$$

This may also be interpreted as the wavefront at the array not being planar. This may be interpreted as spatial diversity, i.e., the correlation between antenna elements decreases with distance, this is seen in the structure of the second moment of $\mathbf{v}$ in (3). The flat fading becomes less severe at the array as the diversity increases, i.e., σ increases. Techniques that make no use of directional information, e.g., [26] efficiently exploit this fact and perform better as the angular spread increases. Methods that are based on directional information, $\mathbf{a}(\theta)$, for estimating the signals [2, 14] will in general deteriorate as the angular spread becomes larger. These methods which are related to traditional *beamforming* techniques, are derived from a point source model. This behavior is not surprising since $\mathbf{v}$ will not correspond to an array response vector for any θ.

In [11], the directional error caused by local scattering is analyzed and characterized for different estimators which make use of array manifold information. The error is in general small and if the goal is to increase the range of a cellular system, this model error is not critical, [29]. However, the situation is quite different when attempting to host multiple mobiles on the same frequency channel. Even a small directional error can cause a significant degradation in the estimates of the signals. Since the array manifold vector in the nominal direction, θ, differs from the spatial signature, an error will be made when determining the copy vectors using the point source model. Consider the case when two signals are present and data is collected during a short period in time so that the users may be considered stationary. In a fading environment, the signal strengths of the two signals can be quite different. Thus, a small error in suppressing the stronger signal will cause a significant decrease in signal to interference and noise ratio (SINR) of the weaker signal. One way of improving the performance in these situations is to modify the array manifold model. The estimate of the signal subspace may be quite accurate and this information can be used to obtain an improved estimate of the spatial signatures.

We will provide a simple modification to the point source model which yields improved estimates of the signal waveforms. For small angular spreads, the spatial signature can be approximated as a linear combination of the array manifold vector and its derivative

$$\mathbf{v} \simeq \mathbf{a}(\theta_i, \rho_i) = \mathbf{a}(\theta_i) + \rho_i \mathbf{d}(\theta_i) \,, \quad \mathbf{d}(\theta_i) = \left.\frac{\mathrm{d}\mathbf{a}(\theta)}{\mathrm{d}\theta}\right|_{\theta=\theta_i} \tag{15}$$

This may be viewed as a generalized array manifold, $\mathbf{a}(\theta_i, \rho_i)$, parameterized by θ_i and ρ_i. Assume that estimates of θ and of the signal subspace, $\mathbf{E}_s$, have been obtained from the data. We may pose the problem of finding the $\boldsymbol{\rho}$ that provides the best fit between the signal subspace and the generalized manifold. This is a subspace fitting problem where $\boldsymbol{\rho}$ is a linear parameter in the manifold and can thus be solved for in a least squares sense. Let $\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho}) = [\mathbf{a}(\theta_1, \rho_1) \ldots \mathbf{a}(\theta_d, \rho_d)]$ where d is the number signals. The following minimization problem

$$\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} \mathrm{Tr}\{\mathbf{A}^*(\boldsymbol{\theta}, \boldsymbol{\rho})(\mathbf{I} - \mathbf{E}_s\mathbf{E}_s^*)\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho})\} \tag{16}$$

can be solved explicitly to provide and estimate of $\boldsymbol{\rho}$.

Consider the following example where two signals 50 dB and 30 dB (on average) above the spatially and temporally white noise are present. An 8 element uniform linear array is used, the nominal directions to the sources are 0° and 20° and assumed known, in each trial 100 snapshots are collected, and the signals are copied (estimated) using the so-called deterministic weight vectors,

$$\hat{\mathbf{s}}(t) = (\hat{\mathbf{A}}^*\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}^*\mathbf{x}(t)\ . \qquad (17)$$

In Figure 3, the average SINR over 500 independent noise and channel realizations[2] is displayed for the weaker signal as a function of the angular spread in degrees. For this case, the improvement obtained by using the generalized array manifold can be as much as 10 dB on average, much larger improvements may be obtained for certain channel realizations. The estimator above is quite straight forward, it is certainly possible to jointly estimate $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$. Alternative generalizations of the array manifold are also possible.

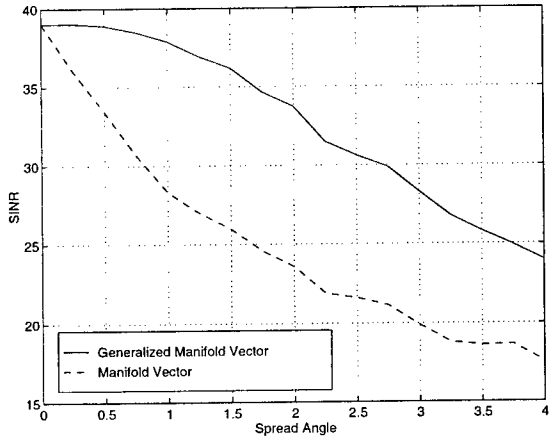


Figure 3. Signal to interference and noise ratio for the weakest signal as a function of angular spread.

The method described above makes no use of available *a priori* information on the source signals, for example, training and preamble sequences in digital communication systems are often present. In [7, 25], an approach is formulated which exploits both temporal and spatial information. In [3], it is shown how this method may be integrated with the Viterbi algorithm to perform accurate symbol detection in mobile communications using the GSM standard.

3.1.2 Non Directional Information

When attempting to separate multiple signals or suppress interference without making use of directional information, some temporal characteristics of the signal must be used. In [2, 26], a reference signal is assumed available which may be correlated with the array output to achieve signal separation. This reference signal may be a known training sequence, a known code sequence [13], or may be generated by feeding back decisions [8]. There are a number of methods that make use of the constant modulus property or finite alphabet of communication signals to separate them [1, 21]. In general, these methods are concerned with the low rank channel model.

As seen in the previous section, also in the case of a high rank channel, the signal is low rank for an appropriate measurement model. In [12, 18] the channel estimation problem for one signal in noise is cast in a subspace framework. Using subspace based methods, estimation algorithms are proposed and evaluated. In [6] the subspace based methods are analyzed and performance bounds are derived. Detecting the transmitted symbol sequence in the presence of multiple users is described in [24]. The row space spanned by the signals is first identified, this removes the effect of the channel (inter symbol interference). The individual transmitted signals may be separated by exploiting the finite alphabet property. In [21] the detection is achieved by alternatively making symbol decisions (from an estimated channel) and then estimating the channel (based on a known symbol sequence). A similar concept is described in [20] where initial symbol decisions are used to reconstruct a reference signal which in turn is used to estimate the channel and so on.

The problem of blind signal separation in this application is difficult although it may be relevant when there is large uncertainty surrounding the transmitted signal. In practical digital communication systems, known bit sequences are always present in some form to identify users and establish a communication link. Training sequences may be used to obtain an initial estimate of the channel and then a tracking mode takes over, updating the channel estimate based on the demodulated and possibly remodulated signal.

3.2. Down Link Processing

Consider the simple low rank channel model in (1) which may be used to characterize the down link spatial channel statistics as well. However, in most current frequency division duplex (FDD) systems the up and down link flat fading may be considered independent. If the main objective is increased range, this does not pose a major problem. However, the unobservable down link channel is one of the main obstacles if the intention is to also increase system capacity. An array could be employed at the mobile site as well, but in many applications this is not considered a feasible solution. Another alternative is to attempt to estimate the channel by employing feedback [5]. This requires a complete redesign of protocols and signaling and is probably only possible in environments which vary very slowly in time. This

[2]The channel realizations are normalized to have norm $\sqrt{8}$.

technique may be feasible for *movable* (rather than mobile) systems such as indoor wireless local area networks.

If we are attempting to increase capacity in current FDD systems in the down link, the information gained from the signal separation techniques in the up link, can not be used directly. Since the channels are not reciprocal, reusing an *optimal* weight vector obtained from receive data in the transmit mode is not advisable. One should at least attempt to *transform* the weights to the transmit frequency. However, this is not a well conditioned problem unless an array response model is introduced. When using an array model to transform weight vectors, directional information is exploited. In [27], the spatial signatures of the users are first estimated using temporal information, then the directions may be extracted by applying a parameterized array manifold. It should be noted that in [15], a transmit scheme is proposed which does not use directional information. The down link scheme is based on statistical information estimated in the up link to take into account the unobservable fading. However, the frequency duplex distance is not compensated for causing the system to degrade in the presence of line of sight propagation.

In time division duplex (TDD) systems, the up and down link channels can be considered reciprocal if there is limited movement between receive and transmit. Up link channel information may then be used to achieve spatially selective transmission and thus increasing capacity [10]. When the channel is high rank, combined spatial-temporal processing may be applied on the down link to increase capacity. The estimated up link channel may be inverted in the sense that the signals are appropriately pre-equalized and spatially multiplexed at the base station to minimize inter symbol and co-channel interference at the users, [9].

The efficient use of the spatial dimension in current FDD cellular systems with high mobility requires the use of directional information. Array response modeling is feasible for medium to large size cells with high placement of the base station antennas avoiding near field scattering.

4 Down Link Capacity

As argued previously, the down link is likely to be the limiting factor when increasing the capacity of cellular systems. This is mainly due to the independence between the up and down link channels when FDD is employed. There are two main approaches for increasing capacity with antenna arrays. The frequency reuse distance may be decreased or multiple mobiles may be allocated to the same cell (or some combination of the above). In [32], the down link capacity problem is studied for FDD systems and a transmission scheme is proposed based on channel information estimated on the up link. It is shown that when inter cell nulling is not employed, multiple mobiles per cell is in general a more efficient way of increasing capacity. Also, capacity depends

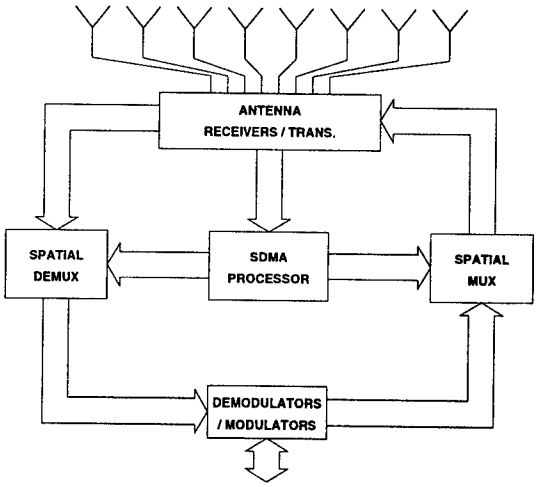


Figure 4. Possible configuration of an SDMA system.

to a large degree on the spread angle of the mobiles. In [30], a down link system using inter cell nulling and slow up link power control is studied. In this case, reduced cluster size provides the largest capacity increases. Note however that this requires power control with good dynamic range and direction estimation to users in neighboring cells. Also, inter cell nulling may be quite difficult even in a synchronous TDMA system when propagation delays are significant. However, reducing the frequency reuse distance in conjunction with frequency hoping and dynamic channel allocation could reduce the requirements on frequency planning.

5 Summary

Providing adequate coverage and sufficient capacity are two challenging problems for wireless communication systems. Antenna arrays at the base stations of cellular systems can increase range compared to current systems. The capacity problem can be significantly mitigated by *spatial division multiple access (SDMA)* techniques. SDMA supports multiple connections on a single conventional channel, based on spatial reception and transmission schemes and/or decreased frequency reuse distance by reducing and rejecting interference. Thus, capacity may be increased over current wireless system implementations.

References

[1] B. Agee, A. Schell, and W. Gardner. "Spectral Self-Coherence Restoral: A New Approach to Blind Adaptive Signal Extraction Using Antenna Arrays". *Proc. IEEE*, 78:753–767, Apr. 1990.

[2] S. Andersson, M. Millnert, M. Viberg, and B. Wahlberg. "An Adaptive Array for Mobile Communication Systems". *IEEE Trans. on Veh. Tec.*, 40(1):230–236, 1991.

[3] D. Asztély. On antenna arrays in mobile communication systems, fast fading and GSM base station receiver algorithms. Technical Report IR-S3-SB-9603, Signals, Sensors and Systems, March 1996.

[4] U. Forssén, J. Karlsson, B. Johannisson, M. Almgren, F. Lotse, and F. Kronestedt. "Adaptive Antenna Arrays for GSM900/DCS1800". In *Proc. IEEE Veh. Technol. Conf.*, pages 605–609, 1994.

[5] D. Gerlach and A. Paulraj. Adaptive transmitting antenna arrays with feedback. *IEEE SP Letters*, 1(10):150–152, October 1994.

[6] M. Kristensson and B. Ottersten. "Asymptotic Analysis of a Subspace Method for Blind Channel Identification". In *Proc. ICASSP 96*, Atlanta, May 1996.

[7] J. Li, B. Halder, P. Stoica, and M. Viberg. Computationally efficient angle estimation for signals with known waveforms. *IEEE Transactions on Signal Processing*, 43:2154–2163, 1995.

[8] E. Lindskog, A. Ahlén, and M. Sternad. "Combined spatial and temporal equalization using an adaptive antenna array and a decision feedback equalization scheme". In *IEEE International Conference on Acoustics, Speech and Signal Processing*, Detroit, USA, May 1995.

[9] H. Liu and G. Xu. "Multiuser Blind Channel Estimation and Spatial Channel Pre-Equalization". In *Proc. ICASSP'95*, pages 1756–1759, Detroit, MI, May 1995.

[10] P. Mogensen, F. Frederiksen, J. Wigard, and S. Petersen. "A Research Study of Antenna Diversity and Data Receivers for DECT". In *Proc. Nordic Radio Symposium*, Saltsjöbaden, Sweden, April 1995.

[11] R. L. Moses, T. Söderström, and J. Sorelius. Effects of multipath-induced angular spread on direction of arrival estimators in array signal processing. In *IEEE/IEE Workshop on Signal Processing Methods in Multipath Environments*, Glasgow, Scotland, Apr. 1995.

[12] E. Moulines, P. Duhamel, J.-F. cois Cardoso, and S. Mayrargue. Subspace methods for the blind identification of multichannel fir filters. *IEEE Transactions on Signal Processing*, 43(2):516–525, February 1995.

[13] A. Naguib, A. Paulraj, and T. Kailath. "Capacity Improvement with Base-Station Antenna Arrays in Cellular CDMA". *IEEE Trans. Vehicular Technology*, 43(3):691–698, Aug. 1994.

[14] B. Ottersten, R. Roy, and T. Kailath. "Signal Waveform Estimation in Sensor Array Processing". In *Proc. 23rd Asilomar Conf. Sig., Syst.,Comput.*, pages 787–791, Nov. 1989.

[15] G. Raleigh, S. Diggavi, V. Jones, and A. Paulraj. A blind adaptive transmit antenna algorithm for wireless communication. In *Proc. ICC*, 1995.

[16] D. Slock. "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction". In *Proc. ICASSP 94*, volume 4, pages 585–588, Adelaide, Australia, April 1994.

[17] D. Slock and C. Papadias. Blind fractionally-spaced equalization based on cyclostationarity. In *Vehicular Technology Conference*, volume 2, pages 1286–1290. IEEE, June 1994.

[18] D. Slock and C. Papadias. Further results on blind identification and equalization of multiple fir channels. In *ICASSP'95*, volume 3, pages 1964–1967, Detroit, Michigan U.S.A., May 1995. IEEE.

[19] S. Swales, M. Beach, D. Edwards, and J. McGeehan. The performance enhancement of multibeam adaptive base-station antennas for cellular land mobile radio systems. *IEEE Trans. Vehicular Technology*, 39:56–67, Feb. 1990.

[20] A. Swindlehurst, S. Daas, and J. Yang. "Analysis of a Decision Directed Beamformer. *IEEE Trans. on Sig. Proc.*, 43(12):2920–2927, December 1995.

[21] S. Talwar, M. Viberg, and A. Paulraj. "Blind Estimation of Multiple Co-Channel Digital Signals Using an Antenna Array". *IEEE SP Letters*, 1:29–31, Feb. 1994.

[22] L. Tong, G. Xu, and T. Kailath. A new approach to blind identification and equalization of multipath channels. In *Proceedings of the 25th Asilomar Conference on Signals, Systems and Computers*, pages 856–860, Pacific Grove, CA, November 1991.

[23] T. Trump and B. Ottersten. Estimation of nominal direction of arrival and angular spread using an array of sensors. To appear in Signal Processing, Elsevier, 1994.

[24] A. van der Veen, S. Talwar, and A. Paulraj. Blind estimation of multiple digital signals transmitted over FIR channels. *IEEE SP Letters*, 2(5):99–102, May 1995.

[25] M. Viberg, P. Stoica, and B. Ottersten. Maximum likelihood array processing in spatially correlated noise fields using parametrized signals. *Submitted to IEEE Trans. on Signal Processing*, 1994.

[26] J. Winters. Optimum combining in digital mobile radio with cochannel interference. *IEEE Trans. Vehicular Technology*, 33(3):144–155, August 1984.

[27] G. Xu and H. Liu. "An Effective Transmission Beamforming Scheme for Frequency-Division-Duplex Digital Wireless Communication Systems". In *Proc. ICASSP'95*, pages 1729–1732, Detroit, MI, May 1995.

[28] G. Xu, H. Liu, W. Vogel, H. Lin, S. Jeng, and G. Torrence. "Experimental Studies of Space-Division-Multiple-Access Schemes for Spectral Efficient Wireless Communications". In *Proc. ICC'94*, New Orleans, LA, May 1994.

[29] J. Yang and A. Swindlehurst. "Maximum SINR Beamforming for Correlated Sources". In *Proc. ICASSP 95*, pages 1916–1919, Detroit, MI, May 1995.

[30] P. Zetterberg. A comparison of two systems for down link communication with antenna arrays at the base. In *IEEE Wireless Communications Systems Symposium*, November 1995.

[31] P. Zetterberg. Mobile communication with base station antenna arrays: Propagation modeling and system capacity. Licentiate thesis TRITA-SB-9502, Signals, Sensors & Systems, February 1995.

[32] P. Zetterberg and B. Ottersten. The spectrum efficiency of a base-station antenna array system for spatially selective transmission. *IEEE Transactions on Vehicular Technology*, 44(3):651–660, August 1995.

WA-1: Signal Processing for Communications II

New Multiple MPSK Trellis Codes for the Mobile Radio Channel and their Complexity

Bernhard J. Mayr, Hans Weinrichter
Vienna University of Techology
Department of Telecommunications
Gußhausstraße 25/389, A-1040 Vienna, Austria
bmayr@email.tuwien.ac.at

Abstract

The characteristics of the mobile radio channel vary between the features of an AWGN-channel and those of a Rayleigh fading channel.

The so far known trellis codes do not meet the requirements of both these channels, because they are either adapted to the AWGN or the Rayleigh fading channel.
In this paper[1] we present new multiple trellis codes, which are well suited for the use in the AWGN-channel and the Rayleigh channel and are thus especially suited for the mobile radio channel.

A new measure of complexity is introduced which allows a fair comparison between multiple trellis codes of different dimensionalities. It is based on the number of algebraic operations per decoded information bit.

1. Introduction

Since their invention in the early 1980ies [1] trellis codes have been used exhaustively to gain noise immunity in communication systems. Trellis codes can be applied in systems with high information rates and $R_b > 1$ information bit per channel symbol, whereas the classical binary block and convolutional codes are constrained to rates $R_b < 1$.

The basic idea of trellis codes is simple: instead of transmitting redundant bits as in the case of block and convolutional codes the size of the transmission alphabet M is increased such that $M > 2^{R_b}$, i.e. there are more symbols for transmission as actually needed.
The redundancy of the increased symbol alphabet is exploited in a way that at each time slot t_k only a subset of all possible symbols is allowed for transmission. The valid symbol sequences are created by the actual information bits $b_{k,\nu}$ at time t_k and the contents of a finite state machine. The state at time slot t_k is defined by the values of the previous information bits $b_{k-1,\nu}, \cdots, b_{k-n,\nu}$ at times $t_{k-1}, \cdots, t_{k-n}$. The remaining task is to define some criterion how to select the allowed symbol sequences and to find appropriate finite state machines.

For the AWGN-channel the problem was solved by G. Ungerböck [1] - [3], who actually invented trellis coded modulation (TCM). In [1] - [3] MPSK- and MQAM-alphabets are used. The optimization criterion for trellis codes in the AWGN-channel is the squared Euclidean distance (Δ^2_{free}) between allowed coded symbol sequences since the error performance at high values of the SNR is lower bounded by:

$$P_b > Q\left(\frac{\Delta_{free}}{\sqrt{2N_0}}\right), \tag{1}$$

where $N_0/2$ is the two-sided spectral density of the additive white Gaussian noise and $Q(\cdot)$ the complementary Gaussian distribution function.
In [4] the idea of trellis codes for the AWGN-channel is extended to multiple symbols, i.e. $\tilde{k}$-tuples of MPSK- or MQAM-symbols are assigned to the trellis branches.

For the Rayleigh fading channel the TCM-problem was solved by D. Divsalar et al. [5],[6]. It turned out that the minimum number of distinct symbols (L_{eff}) between any two coded symbol sequences has to be maximized for minimizing the error rate:

$$P_b > \frac{1}{\beta^2 \cdot \mathrm{SNR}^{L_{eff}}}. \tag{2}$$

The effective length (L_{eff}) dominates the slope of the bit error rate curve as a function of SNR. The parameter β^2 is the so called product distance, i.e. the product of all non-zero Euclidean branch distances along any two trellis paths. It has to be maximized as well. Conventional trellis codes with one symbol per trellis branch and trellis codes with more than

[1]This work was sponsored by the Austrian Science Foundation (FWF), grant 10294 ÖPY.

one symbol per trellis branch (multiple trellis codes), which are especially designed for the Rayleigh fading channel, can be found in [5] - [10].

It can be seen from eqn.(2) that on one hand the minimum squared (free) Euclidean distance Δ^2_{free} is of no primary importance for the error performance in the Rayleigh fading channel and on the other hand the effective length and product distance do not show up in eqn.(1) for the BER of the AWGN-channel. Hence the optimum trellis codes for the two channels are obtained by different optimization criterions. As a consequence trellis codes for the AWGN-channel do not perform well in the Rayleigh fading channel and vice versa.

The mobile radio channel - as met in wireless communications - has been of growing economic interest in the recent years. Any communication system has to cope with short and long term time variant channel characteristics. The short term fluctuations caused by multipath propagation result in deep fades of the signal power. These deep fades can be combated against with equalization, interleaving, and coding techniques. However, there are also long term fluctuations in the characteristics of the mobile radio channel depending on the actual position of the base and the mobile station: in urban areas the signal is affected by multipath propagation and thus the fading channel is Rayleigh like. In rural areas there are only a few scatterers, hence the channel is similar to an AWGN-channel.
Therefore it is clear that none of the so far known trellis codes are suited for the mobile radio channel which requires trellis codes optimized for the AWGN-channel *and* the Rayleigh fading channel simultaneously.

The rest of the paper is organized as follows: In section 2 we will present new construction principles for trellis codes which are well suited for the use in the AWGN-channel and the Rayleigh fading channel. Finally, in section 3 a new definition of trellis complexity will be given.

2. Code construction

Any code that performs well in the AWGN-channel and the Rayleigh fading channel will also perform well in the mobile radio channel which fluctuates between AWGN and Rayleigh characteristics. The optimization criterion for our new trellis codes is to maximize the three parameters $L_{eff}, \Delta^2_{free}, \beta^2$ simultaneously. Our results are based on three new supports:

- The construction rule for the MPSK-subset decomposition which give optimum trellis codes for fading channels presented in [6] have been rearranged. The new rules can be interpreted as a non linear labeling method of the MPSK-signals, in contrary to the linear labeling method presented in [6].

  The new subset decomposition results in an increased free distance Δ^2_{free} and an increased product distance β^2 compared to the so far known codes.

- A completely new mapping of multiple MPSK-symbols to trellis branches increases the effective length L_{eff} and the free distance Δ^2_{free}. The idea behind the new method is to optimize the distances between emerging *and* the distances between reemerging branches of all trellis states (in the so far known codes only the distances between emerging branches are optimized).

- A new bit-to-symbol mapping helps to minimize the number of bit errors per error event. Binary information sequences with a small Hamming distance correspond to coded symbol sequences with a small Euclidean distance and a short effective length.

- Sometimes this three supports are supplemented with a fourth action: instead of doubling the symbol alphabet needed for uncoded transmission as applied in all so far known systems (cf. [1]), a fourfold larger alphabet is used which allows to increase the effective length and minimizes the number of nearest neighbor sequences.

The multiple MPSK trellis codes constructed according to these specified rules outperform the so far known codes. In the AWGN-channel the average gain of SNR is approx. 1 dB at $P_b = 10^{-5}$, in the Rayleigh fading channel it is approx. 3 dB. In fig.(1) the simulation results of two trellis codes are shown. One has been constructed according to our new rules and the other one according to the rules given in [6]. In both cases 4D-QPSK symbols (pairs of QPSK-symbols) are used for transmission at rate $R_b = 1$ information bit per QPSK-symbol. The number of trellis states equals $s = 4$. The code parameters are: $L_{eff} = 4$, $\Delta^2_{free} = 10$, $\beta^2 = 32$ for the new code and $L_{eff} = 3$, $\Delta^2_{free} = 6$, $\beta^2 = 8$ for the other code. It can be seen in fig.(1) that our new code clearly outperforms the other code: at $P_b \approx 10^{-5}$ the coding gain equals 2.5 dB in the AWGN-channel and more than 4 dB in the Rayleigh fading channel.
A more detailed description of our new construction rules and a complete list of all new codes can be found in [11].

3. Code complexity

Complexity is an important aspect of trellis codes. All so far given definitions of TCM-complexity consider the connectivity of the trellis graph regardless whether the connections contain parallel transitions or not. Following the notation of G. Ungerböck [3] this complexity can be written

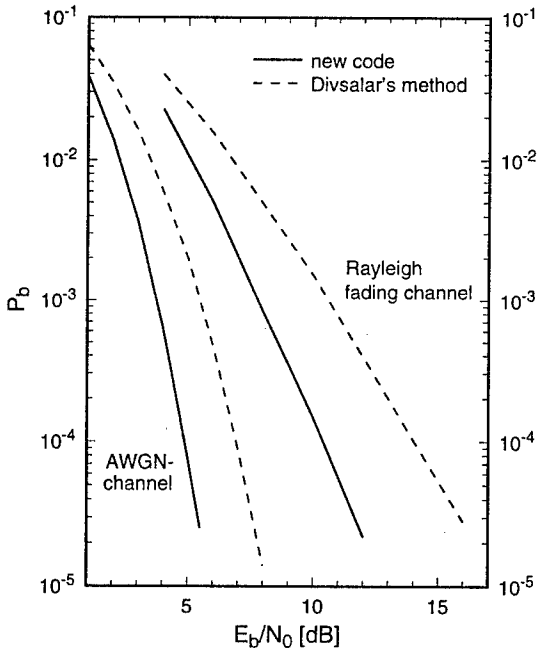


Figure 1. Simulation results of two 4D-QPSK trellis codes with four states in the AWGN-channel and the Rayleigh fading channel.

as:

$$K = \frac{s \cdot 2^{\tilde{n}}}{\tilde{k}}, \tag{3}$$

where s is the number of trellis states, $\tilde{n}$ the number of coded bits per transition, and $\tilde{k}$ the number of channel symbols per transition.

In contrast to this definition we propose a more practical definition of TCM-complexity which emphasizes the computational effort of the decoding process. The newly defined trellis complexity K_B measures the number of algebraic operations per decoded information bit. It can be written as:

$$K_B = \frac{\tilde{k} \cdot M + s \cdot 2^{\tilde{k}R_b}(\tilde{k} + 1/x)}{\tilde{k}R_b}, \tag{4}$$

where M is the seize of the symbol alphabet, and x the number of parallel branches per transition. The derivation of eqn.(4) can be found in [11]. It is also shown in [11] that the complexity K, eqn.(3), favors trellis codes with large $\tilde{k}$ in an unfair way. This is overcome by the new definition of K_B which shows that the computational effort of the decoding process grows exponentially with $\tilde{k}$.
However, this drawback of trellis codes with high multiplicity is all alleviated by the fact that these codes are well suited for efficient parallel implementation in real time systems [12].

References

[1] G. Ungerböck: Channel Coding with Multilevel/Phase Signals. *IEEE Trans. IT*, Vol. 28, No. 1, pp. 55–67, Jannuary 1982.

[2] G. Ungerböck: Trellis Coded Modulation with Redundant Signal Sets, Part I: Introduction. *IEEE Comm. Mag.*, Vol. 25, No. 2, pp. 5–11, February 1987.

[3] G. Ungerböck: Trellis Coded Modulation with Redundant Signal Sets, Part II: State of the Art. *IEEE Comm. Mag.*, Vol. 25, No. 2, pp. 12–21, February 1987.

[4] S.S. Pietrobon, R.H. Deng, A. Lafanchére, G. Ungerböck, D. Costello: Trellis Coded Multidimensional Phase Modulation. *IEEE Trans. IT*, Vol. 36, No. 2, pp. 63–89, January 1990.

[5] D. Divsalar, M. Simon: The Design of Trellis Coded MPSK for Fading Channels: Performance Criteria. *IEEE Trans. COM*, Vol. 36, No. 9, pp. 1004–1011, September 1988.

[6] D. Divsalar, M. Simon: The Design of Trellis Coded MPSK for Fading Channels: Set Partitioning for Optimum Code Design. *IEEE Trans. COM*, Vol. 36, No. 9, pp. 1013–1021, September 1988.

[7] C. Schlegel, D. Costello: Bandwidth Efficient Coding for Fading Channels: Code Construction and Performance Analysis. *IEEE Jour. SAC*, Vol. 7, No. 9, pp. 1356–1368, December 1989.

[8] J. Du, B. Vucetic: New M–PSK Trellis Codes for Fading Channels. Electronic Letters, Vol. 26, No. 16, pp. 1267–1269, August 1990.

[9] S.H. Jamali, T. LeNgoc: A New 4-State 8PSK TCM Scheme for Fast Fading, Shadowed Mobile Radio Channels. *IEEE Trans. VT*, Vol. 40, No. 1, pp. 216–222, February 1991.

[10] S.S. Periyalwar, S.M. Fleisher: A Modified Design of Trellis-Coded MPSK for the Fading Channel. *IEEE Trans. COM*, Vol. 41, No. 6, pp. 874–882, June 1993.

[11] B.J. Mayr: Codiertechniken für den Schwundkanal. *Dissertation, TU-Wien*, 1995, in German.

[12] R. Schweikert, DLR Oberpfaffenhofen Germany, *personal communication* at the ITG-Fachtagung *Mobile Kommunikation* in Ulm, Germany, September 1995.

The Power Spectral Density of Maximum Entropy M-ary (d,k) Codes

Bane Vasić
Rochester Institute of Technology
Rochester, NY 14623

Abstract: A formula for the power spectral density of maximum entropy M-ary (d,k) constrained sequences is given. The formula for spectrum is derived by using the method of a difference equation assigned to a Markov chain generating M-ary (d,k) sequences.

Introduction

Multi-amplitude (d,k) codes are introduced by French and Wolf [4], Earman [3], and McLaughlin [7]. One of the applications of these codes are electron trapping optical recording channels [7]. The M-ary (d,k) code stream $\{a^{(n)}\}_{n\in Z}$ consists of the symbols from an alphabet $A=\{A_1,\ldots,A_M\}$ of size M. The lengths of the sequences of consecutive like symbols (phrases) are constrained, and must be in the range $[d+1,k+1]$. In this letter we consider the power spectral density (spectrum) of the M-ary (d,k) codes. The formula for spectrum of maximum entropy M-ary (d,k) codes is derived.

The generating a constrained sequence $\{a^{(n)}\}_{n\in Z}$ is modeled by reading off the state labels during a random walk through a finite directed Moore-type graph [2, Chpt 3]. When probabilities between states are specified, then a sequence of graph states $\{s^{(n)}\}_{n\in Z}$ become a Markov chain, and a sequence $\{a^{(n)}\}_{n\in Z}$ is a memoryless function of the Markov chain $a^{(n)}=h(s^{(n)})$ [1, Chpt. 12]).

To derive a closed form expression of the power spectral density of a memoryless function of the Markov chain we use the difference equation method described by Vasić in [8] and [9]. This is primarily a numericaly efficient algorithm for spectrum computation, but the D-domain version of this method simplifies the algebraic manipulation with the expression for the autocorrelation function in cases when a graph has some specific structure.

The Markov Chain Generating M-ary (d,k) Sequences

The transition diagram of the Moore-type Markov chain generating M-ary (d,k) sequences is shown in Fig. 1. The diagram consist of M identical branches modeling the generation of the phrases of different symbols A_m, $1\le m\le M$. The states are drawn as a circles and denoted by pairs (m,i), $1\le m\le M$, $1\le i\le k+1$. The label inside the circle denotes the symbol generated from this state ($A_m=h((m,i)$, for all i). The incoming into the state (m,i) means starting the generation of the phrase of length i, of symbols A_m. The edge labels are of the form pD, where p is the transition probability and D is the time delay operator. The Markov chain described by the transition diagram given in Fig. 1 is stationary and ergodic [1, Chpt. 12]. The stationary probabilities of states and transition probability between states of the Markov chain are given by the following theorems.

Lemma 1: The M-ary (d,k) constrained sequence achieves maximal information rate if the phrase lengths are i.i.d. random variables with probability distribution $P_i=(M-1)M^{-Ci}$, $d+1\le i\le k+1$, wherein C is constant satisfying

$$\sum_{i=d+1}^{k+1} M^{-C\cdot i}=1/(M-1). \qquad (1)$$

Proof: The proof is straigforward generalization of the Theorem 1 of Zehavi and Wolf [10]. We can also prove this lemma by considering variable length graph [6] of M-ary (d,k) constraint. This graph contains $k+1-d$ loops, and the labels of loops are of the form $(M-1)D^i$, $d+1\le i\le k+1$ (see [6]).

Remark 1: C is the channel capacity of M-ary (d,k) constraint in M-ary unit of information amount.

Theorem 1: For the Markov chain generating M-ary (d,k) constrained sequences of maximal entropy we have:

0-8186-7576-4/96 $5.00

a) The stationary probability of states are $Pr\{s^{(n)}=(m,i)\}=\pi_i=(1/ML)(P_i+\ldots+P_{k+1})$, $1\le m\le M$, $1\le i\le k+1$, wherein L is average phrase length $L=(d+1)P_{d+1}+\ldots+(k+1)P_{k+1}$.

b) The transition probability to (m,i) from any state $(l,1)$, $1\le l\le M$, $l\neq m$ is $Pr\{s^{(n)}=(m,i)|s^{(n-1)}=(l,1)\}=(M-1)P_i$.

Proof: The proof is given in Appendix A

Power Spectral Density Formula

To derive a closed form expression of the power spectral density of the stream $\{a^{(n)}\}_{n\in Z}$ which is a memoryless function of the Markov chain $\{s^{(n)}\}_{n\in Z}$ given by the transition diagram shown in Fig. 1 we use the difference equation method described by Vasić in [8] and [9]. Let us denote the Markov chain state set by $\mathcal{S}$, and one step transition probabilities between states $u,v\in\mathcal{S}$ by $p_{u|v}=Pr\{s^{(n)}=u|s^{(n-1)}=v\}$. Let $\mathcal{S}_u$ be the set of states from which Markov chain can pass into state u, let p_u be stationary probability of state u, and let symbol A_u be generated from state u i.e. $A_u=h(u)$. Then the difference equation method is summarized in the following theorem.

Theorem 2 [8],[9]: The autocorrelation func-tion $r^{(n)}=E\{a^{(i)}a^{(i+n)}\}$, $j\in Z$, $n\ge 0$, of the stream $\{a^{(n)}\}_{n\in Z}$, $a^{(n)}=h(s^{(n)})$ where $\{s^{(n)}\}_{n\in Z}$ is a Markov chain can be obtained from the following set of difference equations

$$\begin{aligned} f_u^{(0)} &= A_u p_u, \quad u\in\mathcal{S} \\ f_u^{(n)} &= \sum_{v\in\mathcal{S}_u} p_{u|v}\cdot f_v^{(n-1)}, \quad u\in\mathcal{S} \\ r^{(n)} &= \sum_{v\in\mathcal{S}} A_u\cdot f_u^{(n)} \end{aligned} \tag{2}$$

Since the stream $\{a^{(n)}\}_{n\in Z}$ is cyclostationary, but not wide sense stationary random process, for finding its spectrum the Wiener-Kinchine theorem cannot be used. The problem is usually avoided by considering the phase averaged process $\{a^{(n+\vartheta)}\}_{n\in Z}$ [5]. Since the stream $\{a^{(n)}\}_{n\in Z}$ is real valued it follows that $r^{(-n)}=r^{(n)}$, and the spectrum of the stream $\{a^{(n+\vartheta)}\}_{n\in Z}$ is $\Phi(f)=r^{(0)}+R(D)+R(D^{-1})$, where $R(D)$ is D transform of the $r^{(n)}$ ($R(D)=\sum_{n\ge 0} r^{(n)}$), and D=exp(j2πf).

By applying the system of equations (2) given by Theorem 2 on the Markov chain generating M-ary (d,k) constrained sequences in Fig. 1, with stationary and transition probabilities given by statements of Theorem 1, we obtain the formula for the spectrum.

Theorem 3: The power spectral density of the M-ary (d,k) constrained stream $\{a^{(n+\vartheta)}\}_{n\in Z}$ is given by

$$\Phi(f)=\frac{1}{L\sin^2(\pi f)}\,\frac{|G(D)|+(M-1)\,\mathrm{Re}\{G(D)\}-(M-1)}{|M-1+G(D)|^2}$$

where $G(D)=\sum_{i=d+1}^{k+1} P_j\cdot D^j$ and τ is random variable uniformly distributed over $[0,1)$.

Proof: Proof is given in Appendix B.

As an illustration in Fig. 2 the spectra of signals modulated by maxentrropic M-ary (d,k) sequences are shown for different alphabet size M. The rectangular pulse shape is assumed. We can observe that the content of low frequency spectral components increases with the increasing M.

Conclusion

By using the difference equation method, we have derived the clossed form expression for spectrum of M-ary maxentropic (d,k) codes. The expression for $R(D)$ is very similar to one for binary (d,k) codes, obtained by Zehavi and Wolf [10] and Gallopoulos, Heegard and Siegel [5]. Instead of $1+G(D)$ in the denominator of the formula for $R(D)$ for binary codes we have the term $M-1+G(D)$.

Appendix A

Proof of Theorem 1: When a phrase of symbols A_l is completed, the probability of selecting some other amplitude A_m $(A_m\neq A_l)$ is $1/(M-1)$. So, the labels of all edges incoming to the branch m have term $1/(M-1)$. In other words $Pr\{s^{(n)}\in\{(m,i)|1+d\le i\le k+1\}|s^{(n-1)}=(l,1)\}=1/(M-1)$. When a branch is selected, the probability of generating a phrase of the length i is P_i, $d+1\le i\le k+1$. Obviously $P_i=0$ for $1\le i<d+1$,

and for $i>k+1$. According to Lemma 1, for maxentropic sequences we have $P_i=(M-1)M^{-Ci}$, $d+1\le i\le k+1$, so that

$$\Pr\left\{s^{(n)}=(m,i)\right\}\left|s^{(n-1)}=(l,1)\right\}=\frac{1}{M-1}\cdot P_i$$

Since all M branches are equiprobable, according to the Bayes formula, for stationary probabilities $\pi_i=Pr\{s^{(n)}=(m,i)\}$, $1\le i\le k+1$, $1\le m\le M$, we can write (see Fig. 1)

$$\pi_{k+1}=P_{k+1}\cdot\pi_1$$
$$\pi_i=1\cdot\pi_{i+1}+P_i\cdot\pi_1,\quad d+1\le i\le k$$
$$\pi_i=1\cdot\pi_{i+1},\quad 1\le i\le d$$

By solving the system (A.2) we obtain $\pi_i=(1/ML)\sum_{j=i}^{k+1}P_j$, $1\le i\le k+1$, where L is average phrase length. Q.E.D.

Appendix B

For Markov source generating M-ary (d,k) constrained sequences (Fig. 1) the initial conditions in difference equation method [8] are $f_{m,i}^{(0)}=A_m\pi_i$, $1\le i\le k+1$, $1\le m\le M$, wherein π_i are state stationary probabilities given by the statement of Theorem 2. The autoccorelation function at $n=0$, $r^{(0)}$ is $r^{(0)}=(1/M)\cdot(A_1^2+\ldots+A_M^2)$.

The D transforms of difference equations following from Fig. 1 are

$$F_{m,i}(D)=\frac{1}{M-1}\cdot D\cdot W_m(D)\cdot\sum_{j=i}^{k+1}P_i\cdot D^{j-i}+$$
$$+\sum_{j=i}^{k+1}f_{m,i}^{(0)}\cdot D^{j-i}$$

where $W_m(D)=\sum_{l=1,\,l\ne m}^{M}F_{l,1}(D)$; $1\le m\le M$.

By summing the expression for $F_{m,1}(D)$ over all m, and after some algebraic manipulation we have

$$W(D)=\sum_{m=1}^{M}F_{m,i}(D)=G(D)\cdot W(D)+$$
$$+\frac{1}{ML}\cdot\frac{1-G(D)}{1-D}\cdot\sum_{m=1}^{M}A_m \qquad (B.1)$$

where $G(D)=\sum_{i=d+1}^{k+1}P_j\cdot D^j$.

Without loss of generality the bipolar, zero mean stream can be assumed. For maxentropic case all symbols are equiprobable, so we have that $A_1+\ldots+A_M=0$. Since $G(D)\ne0$, from (B.1) we have $W(D)=0$. From this, since $W(D)=W_m(D)+F_{m,1}(D)$, it follows $F_{m,1}(D)=-W_m(D)$.

From the expression (B.1) for $W(D)$ it follows

$$W_m(D)=-(M-1)\cdot\frac{A_m}{ML}\cdot\frac{1}{1-D}\cdot\frac{1-G(D)}{M-1+G(D)}$$

If D transform of the autocorrelation function is written in the form $R(D)=A_1R_1(D)+\ldots+A_MR_M(D)$, where $R_m(D)=F_{m,1}(D)+\ldots+F_{m,k-1}(D)$, then, after some manipulation, we obtain

$$R_m(D)=-\frac{A_m}{ML}\frac{2D}{(1-D)^2}\frac{1-G(D)}{M-1+G(D)}+\frac{A_m}{M}\frac{1}{1-D}$$

The D transform of whole autocorrelation function is

$$R(D)=\frac{1}{1-D}-\frac{1}{L}\frac{2D}{(1-D)^2}\frac{1-G(D)}{M-1+G(D)}$$

From the above expression and definition of the spectrum, the statement of the Theorem 3 follows directly. Q.E.D.

Rererences

[1] Borovkov, *Probability theory*, Moskva, Nauka, 1986 (in Russian).

[2] Cvetković, *Graph theory and applications*, Belgrade, Naučna knjiga, 1981 (in Serbian).

[3] Earman, "Optical data storage with electron trapping materials using M-ary data channel coding," Proceedings of the Optical Data Storage Conference, San Jose, CA, Oct. 1992.

[4] French and J.K. Wolf, "Results involving (D,K) constrained M-ary codes," IEEE Trans. Magnetics, vol. 23, n0. 5, pp. 3678-3680, Sept. 1987.

[5] Gallopoulos, C. Heegard and P. H. Siegel, "The power spectrum of run-length-limited codes," IEEE Trans. Commun., vol. 37, no. 9, pp. 906-917, Sept. 1989.

[6] D. Heegard, B. H. Marcus and P. H. Siegel, "Variable-length state splitting with applications to average runlength-constrained

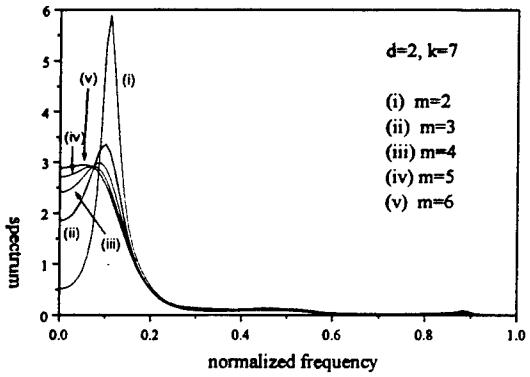
(ARC) codes," IEEE Trans. Inform. Theory, vol. 37, no. 3, pp. 759-777, May 1991.

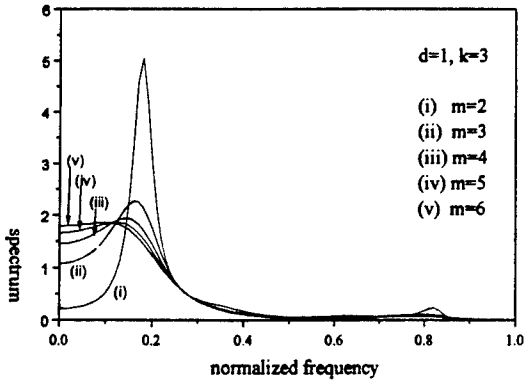
[7] McLaughlin, "Improved distance M-ary (d,k) codes ofr high density recording," IEEE Trans. Magnetics, vol 31, no. 2, March 1995.

[8] Vasić and M. Stefanović, "Spectral analysis of coded digital signals by means of difference equation systems," El. letters, vol. 27, no. 24, pp.2272-2273, Nov. 1991.

[9] Vasić, Spectral analisys of codes for magnetic and optical recording, Ph.D. Thesis, University of Niš, 1993.

[10] Zehavi anxd J. K. Wolf, "On runlength codes," IEEE Trans. Inform. Theory, vol. 34, no. 1, pp. 45-54, Jan. 1988.





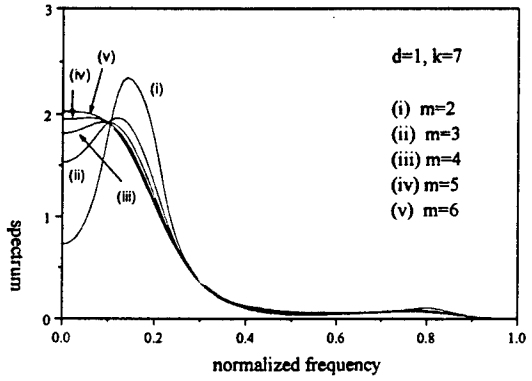


Figure 2:
Spectra of the signals modulated by the M-ary (d,k) sequences for different values of M

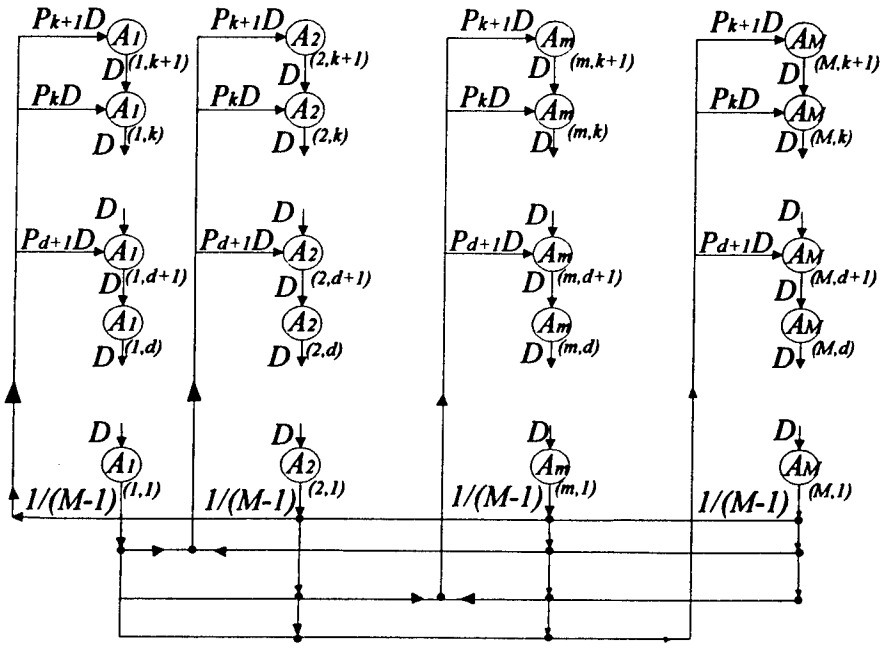


Figure 1
The transition diagram of the Markov chain generating M-ary (d,k) constrained sequences

Performance of a Complex Digital System: Optical Cable-Coaxial Cable

Mihajlo Č. Stefanović, Ivan B.Djordjević, Goran T. Djordjević, Djordje Milošević
Faculty of Electronic Eng., University of Niš, Beogradska 14, 18000 Niš, Yugoslavia
E-mail: ica@efnis.elfak.ni.ac.yu

Abstract

In this paper digital system which contains two parts is considered. The first part is optical communication system with ASK modulation. Optical receiver is constructed as heterodyne asynchronous ASK receiver. Signal is transmitted in baseband through the second part of system (coaxial cable). Probability density functions for both of hypotheses (H_0 and H_1), optimal threshold and performance of system are determined for proposed system. Optimal threshold is determined on condition that error probability is minimal.

1 Introduction

Digital system which is investigated in this paper has two parts. The first part is optical communication system with ASK modulation. Optical receiver is constructed as heterodyne asynchronous ASK receiver. Signal is transmitted in baseband (for example by means of coaxial cable) through the second part of system. System ought to be designed on condition that error probability is less than beforhand given value (for example 10^{-7}). As optical system could be designed that error probability is slight enough, estimator is not strictly behind envelope detector. Signal is transmitted in baseband by means of section of coaxial cable, and estimation is done at the end of this section. Noise components in optical receiver (shot noise, thermal noise etc.) are approximated by white Gaussian noise. Disturbances which are appeared during transmission by means of coaxial cable can be represented as sum of additive white Gaussian noise and crosstalk. Crosstalk is modelled as sinusoidal disturbance with constant amplitude and uniform distribution of phase.

Probability density functions for both of hypotheses (H_0 and H_1), optimal threshold and performance of system are determined. Optimal threshold is determined on condition that error probability is minimal.

2 Performance determination

The first part of system is constructed as optical communication system and it contains optical ASK transmitter, optical cable, 3 dB coupler (balanced detector), two photodiodes, band pass filter and envelope detector. The second part of system, in baseband, contains coaxial cable, low-pass filter, sampler and estimator. Optical receiver is constructed as heterodyne asynchronous ASK receiver.

In Fig.1 $i(t)$ represents crosstalk which is modelled as sinusoidal disturbance $i(t) = A_1 \cos(\omega_0 t + \theta_1)$, where θ_1 has uniform distribution of phase $p(\theta_1) = 1/2\pi$, $\mid \theta_1 \mid < \pi$. LO lightwave addition to incoming optical signal is implemented by using balanced detector. Outcoming signals of balanced detector (X_1 and X_2) illuminate separate photodiodes and generate currents

$$i_1(t) = R \mid X_1 \mid^2 + n_1(t) \tag{1}$$

$$i_2(t) = R \mid X_2 \mid^2 + n_2(t) \tag{2}$$

where $n_1(t)$ and $n_2(t)$ are shot noise components with power spectral densities $S_{n1}(\omega) = qR \mid X_1 \mid^2$ and $S_{n2}(\omega) = qR \mid X_2 \mid^2$; $R = \eta q/(h\upsilon$ is responsivity of photodiode; η is quantum efficiency and q is charge of an electron. Total current is determined as [1]

$$\begin{aligned} i_1(t) - i_2(t) &= R(\mid X_1 \mid^2 - \mid X_2 \mid^2 + n(t) \\ &= 2R_e\{SL^*\} + n(t) \end{aligned} \tag{3}$$

$R_e\{\}$ is real part of complex number, and $n(t) = n_1(t) - n_2(t)$ is approximated as a zero-mean white Gaussian noise with PSD

$$S_n(\omega) = qR(\mid X_1 \mid^2 + \mid X_2 \mid^2) = qR(\mid S \mid^2 + \mid L \mid^2) \tag{4}$$

As incoming optical signal has form

$$S = a\sqrt{P_s}e^{j\omega_0 T}, \quad a \in \{0, 1\} \tag{5}$$

0-8186-7576-4/96 $5.00

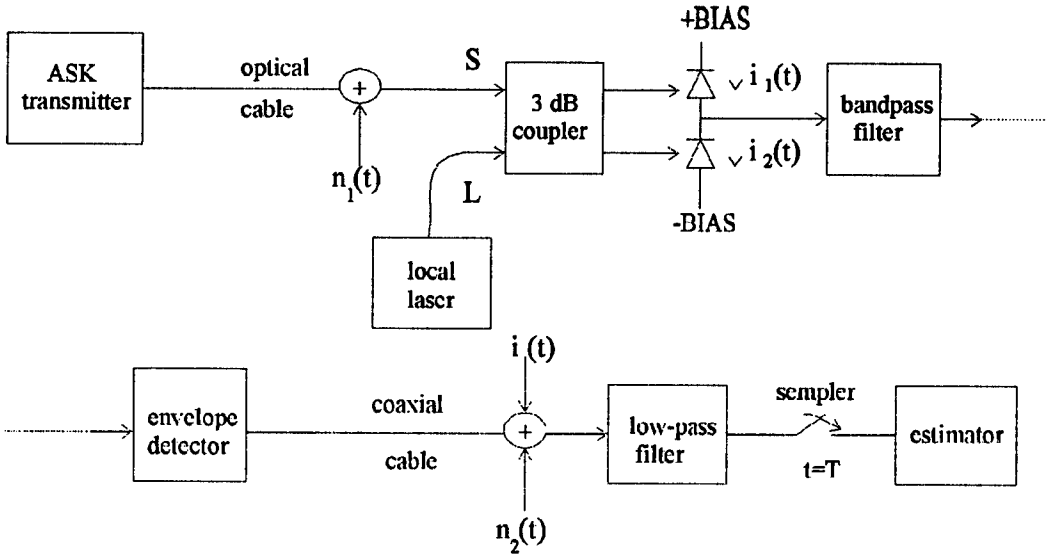


Fig 1.

and LO lightwave is

$$L = \sqrt{P_{LO}} e^{j\omega_{LO} t} \tag{6}$$

total current will be

$$\begin{aligned} i_1(t) - i_2(t) &= a2R\sqrt{P_s P_{LO}} \cos[(\omega_0 - \omega_{LO})t] \\ &= a2R\sqrt{P_s P_{LO}} \cos(\omega_{IF} t) + n(t) \end{aligned} \tag{7}$$

So, problem is reduced to the classical detection problem of a known signal in additive white Gaussian noise [2],[3]. As it has been shown in [3], probability density function for stochastic variable Y (behind envelope detector) has Rayleigh distribution when $a = 0$

$$P_Y(y/H_0) = \frac{y}{\sigma^2} e^{-\frac{y^2}{2\sigma^2}}, \quad y \geq 0 \tag{8}$$

and Ricean distribution when $a = 1$

$$P_Y(y/H_1) = \frac{y}{\sigma^2} e^{-\frac{y^2+A^2}{2\sigma^2}} I_0(\frac{yA}{\sigma^2}, \quad y \geq 0 \tag{9}$$

where $\sigma^2 = qRP_{LO}$ and $A = 2R\sqrt{P_s P_{LO}}$.

Disturbances which are appeared during transmission by means of coaxial cable can be represented as sum of additive zero- mean white Gaussian noise (with variance σ^2) and crosstalk. Crosstalk is modelled as sinusoidal disturbance with constant amplitude (A_1) and uniform distribution of phase (θ_1). So, conditional probability density function for stochastic variable Z (in reception place) for the case of hypothesis H_0 can be determined as

$$P_{Z/H_0}(z/\theta_1, y) = \frac{1}{\sigma_2\sqrt{2\pi}} exp(-\frac{(z - y - A_1 \cos\theta_1)^2}{\sigma_2^2}) \tag{10}$$

where θ_1 has uniform distribution and y has Rayleigh distribution and PDF is

$$\begin{aligned} P_{Z/H_0}(z) = &\int_{-pi}^{\pi} \int_0^{\infty} \frac{1}{\sigma_2\sqrt{2\pi}} \\ &exp(-\frac{(z - y - A_1 \cos\theta_1)^2}{2\sigma_2^2}) \\ &\frac{1}{2\pi}\frac{y}{\sigma^2} exp(-\frac{y^2}{2\sigma^2}) d\theta_1 dy \end{aligned} \tag{11}$$

On the same way PDF for stochastic variable Z (in reception place) for the case of hypothesis H_1 is

$$\begin{aligned} P_{Z/H_1}(z) = &\int_{-pi}^{\pi} \int_0^{\infty} \frac{1}{\sigma_2\sqrt{2\pi}} \\ &exp(-\frac{(z - y - A_1 \cos\theta_1)^2}{2\sigma_2^2}) \\ &\frac{1}{2\pi}\frac{y}{\sigma^2} exp(-\frac{y^2 + A^2}{2\sigma^2}) I_0(\frac{Ay}{\sigma^2}) d\theta_1 dy \end{aligned} \tag{12}$$

If we set $\cos\theta_1 = x$ and use Chebyshev quadratures expressions, (11) and (12) become

$$\begin{aligned} P_{Z/H_0}(z) = &\frac{1}{\sigma^2\sigma_2 N\sqrt{2\pi}} \sum_{k=1}^{N} \\ &\int_0^{\infty} y exp(-\frac{y}{2\sigma^2} - \frac{(z - y - A_1 x_k)^2}{2\sigma_2^2}) dy \end{aligned} \tag{13}$$

$$\begin{aligned} P_{Z/H_1}(z) = &\frac{1}{\sigma^2\sigma_2 N\sqrt{2\pi}} \sum_{k=1}^{N} \int_0^{\infty} y exp \\ &(-\frac{y}{2\sigma^2} - \frac{(z - y - A_1 x_k)^2}{2\sigma_2^2}) I_0(\frac{Ay}{\sigma_2^2}) dy \end{aligned} \tag{14}$$

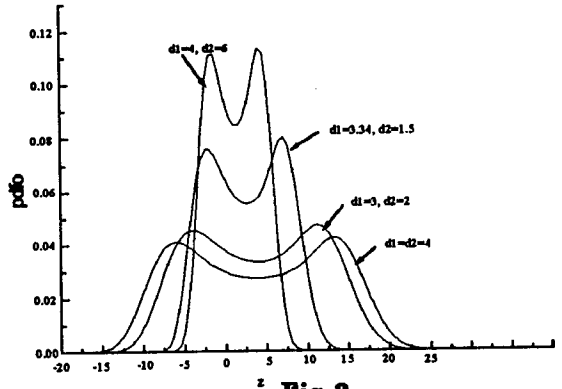


Fig 2.

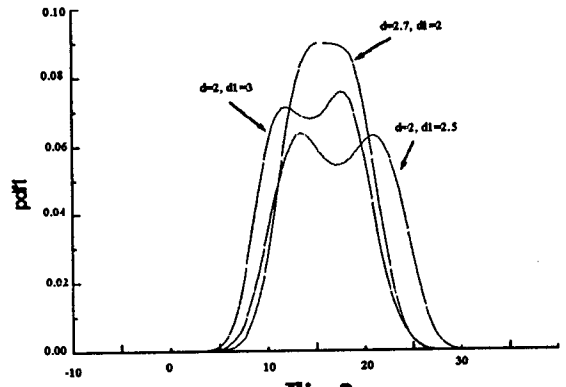


Fig 3.

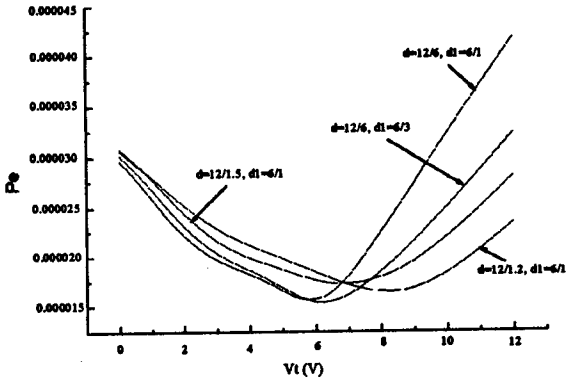


Fig 4.

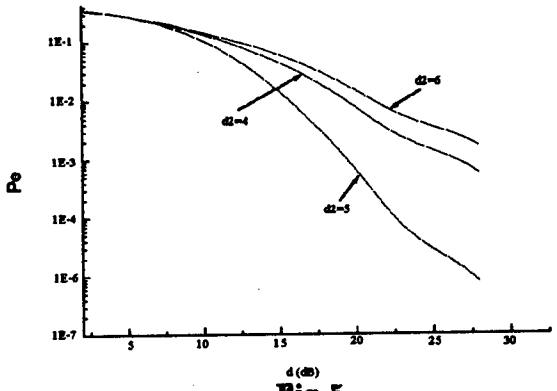


Fig 5.

where

$$x_k = \cos(\frac{2\pi}{N}(2k-1)), \quad (k = 1, 2, ..., N) \qquad (15)$$

PDF for hypothesis H_0 is given graphically in Fig. 2. A_1/σ and A_1/σ_2 are parameters.

PDF for hypothesis H_1 is given graphically in Fig. 3. A/σ and A_1/σ_2 are parameters. System ought to be designed on condition that error probability is not greater than beforehand given value. Error probability dependence on threshold is drawn in Fig.4. Optimal threshold is determined. It is 0.65A. Error probability dependence on signal to noise ratio ($20log(A/\sigma)$, in dB), for this optimal threshold, is drawn in Fig.5.

3 Conclusion

Based on obtained results for the proposed system, we can conclude that this system can be applied in practice.

References

[1] J.H. Bary, E.A. Lee, "Performance of Coherent Optical Receivers" Proceedings of the IEEE, vol.78, no.8, August 1990."

[2] A.D. Whalen, "Detection of Signals in noise", Academic Press, New York and London, 1971.

[3] N.R. Levin, "Teoreticheskie osnovi statisticheskoy radiotehniki", Sovetskoe radio, Moscow 1974.

An Intelligent LMS+F Algorithm

Dimitrios I. Pazaitis and A.G. Constantinides

Signal Processing Section, Dept. Electrical and Electronic Engineering
Imperial College, Exhibition Road, London SW7 2BT, UK
e-mail : {d.pazaitis, a.constantinides}@ic.ac.uk

Abstract

A new techique for combining the LMS and LMF cost sunctions is proposed in this contribution. The resulting stochastic gradient adaptive algorithm uses a time varying mixing parameter to optimise a combination of the above cost functions, taking into consideration the noise statistics. Furthermore, the behaviour of the proposed algorithm is analysed and convergence conditions are established. Simulation results verify the ability of the algorithm to adapt itself to the noise characteristics, illustrate its enhanced performance and support very well the theoretic analysis. The continuous adaptation of the mixing parameter adds flexibility and enables rapid response of the algorithm to non stationarities.

1. Introduction

Stochastic gradient adaptive algorithms are quite popular and have been used in a wide variety of applications including array processing, system identification, channel equalization, echo and other interference cancellations, mainly due to their inherent simplicity. LMS is the mostly known such algorithm and its performance has been thoroughly investigated in the literature (e.g [4, 5, 7, 12]). Its most attractive feature is its amenability to simple implementation, while its main drawback is the degradation of its performance due to eigenvalue spread.

The LMS algorithm belongs to a more general family, which attempts to optimize (minimize) the following cost function [12]

$$J = E\{e^{2K}(n)\} \quad K = 1, 2, \ldots . \tag{1}$$

and is obtained by setting K equal to 1, i.e., $J = E\{e^2(n)\}$. For $K = 2$, we obtain the second member of the family, i.e., $J = E\{e^4(n)\}$, known as the Least Mean Fourth (LMF) algorithm. Walach and Widrow compared the two algorithms above and found out that under certain conditions the LMF outperforms LMS [12]. Furthermore, it is obvious that when far from the optimum (i.e., $e(n) > 1$) the LMF algorithm exhibits faster convergence.

The *LMS+F* algorithm [8] was a first approach towards the combination of the advantages of the two algorithms. More precisely, *LMS+F* aimed at exploiting the faster initial convergence of the LMF algorithm, while retaining the desirable LMS characteristic of low misadjustment and immunity over the different distributions of noise, when around the optimum.

A mixed criterion algorithm of this type, obtained though from a different perspective, was also presented in [3]. The proposed algorithm was derived in an attempt to minimise the variance of the square error subject to a constraint on the mean square error. The constrained minimisation method resulted in a weight update formula with a fixed preselected mixing parameter. However, no information was provided about the optimal value of the mixing parameter or the constraint. The proposed combination appeared also in [1], where a constant mixing parameter λ was used. In this contribution, the λ selection criterion of [8] is modified, so that the evolution of the λ sequence takes into account the noise characteristics, enhancing thus the algorithm's performance. Moreover, the adaptive nature of the λ_n sequence adds flexibility to our algorithm.

2. The LMS+F Adaptive Algorithm

The following type of cost function is proposed

$$J = (1 - \lambda_n)E\{e^2(n)\} + \lambda_n E\{e^4(n)\}, \tag{2}$$

where λ_n is a time varying scalar sequence and α is a scaling factor, selected according to the following rule

$$\lambda_{n+1} = \begin{cases} \lambda_n + \alpha & \text{, if } E\{e_n^4\} \geq 1 \\ \lambda_n - \alpha sgn\{C_e^4(n)\} & \text{, otherwise} \end{cases}, \tag{3}$$

where *sgn* is the *sign* (*signum*) function, $C_e^4(n) = E\{e_n^4\} - 3E\{e_n^2\}$ is the kurtosis (fourth order cumulant) of the assumed zero mean error signal and α is a scaling factor lying in the interval $[0, 1]$. By choosing $\alpha = 1.0$ no "transient" behaviour is exhibited and the algorithm just alternates between LMS and LMF. To preserve the unimodal characteristic of the mixed cost function, as both of the consisting functions are characterised by convexity, λ_n in (2) is confined to the closed interval $[0, 1]$.

As it can be seen from the first part of (3), when far from the optimum weight vector, i.e., $e(n) > 1$, λ_n increases, increasing therefore the convergence speed, but as the weight vector $\boldsymbol{H}_n$ approaches its optimum value ($\boldsymbol{H}^{opt}$), λ_n adjusts itself according to the noise distribution, enhancing therefore the algorithms's performance. The first part of the above formula implicitly assumes that the noise variance is lower than unity. If this is not the case, it can be

This work was supported by a scholarship from the State Scholarship Foundation (I.K.Y) of the Hellenic Republic.

0-8186-7576-4/96 $5.00 © 1996 IEEE

either ignored or automatic gain control or normalisation with respect to the input power could be applied to enable the use of the above update equation, which significantly improves the initial convergence of the *LMS+F* algorithm in non stationary environments [8]. Applying the steepest descent gradient search, and using the instantaneous value of the gradient instead of the mean, the filter coefficient vector update equation for the cost function (2) becomes

$$\boldsymbol{H}_{n+1} = \boldsymbol{H}_n + 2\mu \left((1-\lambda_n)e(n) + 2\lambda_n e^3(n)\right) \boldsymbol{X}_n, \quad (4)$$

where

$$e(n) \triangleq w_n - \boldsymbol{X}_n^T \boldsymbol{V}_n \quad (5)$$

and the weight error vector is given by

$$\boldsymbol{V}_n = \boldsymbol{H}_n - \boldsymbol{H}_n^{opt} \quad (6)$$

We can now proceed to analyse the proposed algorithm.

3. Convergence analysis of the LMS+F algorithm

To facilitate our analysis we introduce the commonly used assumption that the various input vectors come from mutually independent zero mean gaussian distributed sequences [2, 4, 5, 12]. Although this is untrue in many applications, since consecutive input vectors share $N-1$ entries, it is widely accepted to capture the first order behaviour and is extensively used in the literature to simplify the analysis producing at the same time reliable results [7]. In our case, the assumption can be relaxed, in that we only require that the input sequence is uncorrelated with the filter weights. Other than that, no restriction applies to the nature of the input autocorrelation matrix $\boldsymbol{R}$.

We also approximate the conditional expectation terms of the form $E\{e_n^2|\boldsymbol{V}_n\}$ with the unconditional mean square estimation error. For slow adaptation conditions, the weight error vector oscilates around the mean value justifying partially the above assumption. Furthermore, according to the central limit theorem, as the filter length increases, the distribution of the error signal gets closer to the gaussian one. This assumption is not new and has led to analytic results which agreed well with the simulated behaviour of nonlinear algorithms [2, 6].

Finally, the estimation error e_n is assumed to follow a gaussian distribution. The above assumption is justified, when the weight vector $\boldsymbol{H}_n$ varries much slower than the input vector $\boldsymbol{X}_n$; a condition corresponding to the slow adaptation case. This assumption has produced reliable results and was used succesfully in [6, 11]. The approximate validity of the above assumptions will be confirmed by the simulation results.

By substracting $\boldsymbol{H}^{opt}$ from both sides of equation (4) using (6) and (5) results in

$$\begin{aligned} \boldsymbol{V}_{n+1} &= \boldsymbol{V}_n + \mu\left((1-\lambda_n)(w_n - \boldsymbol{X}_n^T\boldsymbol{V}_n)\right) \\ &+ \mu\left(2\lambda_n(w_n - \boldsymbol{X}_n^T\boldsymbol{V}_n)^3\right)\boldsymbol{X}_n. \end{aligned} \quad (7)$$

We wish now to develop a recursive equation for the time evolution of the correlation matrix of the weight error vector $\boldsymbol{V}_n$. Using $\boldsymbol{K}_n$ to denote this correlation matrix at time instant n, we have, by definition,

$$\boldsymbol{K}_n \triangleq E\{\boldsymbol{V}_n\boldsymbol{V}_n^T\}. \quad (8)$$

Substituting (7) in (8) we obtain

$$\begin{aligned} E\{\boldsymbol{V}_{n+1}\boldsymbol{V}_{n+1}^T\} &= E\{\boldsymbol{V}_n\boldsymbol{V}_n^T\} \\ &+ E\left\{\mu\left((1-\lambda_n)e_n + 2\lambda_n e_n^3\right)\left[\boldsymbol{V}_n\boldsymbol{X}_n^T + \boldsymbol{X}_n\boldsymbol{V}_n^T\right]\right\} \\ &+ E\left\{\mu^2\left((1-\lambda_n)e_n + 2\lambda_n e_n^3\right)^2 \boldsymbol{X}_n\boldsymbol{X}_n^T\right\}. \end{aligned} \quad (9)$$

To obtain the individual terms on the right hand of the above equation we take conditional expectations (with respect to $\boldsymbol{V}_n$) and then average over all $\boldsymbol{V}_n$. We thus obtain

$$E\left\{e_n\left(\boldsymbol{V}_n\boldsymbol{X}_n^T + \boldsymbol{X}_n\boldsymbol{V}_n^T\right)\right\} = -\left(\boldsymbol{K}_n\boldsymbol{R} + \boldsymbol{R}\boldsymbol{K}_n\right). \quad (10)$$

Adopting a similar approach and applying Price's [10] theorem on the right hand terms of (9), we obtain [9]

$$E\left\{e_n^3\left(\boldsymbol{V}_n\boldsymbol{X}_n^T + \boldsymbol{X}_n\boldsymbol{V}_n^T\right)\right\} = -3\sigma_{e_n}^2\left(\boldsymbol{K}_n\boldsymbol{R} + \boldsymbol{R}\boldsymbol{K}_n\right), \quad (11)$$

$$E\left\{e_n^2\boldsymbol{X}_n\boldsymbol{X}_n^T\right\} = \sigma_{e_n}^2\boldsymbol{R} + 2\boldsymbol{R}\boldsymbol{K}_n\boldsymbol{R}, \quad (12)$$

$$E\left\{e_n^4\boldsymbol{X}_n\boldsymbol{X}_n^T\right\} = 3\sigma_{e_n}^4\boldsymbol{R} + 12\sigma_{e_n}^2\boldsymbol{R}\boldsymbol{K}_n\boldsymbol{R}, \quad (13)$$

and

$$E\{e_n^6\boldsymbol{X}_n\boldsymbol{X}_n^T\} = 15\sigma_{e_n}^6\boldsymbol{R} + 90\sigma_{e_n}^4\boldsymbol{R}\boldsymbol{K}_n\boldsymbol{R}. \quad (14)$$

Substituting (10), (11), (12), (13) and (14) in (9) we obtain the following equation for the error correlation matrix

$$\begin{aligned} \boldsymbol{K}_{n+1} &= \boldsymbol{K}_n - \mu\left((1-\lambda_n) + 6\lambda_n\sigma_{e_n}^2\right)\left(\boldsymbol{K}_n\boldsymbol{R} + \boldsymbol{R}\boldsymbol{K}_n\right) \\ &+ \mu^2\left((1-\lambda_n)^2\sigma_{e_n}^2 + 12\lambda_n(1-\lambda_n)\sigma_{e_n}^4 + 60\lambda_n^2\sigma_{e_n}^6\right)\boldsymbol{R} \\ &+ \mu^2\left(2(1-\lambda_n)^2 + 48\lambda(1-\lambda_n) + 360\lambda^2\sigma_{e_n}^4\right)\boldsymbol{R}\boldsymbol{K}_n\boldsymbol{R}, \end{aligned} \quad (15)$$

Due to the form of the above equation, i.e., non linear, an exact convergence condition is difficult to find. To fascilitate our analysis we introduce the concept of the distance T_n [11]

$$T_n = tr[\boldsymbol{\Lambda}\boldsymbol{K}_n'] = tr[\boldsymbol{R}\boldsymbol{K}_n] = E\{(\boldsymbol{V}_n^T\boldsymbol{X}_n)^2\}. \quad (16)$$

Adopting the approach in [4] we develop the following sufficient and necessary condition on the step size parameter that ensures mean square convergence

$$0 < \mu_{max} < \frac{2\left((1-\lambda_n) + 6\lambda_n\left(\sigma_w^2 + T_n\right)\right)}{(I_1 + I_2)(N + 2I_3)\gamma_{max}}, \quad (17)$$

where

$$I_1 = (1-\lambda_n)^2 + 12\lambda_n(1-\lambda_n)(2\sigma_w^2 + T_n), \quad (18)$$

$$I_2 = 60\lambda_n^2\left(3\sigma_w^4 + 3\sigma_w^2T_n + T_n^2\right), \quad (19)$$

$$I_3 = (1-\lambda_n)^2 + 24\lambda(1-\lambda_n) + 180\lambda^2\left(\sigma_w^2 + T_n\right)^2 \quad (20)$$

and γ_{max} stands for the maximum eigenvalue of the input correlation matrix $\boldsymbol{R}$.

4. Simulation results

In this final section we present and analyse the results obtained from simulations. The algorithm is applied to a system identification problem, where the system to be identified is considered non stationary. The optimum filter coefficients assume the following initial values $\boldsymbol{H}_0^{opt} = [0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0,\ 1.0,\ 0.8,\ 0.6,\ 0.4,\ 0.2]$ and, after that, experiences random disturbances. The null vector $(\mathbf{0})$ is chosen as the initial vector - starting point - $\boldsymbol{H}_0$, and all the results are obtained by averaging over an ensemble of 100 runs. The parameter λ is initialised to 1, i.e. we start with the LMF algorithm. The input is assumed gaussian distributed and both noise and input sequences are assumed to be zero mean i.i.d with input and noise variance equal to *unity* and 0.1 respectively. The *System mismatch* $\left(E\{\boldsymbol{V}_n^T \boldsymbol{V}_n\}\right)$ is selected as a performance measure. The mean square and the mean fourth value of the error are estimated using the following formula

$$E\{e_n^k\} = \beta E\{e_{n-1}^k\} + (1-\beta)e_n^k, \qquad (21)$$

where $k = 2, 4$ and the constant β is a memory controlling factor. The larger the value it assumes the "stronger" the memory of the system. Alternatively, a finite length moving window could be used.

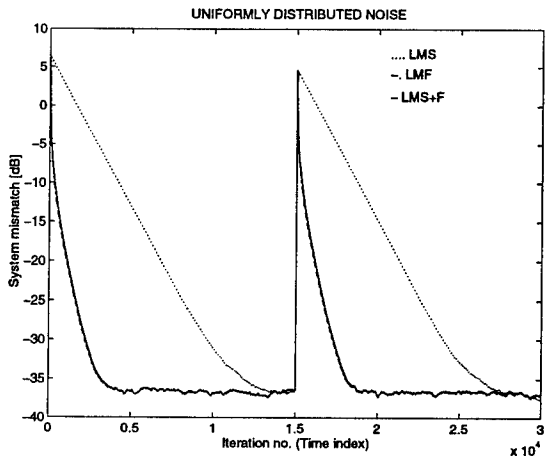


Figure 1: Performance comparison between the LMS (dotted line), the LMF (dash dot line) and the $LMS + F$ algorithm (solid line) under uniformly distributed noise conditions

Figures 1,2 and 3 depict the performance behaviour of all the algorithms, namely LMS, LMF and $LMS+F$, under various noise conditions. The parameters of the above algorithms were chosen so as either to match the steady state error (misadjustment) or the convergence rate, depending on the insight in the performance of the mixed algorithm they provide. The chosen values are as follows: $\mu_{LMS} = 4.5 \cdot 10^{-4}$, $\mu_{LMF} = \mu_{LMS+F} = 1.8 \cdot 10^{-3}$ for the uniformly distributed noise case (figure 1) and $\mu_{LMS} = \mu_{LMF} = \mu_{LMS+F} = 1.8 \cdot 10^{-3}$ for the gaussian and laplacian distributed noise cases in figures 2 and 3 respectively. The random disturbances are assumed to follow a uniform distribution with variance $\sigma_A^2 = 0.3$ in figure 1 and $\sigma_A^2 = 0.25$ in figures 2 and 3.

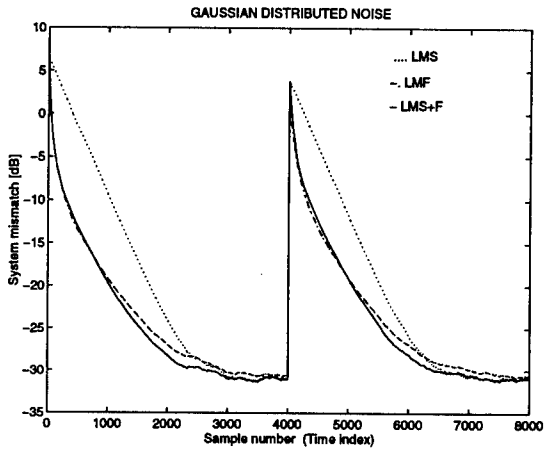


Figure 2: Performance comparison between the LMS (dotted line), the LMF (dash dot line) and the $LMS + F$ algorithm (solid line) under gaussian distributed noise conditions

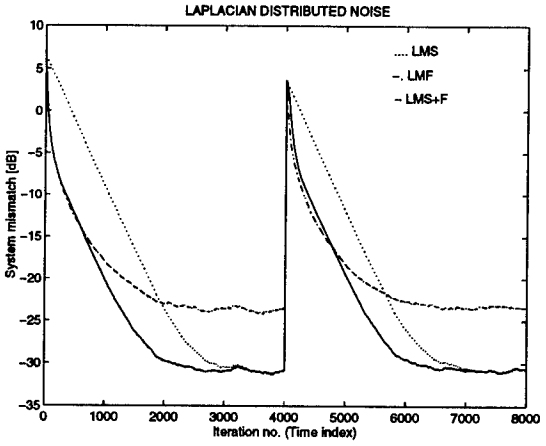


Figure 3: Performance comparison between the LMS (dotted line), the LMF (dash-dot line) and the $LMS + F$ algorithm (solid line) under laplacian distributed noise conditions

In figure 1 we see that the $LMS+F$ algorithm behaves as the LMF providing fast convergence and low steady state error. To achieve the same steady state error with the LMS algorithm, we had to significantly decrease the convergence factor μ.

In figures 2 and 3 we observe that the $LMS+F$ adopts the initial high tracking speed of the LMF algorithm but, as it approaches the optimum, it gradually changes to the LMS algorithm (laplacian noise case) or combines both criteria (gaussian noise case) to achieve a lower steady state error.

The adaptability of the $LMS+F$ algorithm to time varying noise distributions is shown in figure 4, where its performance is also compared with that of LMS and LMF. As in the previous figures, the $LMS+F$ exploits the fast initial convergence of the LMF algorithm and after that adopts the

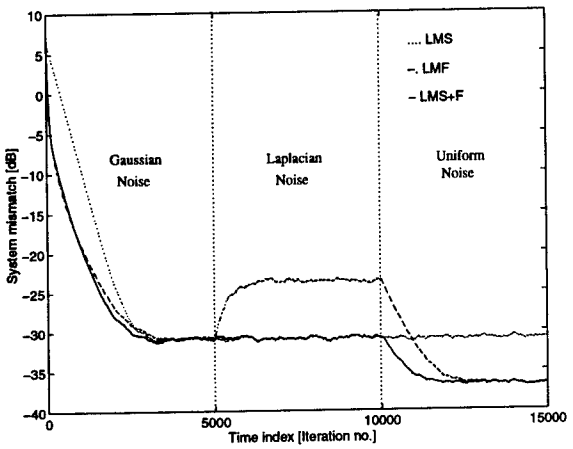


Figure 4: Performance curves of the algorithms under time varying noise conditions (LMS: dotted, LMF: dash dot, $LMS+F$: solid)

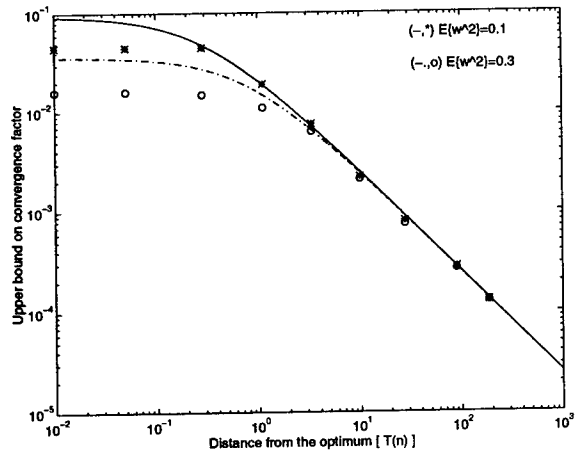


Figure 5: Comparison between the theoretically obtained upper bound on μ (solid ($\sigma_w^2 = 0.1$) and dashed ($\sigma_w^2 = 0.3$) lines) and and the ones obtained from simulations (stars ($\sigma_w^2 = 0.1$) and circles ($\sigma_w^2 = 0.3$)) ($\lambda = \delta = 1/2$).

performance criterion that exhibits the lower steady state error.

Next, we focus our attention on the theoretically obtained upper allowable value for the convergence factor, which is a function of the parameter λ. When λ is allowed to take any value in the closed interval $[0, 1]$ the convergence condition is dictated by that of LMF (i.e.,$\lambda = 1$), since it poses more stringent convergence bounds. However, if, for any reason, the λ_n sequence is constrained to $[0, \delta]$, where $\delta < 1$, then the necessary and sufficient condition is obtained from equation (17) by setting λ equal to δ. This latter case is depicted in figure 5, where a comparison between the expected (Eq.17) and observed upper bounds is presented for two values of noise variance and δ equal to $1/2$. It is easily observed that the simulation results are in good agreement with the theoretical ones, especially when far from the optimum. The observed deviation from the theoretic curves near the optimum is justified by noting that the theoretical results are obtained using averages, whereas in simulations, we deal with the instantaneous values of the stochastic variables. It can be also observed that, when far from the optimum, the condition on the step size is dominated by the distance T_n and is almost independent of the noise variance.

5. Conclusions

In this contribution, a new technique for mixing the LMS and LMF cost functions was presented. It differs from what was previously suggested in that it is time varying and takes into account the noise distribution. The proposed algorithm was analysed and conditions regarding the behaviour and stability were established. Simulation results verify the improved performance of the proposed algorithm and support well the theoretic results.

6. References

[1] J.A. Chambers, O. Tanrikulu, and A.G. Constantinides. Least mean mixed-norm adaptive filtering. *Electronics Letters*, 30(19):1574–1575, September 1994.

[2] S. C. Douglas and T. H.-Y. Meng. Stochastic Gradient Adaptation Under General Error Criteria. *IEEE Trans. Signal Processing*, 42(6):1335–1351, June 1994.

[3] J. D. Gibson and S. D. Gray. MSVE Adaptive Filtering Subject to a Constraint on MSE. *IEEE Trans. Circuits and Systems*, 35(5):603–608, May 1988.

[4] Simon Haykin. *Adaptive Filter Theory.* Englewood Cliffs : Prentice-Hall, 2nd edition, 1991.

[5] L. L. Horowitz and K.D. Senne. Performance advantage of complex LMS for controlling narrow-band adaptive arrays. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-29(3):722–735, June 1981.

[6] V. J. Mathews. Performance Analysis of Adaptive Filters Equipped with the Dual Sign Algorithm. *IEEE Trans. Signal Processing*, 39(1):85–91, January 1991.

[7] J.E. Mazo. On the Independence Theory of Equalizer Convergence. *The Bell System Technical Journal*, 58(5):963–993, 1979.

[8] D. I. Pazaitis and A. G. Constantinides. LMS+F Algorithm. *Electronics Letters*, 31(17):1423–1424, August 1995.

[9] D.I. Pazaitis. *Adaptive Signal Processing Techniques for Improved Convergence and Noise Robustness.* MPhil/PhD transfer report, Imperial College, March 1996.

[10] R. Price. A Useful Theorem for Nonlinear Devices Having Gaussian Inputs. *IRE Trans. Information Theory*, IT. 4:69–72, June 1958.

[11] O. Tanrikulu and A.G. Constantinides. Mean square convergence analysis and variable step-size of the LMF adaptive algorithm. *Submitted in Trans. Circuits and Systems*, 1995.

[12] E. Walach and B. Widrow. The Least Mean Fourth (LMF) Adaptive Algorithm and its Family. *IEEE Transactions on Information Theory*, 30(2), March 1984.

An Adaptive Hidden Markov Model/Kalman Filter Algorithm for Narrowband Interference Suppression with Applications in Multiple Access Communications*

A. Logothetis and V. Krishnamurthy

Dept.of Electrical and Electronic Eng.
University of Melbourne
Parkville, Victoria 3052
Australia

a.logothetis@ee.mu.oz.au v.krishnamurthy@ee.mu.oz.au

Abstract

This paper presents a novel nonlinear filter for narrowband interference suppression in multiple access communication systems, in particular code division multiple access spread spectrum. The proposed algorithm combines a recursive Hidden Markov Model estimator, Kalman filter and the recursive Expectation Maximization algorithm. It is shown that the proposed algorithm outperforms current linear and nonlinear filtering techniques, presented in Rusch and Poor [3].

1 Introduction

Code Division Multiple Access (CDMA), also known as Spread Spectrum Multiple Access (SSMA), provides a means of separating the signals of multiple users transmitting simultaneously and occupying the same RF bandwidth. In a Direct Sequence (DS) CDMA system each user has a distinct pseudonoise (PN) code (or sequence). The message from each user is modulated with the corresponding PN code, resulting in a transmission bandwidth much greater than the message bandwidth.

One of many reasons for spreading the spectrum is the inherent immunity of the communication system to interference. It is well known that system performance is greatly enhanced if the receiver employs some means of suppressing narrowband interference prior to signal "despreading". Such techniques are possible since they exploit the different nature of the received signals. The power spectrum of a narrowband interference is highly peaked, while the message signals are spread over a wide bandwidth.

Given T observations of the received noisy signal computing optimal estimates of the narrowband interference and finite-state spread spectrum signal, it is necessary to consider all N^T realizations of the N-state T-point spread spectrum signal. This is computationally prohibitive, thus the only feasible estimators are suboptimal.

Nonlinear suboptimal techniques for narrowband interference suppression in spread spectrum systems are presented in [1, 2, 3]. These offer improved performance over the time domain linear methods summarized in [1]. The narrowband interference is modeled as a Gaussian autoregressive process. Assuming known parameters, the interference signal is estimated using an Approximate Conditional Mean (ACM) filter [4]. For the specified assumptions on the observation process in [3], the ACM filter for interference estimation turns out to be a Kalman-type recursive filter with nonlinearities to deal with the finite state spread spectrum signal. The nonlinearities take the form of a soft decision feedback which seeks to remove the spread spectrum signal from the estimation of the narrowband interference. For the general case of unknown interference statistics the soft decision feedback was incorporated into a LMS adaptive filter [2]. A modification and an enhancement to this approach is presented in [3], where multiple spread spectrum users are considered. The proposed Enhanced Nonlinear Adaptive al-

*Partially supported by the Australian Telecommunications and Electronics Research Board (ATERB), Australian Research Council (ARC) and the Co-operative Research Centre for Sensor Signal and Information Processing (CSSIP)

0-8186-7576-4/96 $5.00 © 1996 IEEE

gorithm (ENA) of [3] shows a significant improvement over existing linear and nonlinear adaptive filters.

In this paper, we present a new nonlinear filter and parameter estimator for narrowband interference suppression in spread spectrum systems. We consider a slightly more general signal model than [3]: In particular, we model the received sampled signal as the sum of the spread spectrum signal (modeled as a finite state Markov chain), narrowband interference (modeled as a Gaussian autoregressive (AR) process) and observation noise (modeled as a zero mean white Gaussian process).

There are at least two reasons that justify modeling the spread spectrum signal as a finite state Markov chain: i) Asynchronous multi user transmission and ii) Oversampling. Each of these cases induce correlation in the received spread spectrum signal, and hence the first-order Markov chain assumption is more realistic than the iid assumption. In examples where the above two reasons do not hold and the spread spectrum signal is iid, simulations show that our algorithm still performs as well as or better than the ENA/ACM algorithm in [3]. Moreover, our algorithm has a comparable computational cost to the algorithms in [3].

Highlights of our HMM-KF algorithm

Our algorithm cross-couples two optimal filters - a Hidden Markov Model (HMM) filters and a Kalman Filter (KF). As described below, together with the recursive EM algorithm, these coupled filters yield estimates of the narrowband interference, spread spectrum signal and their parameters. We call our algorithm the HMM-KF algorithm.

Methedology: We first explain the motivation for our algorithm: If the narrowband interference is exactly known at a given time, then the estimation task reduces to the problem of extracting the finite-state Markov chain (spread spectrum) in additive white Gaussian observation noise. The Markov chain can be optimally extracted using the well known Hidden Markov Model filter. Furthermore, on-line (or adaptive) parameter estimates of the HMM (including transition probabilities and noise variances) can be obtained via the recursive EM algorithm presented in [5].

On the other hand, if the spread spectrum signal is exactly known at a given time, then the estimation problem reduces to the problem of extracting a Gaussian autoregressive process (narrowband interfence) embedded in white Gaussian noise. Then optimal state estimates of the narrowband interference are obtained using a Kalman filter [5]. Furthermore, on-line parameter estimates (autoregressive coefficients, process and observation variances) can be achieved via the recursive EM algorithm [5, 6].

Since we do not have an exact knowledge of the narrowband interference or the spread spectrum signal, we propose the following scheme: cross-couple the above two recursive EM algorithms, one algorithm for the HMM and the other for the noisy AR model. The proposed algorithm is called the HMM-KF algorithm. It is schematically shown in Fig.1.

2 Problem Formulation

We now present the signal model in detail and state our estimation objectives.

2.1 A Model for the Received Signal

We assume a similar spread spectrum and narrowband interference signal model to [2]. In particular, assume that the continuous-time received signal $y(t)$ consist of the spread spectrum signal $s(t)$ from all N users, the narrowband interference $i(t)$ and the white Gaussian observation noise $n(t)$. That is: $y(t) = s(t) + i(t) + n(t)$. If $y(t)$ is sampled at chip rate of the PN sequence, the resulting discrete time observations [3] are:

$$y_k = s_k + i_k + n_k \tag{1}$$

where n_k is a zero mean white Gaussian process with variance σ_n^2 and s_k is a discrete-state iid process.

The unknown narrowband interference i_k is modeled as a Gaussian autoregressive process of known order p, which can be written as

$$i_k = -d_1 i_{k-1} - \cdots - d_p i_{k-p} + e_k \tag{2}$$

where e_k is white Gaussian noise, independent of s_k and n_k, with zero mean and variance σ_e^2.

In [3], the spread spectrum signal s_k is the sum of N independent and identically distributed (iid), equiprobable, binary random variables. This is so, since the user message and the PN sequence are assumed purely random. Furthermore, it is also assumed that each user is received at the same (normalized) power and are chip synchronous.

In this paper we assume that the received spread spectrum signal $s(t)$ are sampled at a rate higher than the chip rate of the PN sequence. This yields samples which are correlated in time. Hence, we assume s_k is a finite-state, discrete-time, homogeneous, first-order Markov chain. Consequently, the state s_k at time k is one of a finite number M of states $q = (q_1, q_2, \ldots, q_M)$. The transition probability matrix is $A = (a_{mn})$ where $a_{mn} = P(s_{t+1} = q_n | s_t = q_m)$ and $m, n \in \{1, \ldots, M\}$. Of course $a_{mn} \geq 0$, $\sum_{n=1}^{N} a_{mn} = 1$, for each m. Let π

denote the initial state probability vector: $\pi = (\pi_m)$, $\pi_m = P(s_1 = q_m)$.

Remark : The number of levels M are equal to $N+1$ when all N users are received at the same power. If each user transmits ± 1, then $q_j = -N + 2(j-1), j \in \{1, \ldots, N+1\}$. Increasing the sampling rate will yield a transition probability matrix A with an increasing diagonally dominant elements. Setting the sampling rate to the chip rate yields the signal model in [3]. This is merely a special case of our signal model with all the rows in A identical.

2.2 Estimation Objectives

Let $\phi_0 = (A, q, D, \sigma_e^2, \sigma_n^2)$ denote the true parameter vector that characterizes the narrowband interference (AR signal) and the spread spectrum signal (Markov chain).

Given the observations $Y_k = (y_1, \ldots, y_k)$, our aim is twofold:

1. **State Estimation**: Compute estimates of the narrowband interference i_k and and the spread spectrum signal s_k.
2. **Parameter Estimation**: Derive a recursive estimator $\phi^{(k)}$ for ϕ_0, where $\phi^{(k)} = (A^{(k)}, q^{(k)}, D^{(k)}, \sigma_e^{2(k)}, \sigma_n^{2(k)})$, for $k > 1$, given the observations Y_k.

Why estimate A and q? When the received power levels are time-varying and asynchronous data transmission is used, A and q may not be known apriori. Indeed, A is a complicated function of the number of levels M, the transmitted power of each user, sampling rate and transmission mode (synchronous or asynchronous). Hence the motivation for estimating A and q.

Remark: We assume that the number of states M of the Markov chain is known. Also, for convenience we assume that $\pi_m = 1/M$, for $m = 1 \ldots, M$,

From (1) and (2) it is quite clear that estimating optimal (maximum a posteriori MAP) state estimates of i_k and s_k and computing optimal (maximum likelihood ML) parameter estimates of the signal model is computationally infeasible since the computational cost is exponential in the data length. In the following subsection we present our sub-optimal nonlinear algorithm for narrowband interference suppression, which combines two optimal filters to achieve both state and parameter estimation.

3 HMM-KF Narrowband Interference Suppression Algorithm

In this section we present our HMM-KF algorithm. The HMM-KF algorithm cross couples two recursive EM algorithms, one algorithm for a HMM and the other for a noisy AR model. The algorithm is schematically shown in Fig.1:

1. At time k, the Kalman filter and recursive EM parameter estimator for the narrowband interference yield estimates of the state of i_k, process noise variance σ_e^2, observation noise variance σ_n^2, and the AR coefficients $d_1, \ldots, d_p$.
2. The Hidden Markov Model filter and recursive EM parameter estimator for the spread spectrum signal gives on-line estimates of the state of s_k, transition probability matrix A and Markov chain levels q.

We now present Steps 1 and 2 in more detail.

3.1 Spread Spectrum Signal Estimator Using Recursive HMMs

At time k, we have available the predicted narrowband interference $i_{k|k-1}$, and variance $p_{i_{k|k-1}}$ of the predicted error $w_k \stackrel{\Delta}{=} i_k - i_{k|k-1}$ obtained from the KF of Sec. 3.2. Therefore, the HMM to be estimated is:

HMM Signal Model:

$$y_k - i_{k|k-1} = s_k + w_k + n_k \qquad (3)$$

We assume that the Kalman predicted error w_k is modeled as a zero mean white Gaussian process with variance $p_{i_{k|k-1}}$ and is independent of the observation noise n_k.

The recursive HMM estimator recursively updates the state and parameter estimates of the HMM. The recursive HMM parameter vector estimate at k is denoted as: $\phi_{\text{HMM}}^{(k)} \stackrel{\Delta}{=} (A^{(k)}, q^{(k)}, \sigma_n^{2(k)})$.

Given the signal model (3), the state and adaptive parameter estimation procedure for the spread spectrum signal s_k is straightforward. Details can be found in [7].

3.2 Recursive Narrowband Interference Estimation Using Kalman Filtering

The HMM estimator described in Sec. 3.1 yields filtered state estimates $s_{k|k}$ of the spread spectrum signal. Given the spread spectrum signal estimate $s_{k|k}$ and the associated error variance $p_{s_{k|k}}$ of $w_k \stackrel{\Delta}{=} s_k - s_{k|k}$, our aim now is to compute state and parameter estimates of the narrowband interference. The signal model is given by:

$$y_k - s_{k|k} = i_k + w_k + n_k \qquad (4)$$

where y_k is the observation and i_k is the narrowband interference signal. $w_k \sim N(0, p_{s_{k|k}})$ is modeled as a zero mean white Gaussian process with variance $p_{s_{k|k}}$

and assumed independent of the observation noise $n_k \sim N(0, \sigma_n^2)$ and process noise $e_k \sim N(0, \sigma_e^2)$. (4) can be represented as the following state space model:

State Space Model:

$$\begin{aligned} x_k &= Fx_{k-1} + Ge_k \\ y_k - s_{k|k} &= Hx_k + w_k + n_k \end{aligned} \quad (5)$$

where the state vector $x_k = (i_k, i_{k-1}, \ldots, i_{k-p})'$,

$$\begin{aligned} F &= \begin{pmatrix} -D' & 0 \\ I_{p\times p} & 0_{p\times 1} \end{pmatrix}, \; D = (d_1, \ldots, d_p)', \\ G &= (1\; 0_{1\times p})', \; H = (1\; 0_{1\times p}), \end{aligned} \quad (6)$$

The recursive EM estimator recursively updates the narrowband interference autoregressive coefficients, the narrowband interference process noise and observation noise. The recursive EM parameter estimate at k is denoted as: $\phi_{\mathrm{KF}}^{(k)} \triangleq (D^{(k)}, \sigma_e^{2(k)}, \sigma_n^{2(k)})$.

Given the signal model of (5), the state and adaptive parameter estimation procedure for the narrowband interference i_k is straightforward. Details can be found in [7].

4 Conclusion

We have presented a new solution to the problem of narrowband interference suppression in CDMA spread spectrum communication systems. Nonlinear Hidden Markov Model (HMM) signal processing techniques together with a Kalman Filter (KF) was used to derive a high-performance algorithm for suppressing the narrowband interference and to simultaneously yield estimates of the spread spectrum signal. Computer simulations show that the proposed algorithm outperforms current linear and other nonlinear filtering techniques. Although, the algorithm is difficult to analyze theoretically, simulation studies in [7] have shown satisfactory estimates in several cases.

References

[1] H.V. Poor and L.A. Rusch, *Narrowband Interference Suppression in Spread Spectrum CDMA*, IEEE Personal Communications, Third Quarter 1994.

[2] R. Vijayan and H.V. Poor, *Nonlinear Techniques for Interference Suppression in Spread-Spectrum Systems*, IEEE Transactions on Communications, Vol.38, No.7, pp. 1060-1065, July 1990.

[3] L.A. Rusch and H.V. Poor, *Narrowband Interference Suppression in CDMA Spread Spectrum Communications*, IEEE Transactions on Communications, Vol.42, No.2/3/4, pp. 1969-1979, February/March/April 1994.

[4] C.J. Masreliez, *Approximate Non-Gaussian Filtering with Linear State and Observation Relations*, IEEE Transactions on Automatic Control, pp. 107-110, February 1975.

[5] V. Krishnamurthy, J.B. Moore, *On-line Estimation of Hidden Markov Parameters based on the Kullback-Leibler Information Measure*, IEEE Trans on Signal Processing, Vol 41, No 8, pp. 2557-2573, August, 1993.

[6] E. Wienstein, A.V. Oppenheim, M. Feder and J.R. Buck, *Iterative and Sequential Algorithms for Multisensor Signal Enhancement*, IEEE Trans. on Signal Processing, Vol.42, No.4, pp 846-859, April 1994.

[7] A. Logothetis and V. Krishnamurthy, *Adaptive Nonlinear Filters for Narrowband Interference Suppression in Spread Spectrum CDMA Systems*, submitted to IEEE Trans. on Communications.

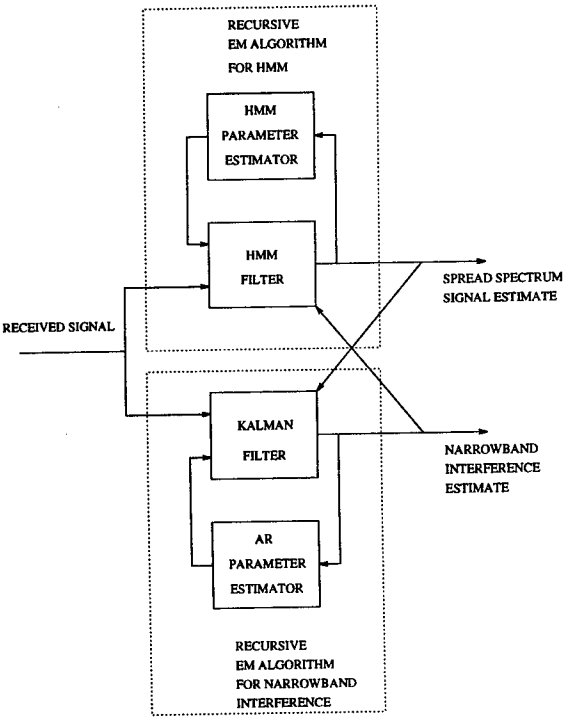


Figure 1. Proposed Adaptive Narrowband Interference Suppression Algorithm.

FIR Filters for Blind Detection of CDMA Signals

Dimitris A. Pados
Dept. of Electrical and Computer Engineering
University of Louisiana
Lafayette, LA 70504-3890
E–mail:pados@usl.edu

Stella N. Batalama
Dept. of Electrical and Computer Engineering
State University of New York at Buffalo
Buffalo, NY 14260
E–mail:batalama@eng.buffalo.edu

Abstract

A new blind FIR filter receiver is proposed for the detection of DS-CDMA signals in unknown Multi-User Interference (MUI) and Additive White Gaussian Noise. The proposed receiver is motivated by the Wiener signal reconstruction theory and it is a very low complexity alternative to the Minimum-Output-Energy (MOE) receiver [8]. At only a minimal increase of the computational cost, the proposed detector outperforms significantly the conventional Matched Filter (MF) receiver. It also compares favorably to the decorrelating detector [2] (with similar near-far resistance), despite the fact that the latter utilizes the assumed known MUI spreading codes. The novel characteristic of the proposed receiver is the incorporation of an auxiliary vector component that allows statistically optimal, adaptive steering of the filter with respect to the incoming DS-CDMA signal.

1. Introduction

Spreading the spectrum of a signal to make it virtually indistinguishable from background noise has served as the basic principle that led to the development of spread spectrum communication systems. The main motivation for the development of such systems emerged from military communication needs to ensure effective suppression of intentional interference as well as to increase security in signal transmission. Currently, Direct-Sequence Code Division Multiple Access (DS-CDMA), a specific form of spread spectrum transmission, is receiving considerable interest in response to an ever-increasing demand for better utilization of the available resources in mobile radio and personal communication environments.

While the overall capacity of a CDMA system is determined by both the forward (base-to-mobile) and the reverse (mobile-to-base) link, most of the recent research focused on the reverse connection and dealt with processing at the base station under the assumption of known active user population. Given the unrealizable complexity and the prohibitive computational requirements that the optimal multiuser detector exhibits [1], any proposal for a suboptimal reduced complexity receiver is well justified. Arguably, the list of such successful proposals includes the decorrelating receiver [2], multistage architectures [3]-[5], and decision feedback detectors [6].

In cellular systems, multiuser detection is envisioned at the base station for the simultaneous recovery of all signals of the known intracell users. While this is the situation in the reverse link, in the forward link the mobile user faces an even more challenging problem, namely the detection of its own signal in the presence of *unknown* multiuser interference and additive white Gaussian channel noise. In addition, processing at the mobile should meet even tighter complexity, size, and weight requirements than the base station. The handy matched filter (MF) solution exhibits unacceptable performance degradation in the presence of one or more high-power interferers (the "near-far" problem [1]). Therefore, it can only be used with some form of stringent and costly power control. Recently two interesting linear-filter alternatives were proposed in the form of a Minimum-Mean-Square-Error (MMSE) filter that requires a separate training sequence [7] and a blind "minimum energy" linear receiver [8]. If L is the system processing gain (as high as 127), then the latter blind receiver may require frequent inversion of the LxL sample autocorrelation matrix of the received signal. Equivalently, the adaptive implementation requires steepest descent in the R^L space.

In this paper we reconsider the issues of single-user detection in *unknown* spread-spectrum MUI and AWGN from the Wiener signal reconstruction viewpoint and we propose a new blind low-complexity receiver for the forward link of DS-CDMA communication systems. We examine low-complexity alternatives to the work

This work was supported in part by the LEQSF contract LEQSF(1995-97)-RD-A-35

0-8186-7576-4/96 $5.00 © 1996 IEEE

presented in [8] that maximize the Signal - to - Interference - plus - Noise - Ratio (SINR) [10]. The proposed receiver is a very low complexity alternative to the MOE detector and at only a minimal increase of the computational cost outperforms significantly the conventional MF detector. It also compares favorably to the decorrelating detector (with similar near-far resistance), although the latter utilizes the assumed *known* MUI spreading codes.

2. System Modeling

Although the results that follow are directly applicable to the asynchronous case, for the sake of brevity and clarity of presentation we choose to present our work in the context of synchronous CDMA. We consider a CDMA system where K users transmit synchronously over an AWGN channel. The continuous-time received signal is modeled as follows:

$$r(t)=\sum_{i}\sum_{k=0}^{K-1}\sqrt{E_k}\,b_k(i)s_k(t-iT)+n(t)\,, \qquad (1)$$

where for the *k-th* user E_k is the *received* energy, $b_k(i)\,\varepsilon\{-1,1\}$ is the *i-th* information bit, and $s_k(t)$ is the signature (spreading code). T is the symbol (bit) period and $n(t)$ is the channel AWGN. The signature of every user is composed of L spreading chips and it is of the form $s_k(t)=\sum_{j=1}^{L}c_k(j)P_{T_c}[t-(j-1)T_c]$, where L is the so called *system processing gain*, $c_k(j)\,\varepsilon\{-1,1\}$, $j=1,\cdots,L$ are the assigned signature bits for the *k-th* user, and $P_{T_c}(t)$ is the spreading pulse with duration $T_c=T/L$. Without loss of generality the signatures are assumed to be normalized. Since we consider synchronous transmission we concentrate on a single information bit interval of length T. The R^L discrete-time version of (1) is

$$\mathbf{r}=\sum_{k=0}^{K-1}\sqrt{E_k}\,b_k\mathbf{S}_k+\mathbf{n}\,. \qquad (2)$$

The random vector $\mathbf{n}$ is assumed to be WG with autocorrelation matrix $E\{\mathbf{n}^T\mathbf{n}\}=\sigma^2 1_{LxL}$. Assuming that the user of interest is user o, it is convenient to define the multiuser interference (MUI) term $\mathbf{I}=\sum_{k=1}^{K-1}\sqrt{E_k}\,b_k\mathbf{S}_k$. This allows us to write (2) as follows:

$$\mathbf{r}=\sqrt{E_o}\,b_o\mathbf{S}_o+\mathbf{I}+\mathbf{n}. \qquad (3)$$

3. Blind Low-Complexity Detectors

In this section we use the theory of Wiener signal-reconstruction filters to derive simple linear receivers that maximize the output SINR. Before we proceed with the proposed scheme, we opt to present a theorem that identifies a Wiener reconstruction filter with an inherent MUI cancellation property.

Let $\{\mathbf{S}_o,\mathbf{G}_1,\cdots,\mathbf{G}_{L-1}\}$ be a set of orthonormal vectors in R^L (an orthonormal basis of R^L that includes $\mathbf{S}_o$). For some arbitrary scalar $k\neq 0$ and $\forall\mathbf{G}_i$ we define

$$\mathbf{G}_i^*=\mathbf{G}_i+k\mathbf{S}_o\,,\ \ i=1,\cdots,L-1\,. \qquad (4)$$

Theorem 1 [10]: Let $\mathbf{S}_o$ and $\mathbf{G}_1^*,\mathbf{G}_2^*\cdots,\mathbf{G}_{L-1}^*$ defined by (4) for some $k\neq 0$ be vectors in the R^L space used as input data sequences for the Wiener reconstruction of the received signal vector $\mathbf{r}$ in (2) with corresponding tap-weights $w_0,w_1,\cdots,w_{L-1}$. Then,
(i) the optimal weighting coefficient w_o is

$$w_o=E_n\{<\mathbf{r},\mathbf{S}_o-k\sum_{i=1}^{L-1}\mathbf{G}_i>\}\,,\ \text{and} \qquad (5)$$

(ii) for any instance of the received signal $\mathbf{r}$ the filter $<\mathbf{r},\mathbf{S}_o-k\sum_{i=1}^{L-1}\mathbf{G}_i>$ cancels completely all MUI vectors $\mathbf{I}$ within the interference subspace V_I spanned by $\{\mathbf{G}_1^*,\cdots,\mathbf{G}_{L-1}^*\}$. □

Part (ii) of Theorem 1 motivates the proposal for a detector of the form

$$\hat{b}_o=\mathrm{sgn}(w_o)\,. \qquad (6)$$

The following proposition places the conventional MF and the decorrelating receiver in the Wiener filter context.

Proposition 1: (i) The matched filter (MF) receiver $<\mathbf{r},\mathbf{S}_o>$ is the result of optimal *single-tap* Wiener reconstruction of the received signal vector $\mathbf{r}$ with input data sequence $\mathbf{S}_o$. *(ii)* In a $K-$user CDMA environment, the decorrelating receiver is the result of optimal $K-$tap Wiener reconstruction of the received signal vector $\mathbf{r}$ with input data sequences the user spreading codes $\mathbf{S}_o,\mathbf{S}_1,\cdots,\mathbf{S}_{K-1}$, that are all assumed known. □

Remarks: 1) Perfect interference cancellation for any *arbitrary* parameter $k\neq 0$ is guaranteed *only* for interference vectors $\mathbf{I}$ on the V_I hyper-plane. Cancellation of interferers $\mathbf{I}$ not on V_I requires tuning of the parameter $k\neq 0$. 2) The size (length) L of the adaptive Wiener reconstruction implementation may render the proposed detector unrealizable for a mobile user. Therefore, careful consideration of lower size Wiener reconstruction filters appears well motivated. 3) A result similar to part (ii) of Proposition 1 was given in [9] and the statistically equivalent context of Least Squares.

Following our notation, let $\mathbf{G}$ be an arbitrary normalized vector in R^L orthogonal to the spreading code of the user of interest $\mathbf{S}_o$. In other words

$$<\mathbf{G},\mathbf{S}_o>=0 \quad \text{and} \quad <\mathbf{G},\mathbf{G}>=1 \,. \tag{7}$$

The following result is a direct corollary to Theorem 1:

Corollary 1: Let $\mathbf{S}_o$ and $\mathbf{G}$ be as in (7) and let k be a non-zero scalar. Consider the 2-tap Wiener reconstruction filter with *input data sequences* $\mathbf{S}_o$ and $\mathbf{G}+k\mathbf{S}_o$, and corresponding tap-weights w_o and w_1. Then,
(i) the MS optimal value of w_o is

$$w_o = E_n\{<\mathbf{r},\mathbf{S}_o - k\mathbf{G}>\} \,, \text{ and} \tag{8}$$

(ii) the linear filter $\mathbf{S}_o - k\mathbf{G}$ cancels all MUI vectors $\mathbf{I}$ in the direction of $\mathbf{G}+k\mathbf{S}_o$. □

Again, the proposed decision statistic is the tap-weight w_o itself and the detector is $\hat{b}_o = \text{sgn}(w_o)$ as in (6). Sample average can be used in place of the expectation with respect to $\mathbf{n}$ in (8) assuming that multiple samples of $\mathbf{r}$ are available. Therefore, for notational simplicity we can drop the expectation from (8) without any loss of generality. With input $\mathbf{r}$ given by (3), the output of the filter in (8) is

$$<\mathbf{r},\mathbf{S}_o - k\mathbf{G}> = \sqrt{E_o}\, b_o + <\mathbf{I},\mathbf{S}_o> - k<\mathbf{I},\mathbf{G}> +$$

$$+ <\mathbf{n},\mathbf{S}_o - k\mathbf{G}> \,. \tag{9}$$

Then for fixed *auxiliary* vector $\mathbf{G}$ the *average* variance of the output (the expectation is taken with respect to both b_o, $\mathbf{I}$, and $\mathbf{n}$) is

$$E\{<\mathbf{r},\mathbf{S}_o - k\mathbf{G}>^2\} = E_o + E\{[<\mathbf{I},\mathbf{S}_o> - k<\mathbf{I},\mathbf{G}>]^2\} +$$

$$+ (1+k^2)\sigma^2 \,. \tag{10}$$

While the MUI cancellation property of the receiver described in part (ii) of Corollary 1 is unsatisfactory, (9) shows that we can still succeed in canceling effective MUI if we choose the *auxiliary* vector $\mathbf{G}$ to be the average (normalized) projection of $\mathbf{r}$ onto the subspace orthogonal to $\mathbf{S}_o$ and we use k as a steering parameter that places the filter $\mathbf{S}_o - k\mathbf{G}$ orthogonally to the interference vector $\mathbf{I}$. Moreover, (10) shows that classical blind minimum variance optimization of the filter $\mathbf{S}_o - k\mathbf{G}$ (which is distortionless in the direction of interest $\mathbf{S}_o$) leads in fact to a maximum Signal to Interference plus Noise Ratio (SINR) receiver. The following proposition optimizes the filter with respect to the scalar k. The auxiliary vector $\mathbf{G}$ is defined immediately after.

Proposition 2: If $\mathbf{S}_o$ is the spreading code of the user of interest and $\mathbf{G}$ is some *auxiliary* vector with realizations constrained by (7), then the value of the steering scalar k that minimizes the variance expression $E\{<\mathbf{r},\mathbf{S}_o - k\mathbf{G}>^2\}$ (maximizes the average output SINR) is

$$k_{MVDR} = E\{<\mathbf{r},\mathbf{S}_o><\mathbf{r},\mathbf{G}>\} / E\{<\mathbf{r},\mathbf{G}>^2\} \,. \; \square \tag{11}$$

The auxiliary vector $\mathbf{G}$ with realizations constrained by (7) is -within a sign ambiguity- the average normalized projection of the received signal vector $\mathbf{r}$ onto the subspace orthogonal to the spreading code $\mathbf{S}_o$. The sign of this projection can be either $\text{sgn}(<\mathbf{r},\mathbf{S}_o>)$ or $-\text{sgn}(<\mathbf{r},\mathbf{S}_o>)$. Therefore, without loss of generality, if we write

$$\mathbf{r}_{IS_o} = sgn(<\mathbf{r},\mathbf{S}_o>) \frac{\mathbf{r} - <\mathbf{r},\mathbf{S}_o>\mathbf{S}_o}{\sqrt{||\mathbf{r}||^2 - <\mathbf{r},\mathbf{S}_o>^2}} \,, \tag{12}$$

then

$$\mathbf{G} = E\{\mathbf{r}_{IS_o}\} / ||E\{\mathbf{r}_{IS_o}\}|| \,. \tag{13}$$

The proposed receiver is completely defined by equations (6), (8), (11), (12), and (13). In the next section we present some numerical results and comparisons that support our theoretical arguments.

4. Numerical Results and Simulations

We examine a scenario of four users each equipped with a signature of length $L=15$ and signature cross correlation matrix given by

$$R_S = \frac{1}{\sqrt{15}} \begin{bmatrix} 15 & 11 & 11 & 11 \\ 11 & 15 & 7 & 7 \\ 11 & 7 & 15 & 7 \\ 11 & 7 & 7 & 15 \end{bmatrix} . \tag{14}$$

We compare the Bit Error Rate (BER) performance of the input-driven auxiliary-vector receiver with the conventional matched filter, the MOE, and the decorrelating receiver for synchronous CDMA transmission over an AWGN channel. The results are shown in Fig. 1 and Fig. 2. Near-far resistance comparisons are shown in Fig. 3.

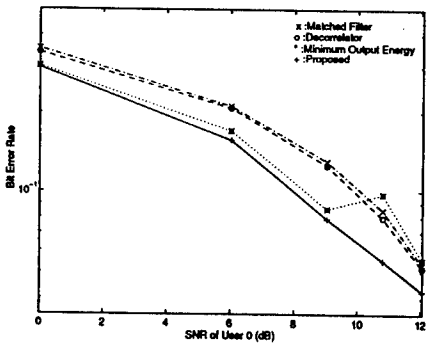


Fig. 1: Bit error rate as a function of the SNR of the user of interest in the presence of weak interferers (SNR_1=2, SNR_2=3, SNR_3=4).

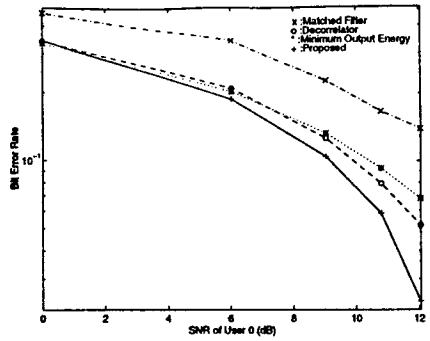


Fig. 2: Bit error rate as a function of the SNR of the user of interest in the presence of strong interferers (SNR_1=8, SNR_2=9, SNR_3=10).

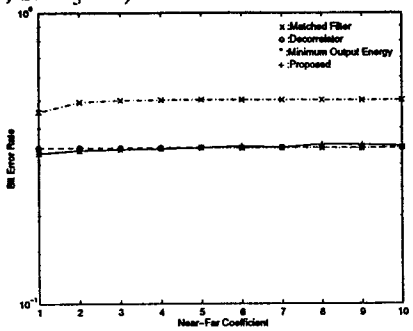


Fig. 3: Bit error rate as a function of the Near-Far coefficient (SNR_1=1, SNR_1=8×*NFC*, SNR_2=9×*NFC*, SNR_3=10×*NFC*).

V. Conclusions

We reconsidered the concept of multiuser detection for DS/CDMA communication systems from the point of view of Wiener signal reconstruction filters. We identified both the decorrelating detector and the signature matched filter receiver as a direct special case of Wiener signal reconstruction. Generalizing this result we proposed an L-tap Wiener adaptive receiver with a powerful inherent MUI canceling property. However, the size of the filter (L taps where L is the system processing gain) appears prohibitive and may restrict severely the practicality of this approach. In view of these observations the natural, low complexity outcome of this line of work is a linear, scalar parameterized, auxiliary-vector receiver. The conceptual and computational simplicity of this receiver promises some immediate practical utility. The optimization can be carried out easily in a variety of different ways. In this work we chose to develop a blind (unsupervised) solution that maximizes the output Signal-to-Interference-plus-Noise Ratio (SINR). In future work we will consider optimization in the minimum probability of error sense (non-least-squares supervised learning [11]).

The newly proposed blind auxiliary-vector receiver compares favorably, both complexity-wise and performance-wise, to the decorrelating detector [2], although the latter utilizes the assumed known signatures of the interferers. This is because the blind maximum SINR criterion, in contrast to the "decorrelating" criterion, strives to achieve the perfect balance between MUI and channel noise suppression. The optimal near-far resistance of the decorrelator appears closely matched by the auxiliary-vector receiver over a wide range of realistic near-far ratios.

In view of these results, the linear, blind, auxiliary-vector filter becomes a candidate for the receiver of choice for the forward link of mobile cellular DS-CDMA communication systems. On the other hand, a bank of blindly optimized auxiliary-vector filters may be deployed as the reverse-link, base-station receiver.

References

[1] S. Verdu, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inform. Theory,* vol.32, pp. 85-96, Jan. 1986.

[2] R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory,* vol. 35, pp. 123-136, Jan. 1989.

[3] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple-access communications," *IEEE Trans. Commun.,* vol. 38, pp. 509-519, April 1990.

[4] M. K. Varanasi and B. Aazhang, "Near-optimum detection in synchronous code-division multiple-access systems," *IEEE Trans. Commun.,* vol. 39, pp. 725-736, May 1991.

[5] Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Select. Areas Commun.,* vol. 8, pp. 683-690, May 1990.

[6] A. Duel-Hallen, "A family of multiuser decision-feedback detectors for asynchronous code-division multiple access channels," *IEEE Trans. Commun.,* vol. 43, pp. 421-434, Feb./March/Apr. 1995.

[7] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.,* vol. 42, pp. 3178-3188, Dec. 1994.

[8] M. L. Honig, U. Madhow, and S. Verdu, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory,* vol. 41, pp. 944-960, July 1995.

[9] D. S. Chen and S. Roy, "An adaptive multiuser receiver for CDMA systems," *IEEE J. Select. Areas Commun.,* vol. 12, pp. 808-816, June 1994.

[10] D. A. Pados and S. N. Batalama, "Low complexity blind detection of DS/CDMA signals: Auxiliary-vector receivers," *IEEE Trans. Commun.,* Oct. 1995, submitted.

[11] D. A. Pados and P. Papantoni-Kazakos, "New nonleast-squares neural network learning algorithms for hypothesis testing," *IEEE Trans. Neural Networks,* vol. 6, pp. 596-609, May 1995.

Performance Analysis on the Linear Minimum Mean Squared Error (LMMSE) Estimate-based Multiuser Detectors for CDMA Communications *

Hongya Ge Y. Bar-Ness
Center for Communications and Signal Processing Research
New Jersey Institute of Technology
University Heights, Newark, NJ 07102

Abstract

In this work, we analyze the performance of linear minimum mean squared error (LMMSE) estimate-based multiuser detector for CDMA communication systems. Remarkable consistency is achieved through numerical evaluation of our analytical results and computer simulations. We also compare the performance of the LMMSE detector [1] *and the adaptive bootstrap multiuser detector* [2]. *Our analyses and simulations show that even though these two detectors were proposed based on different optimization criteria, they exhibit approximately equal performance in multiuser CDMA communication applications.*

1. Introduction

Multiuser separation and interference suppression is an active research topic in CDMA communications area. Various detectors have been proposed to balance the computational simplicity and reliable detection performance [1 – 6]. Most of the proposed detectors treat the multiuser signal vector as a deterministic vector, and work only on the *conditional* probability density function (PDF) of the data given the multiuser signal vector. We notice that the information bit of multiuser is actually a *random* vector with known statistics. Therefore, by incorporating prior knowledge (in a statistical sense) of the multiuser signal vector, one can always improve the overall detection performance [6]. The availability of the prior knowledge depends on the specific communication systems. If we constrain our detector to be in the linear class, we can find, as proposed in [1] and not accurately termed as minimum mean squared error (MMSE) detector, the sub-optimal *linear* MMSE (LMMSE) detector. In this paper, we further analyze the performance of the LMMSE detector in detail; verify our results through numerical evaluation and computer simulations. Comparative study of performances of the linear class of decorrelating detector, LMMSE detector, and adaptive bootstrap multiuser detector is also provided in this work.

*This work was supported in part by the Office of Sponsored Research, NJIT, and Rome Air Force Lab under contract F30602-94-C-0135.

2. Problem Formulation

Due to the multiple access (MA) scheme used in CDMA systems, the data $r(t)$ available at the receiver is actually a mixture of multiuser data embedded in additive noise. That is,

$$r(t) = \sum_i \sum_{k=1}^{K} \sqrt{a_k(i)}\, b_k(i)\, s_k(t - iT - \tau_k) + n(t)\,, \quad (1)$$

where $a_k(i)$, $b_k(i)$, $s_k(t)$, and τ_k are the bit energy, information bit, signature waveform, and the transmission delay of the kth user in the ith bit symbol interval (of duration T), respectively. $n(t)$ is a white Gaussian process, with two-sided power spectral density of σ^2.
In this work, we consider the case when $\tau_k = 0$ ($k = 1, \ldots, K$), which corresponds to the synchronous channel. The synchronous channel model is valid for down-link channel (from base station to mobiles). Note that once the channel is synchronized, all the information bits of the multiusers in the ith symbol interval are completely contained in the data $r(t)$ within the ith symbol interval. Therefore, we can concentrate on solving multiuser separation problem within a specific symbol interval, say $i = 0$ th, without lose of generality. After ignoring index i, we rewrite data in (1) into a matrix form,

$$r(t) = \mathbf{S}^T(t) \cdot \mathbf{A} \cdot \mathbf{b} + n(t)\,, \qquad 0 \le t \le T\,. \quad (2)$$

with $\mathbf{S}^T(t) = [\ s_1(t)\ s_2(t)\ \cdots\ s_K(t)\ ]$;
$\mathbf{A} = \mathrm{diag}\,\{\sqrt{a_1},\ \sqrt{a_2},\ \cdots,\ \sqrt{a_K}\}$
and $\mathbf{b} = [\ b_1\ b_2\ \cdots\ b_K\ ]^T$.

At the receiver end, we first filter data $r(t)$ with a bank of matched filters, whose impulse responses are given by

0-8186-7576-4/96 $5.00 © 1996 IEEE

$h_k(t) = s_k(T-t)$, $(k = 1, 2, \ldots, K)$. We then stack the outputs of the bank of matched filters, sampled at $t = T$, into a vector and get the following matrix notation,

$$\mathbf{x} = \mathbf{P} \cdot \mathbf{A} \cdot \mathbf{b} + \mathbf{n}, \tag{3}$$

where $\mathbf{x} = [\, x_1(T)\ x_2(T)\ \cdots\ x_K(T)\,]^T$, with $x_k(T) = r(t) * h_k(t)\,|_{t=T}$ being the kth matched filter output sampled at time instant $t = T$; and $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{P})$ being the *colored* Gaussian noise due to the matched filtering. Note that matrix $\mathbf{P}$ in (3) is the correlation matrix of the signature waveforms. $\mathbf{P}$ is symmetric and positive definite and its elements are given by,

$$P[i, j] = \int_0^T s_i(t)\, s_j(t)\, dt \triangleq \rho_{ij}\,. \qquad \begin{matrix} i = 1, 2, \ldots, K; \\ j = 1, 2, \ldots, K. \end{matrix}$$

In practice, due to the finite bandwidth constraint and large number of users, the signature waveforms are not idealy orthonormal. Therefore the matrix $\mathbf{P}$ will not be an identity matrix in general. The non-diagonal nature of the $\mathbf{P}$ matrix will cause the interference between multiusers. In order to remove the multiple access interference (MAI), various detectors have been proposed [1 – 6]. One major effort of proposing various detectors is trying to balance the computational simplicity and reliable detection performance.

3. Linear Class of Multiuser Detectors and Their Performances

In this work, a comparative performance study of various *linear* multiuser detectors is conducted, with emphasis on LMMSE detector and the adaptive bootstrap multiuser detector.

3.1. Simple Decorrelating Detector

The simple decorrelating detector [4] is originated by finding a *linear* conditional maximum likelihood estimate (MLE) of the signal vector $\theta = \mathbf{A} \cdot \mathbf{b}$ from the *conditional* PDF $p(\mathbf{x}|\theta)$ obtained from (3). It then detects the multiuser information bit $\mathbf{b}$ by directly making decision on the *linear* conditional MLE $\hat{\theta} = \mathbf{P}^{-1} \cdot \mathbf{x}$,

$$\hat{\mathbf{b}} = \text{sign}\{\mathbf{z}\} = \text{sign}\{\mathbf{P}^{-1} \cdot \mathbf{x}\}\,. \tag{4}$$

This detector has the advantage of structural simplicity. It is also near-far resistant. But a potential problem associated with this detector is that noise is enhanced by the $\mathbf{P}^{-1}$ inverse filtering. We have shown in [6] that this detector has its limited performance with an error probability of the kth user,

$$P_e(k) = Q\left(\sqrt{\frac{a_k\,(1 - \underline{\rho}_k^T\, \mathbf{P}_k^{-1}\, \underline{\rho}_k)}{\sigma^2}}\right) \le Q\left(\sqrt{\frac{a_k}{\sigma^2}}\right), \tag{5}$$

Note that in (5), the matrix $\mathbf{P}_k$ is a $(K-1)\times(K-1)$ matrix, constructed from $\mathbf{P}$ matrix after removing the contribution of the kth user. $\underline{\rho}_k$ is the kth column of matrix $\mathbf{P}_k$. The inequality $0 < (1 - \underline{\rho}_k^T\, \mathbf{P}_k^{-1}\, \underline{\rho}_k) \le 1$ always holds, with the equality holds if and only if $\mathbf{P}$ is a diagonal matrix. This corresponds to the case of using a sets of perfect orthonormal signature waveforms. Hence, the performance of the decorrelating detector in (4) is always worse than the BPSK limit, as shown in the inequality in (5). The near-far resistance of the decorrelating detector can be easily seen from its $P_e(k)$ expression in (5), since $P_e(k)$ is invariant to $\mathbf{A}_k \triangleq \text{diag}\{\sqrt{a_1}, \cdots, \sqrt{a_{k-1}}, \sqrt{a_{k+1}}, \cdots, \sqrt{a_K}\}$.

We further notice the following fact that the above decorrelating detector is also a linear estimate based detector. Part of its limited performance is due to the fact that linear estimate $\hat{\theta} = \mathbf{P}^{-1}\mathbf{x}$ is based on the *conditional* PDF $p(\mathbf{x}|\theta)$. Therefore, we can expect to further improve its detection performance and maintain its linear feature, by incorporating the joint statistics of both θ and $\mathbf{x}$ into the estimate as demonstrated in the following linear minimum mean squared error (LMMSE) estimate-based detector.

3.2. Linear Minimum Mean Squared Error (LMMSE) Detector

It is well known that among the linear class of estimates, LMMSE estimate exhibits the minimum mean squared estimation error. Therefore we can improve the detection performance of the above *linear* decorrelating detector by deriving a LMMSE estimate-based multiuser detector. Specifically, let us rewrite formula (3) as,

$$\mathbf{x} = \mathbf{P}\,\mathbf{A}\,\mathbf{b} + \mathbf{n} = \mathbf{P}\,\theta + \mathbf{n}\,. \tag{6}$$

We assume that all the components of the random multiuser information bit $\mathbf{b}$ are independent and identically distributed (i.i.d.) with zero mean and unit variance. Even further, the random vector θ and noise vector $\mathbf{n}$ are statistically independent. For most communication applications, these assumptions are reasonable ones. We then define a new random vector as follows,

$$\mathbf{y} \triangleq \begin{bmatrix} \theta \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\,\mathbf{b} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{P}\,\mathbf{A} & \mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{b} \\ \mathbf{n} \end{bmatrix}.$$

The expectation and covariance matrix of the above newly defined vector $\mathbf{y}$ can be found as,

$$E(\mathbf{y}) = \mathbf{0}\,,$$

$$\text{cov}(\mathbf{y}) = \begin{bmatrix} \mathbf{A}^2 & \mathbf{A}^2\mathbf{P} \\ \mathbf{P}\,\mathbf{A}^2 & \mathbf{P}\,\mathbf{A}^2\mathbf{P} + \sigma^2\mathbf{P} \end{bmatrix} \triangleq \begin{bmatrix} \sum_{\theta\theta} & \sum_{\theta\mathbf{x}} \\ \sum_{\mathbf{x}\theta} & \sum_{\mathbf{x}\mathbf{x}} \end{bmatrix}.$$

Given $\mathbf{x}$, $E(\mathbf{y})$, and $\mathrm{cov}(\mathbf{y})$, the LMMSE estimate of θ can then be derived from the following [7, 8],

$$\begin{aligned}\hat{\theta}_{LMMSE} &= E(\theta) + \sum_{\theta\mathbf{x}} \sum_{\mathbf{xx}}^{-1} (\mathbf{x} - E(\mathbf{x})) \\ &= \underbrace{\left(\mathbf{P} + \sigma^2 \mathbf{A}^{-2}\right)^{-1}}_{\mathbf{W}} \cdot \mathbf{x} . \end{aligned} \tag{7}$$

The LMMSE estimate-based multiuser detector makes its decision based on the decision rule,

$$\hat{\mathbf{b}} = \mathrm{sign}\left(\hat{\theta}_{LMMSE}\right) . \tag{8}$$

Note that the diagonal matrix $\sigma^2 \mathbf{A}^{-2} = \mathrm{diag}\{\sigma^2/a_1, \sigma^2/a_2, \cdots, \sigma^2/a_K\}$ involved in $\mathbf{W}$ of formula (7) is actually the inverse SNR matrix. When the interferences of other users are very small compared to the noise level, $\mathbf{W}$ reduces into a *diagonal* matrix, or $\mathbf{W} = (\mathbf{P} + \sigma^2 \mathbf{A}^{-2})^{-1} \approx (\mathbf{I} + \sigma^{-2} \mathbf{A}^{-2})^{-1}$. In this case, LMMSE detector has the same performance as that of the single user detector, which is the BPSK limit. When the interference levels are very large compared to the noise level, then $\mathbf{W} = (\mathbf{P} + \sigma^2 \mathbf{A}^{-2})^{-1} \approx \mathbf{P}^{-1}$, LMMSE detector's performance is compariable to that of the decorrelating detector. Therefore, the overall performance of LMMSE detector is better than that of the decorrelating detector. Even further, we calculate the probability of error of the kth user based on the following observation. We arrange the order of all the users such that $\theta^T = [\ \theta_k \quad \underline{\theta}_k^T\ ]$. We then decompose the above derived $\hat{\theta}_{LMMSE}$ as follows,

$$\begin{aligned}\hat{\theta}_{LMMSE} &= \mathbf{W}\mathbf{x} = (\mathbf{P} + \sigma^2 \mathbf{A}^{-2})^{-1} \mathbf{x} , \\ &= (\mathbf{P} + \sigma^2 \mathbf{A}^{-2})^{-1} \cdot (\mathbf{P}\mathbf{A}\mathbf{b} + \mathbf{n}) , \\ &= \theta - \sigma^2 \mathbf{W}\mathbf{A}^{-1}\mathbf{b} + \mathbf{W}\mathbf{n} , \\ &= \theta + \mathbf{e} , \end{aligned} \tag{9}$$

where the estimation error, $\mathbf{e} = -\sigma^2 \mathbf{W}\mathbf{A}^{-1}\mathbf{b} + \mathbf{W}\mathbf{n}$, contains both *bias* and noise components. We will notice later, the improved performance of the LMMSE detector is achieved by trading in a little bias for less noise variance, which finally results in less overall mean squared error (MSE).

For a given information bit $\mathbf{b}$, LMMSE detector of (8) makes an errorous decision on the kth user's information bit b_k whenever,

$$e_k < -\sqrt{a_k} , \quad \text{when } b_k = +1 ,$$

OR

$$e_k > +\sqrt{a_k} , \quad \text{when } b_k = -1 .$$

Therefore, the probability of error for the kth user can be expressed as,

$$\begin{aligned} P_e(k) &= P_e\left(k \,|\, \underline{\mathbf{b}}_k\right) \cdot P\left(\underline{\mathbf{b}}_k\right) , \\ &= \frac{1}{2}\left\{ P(e_k > +\sqrt{a_k} \,|\, b_k = -1, \underline{\mathbf{b}}_k) + \right. \\ &\qquad \left. P(e_k < -\sqrt{a_k} \,|\, b_k = +1, \underline{\mathbf{b}}_k) \ \right\} \cdot P\left(\underline{\mathbf{b}}_k\right) \\ &= \frac{1}{2^K} \sum_{\underline{\mathbf{b}}_k} \left[Q\left(\frac{\sqrt{a_k} - R_k(b_k = -1, \underline{\mathbf{b}}_k)}{\sigma_k} \right) + \right. \\ &\qquad \left. Q\left(\frac{\sqrt{a_k} + R_k(b_k = +1, \underline{\mathbf{b}}_k)}{\sigma_k} \right) \right] , \end{aligned} \tag{10}$$

where $R_k(\cdot) = -\sigma^2 \mathbf{w}_k^T \mathbf{A}^{-1} \mathbf{b}$; $\sigma_k^2 = \sigma^2 \mathbf{w}_k^T \mathbf{P}\, \mathbf{w}_k$; and $\mathbf{w}_k^T$ is the kth row of $\mathbf{W}$ matrix.

We also verified the following facts that under various interference conditions, the above error probability expression will reduce either into the single user BPSK limit or into that of the decorrelating detector as follows,

$$P_e(k) = \begin{cases} Q\left(\dfrac{\sqrt{a_k}}{\sigma}\right) , & \text{small interferences} \\ Q\left(\dfrac{\sqrt{a_k\,(1 - \underline{\rho}_k^T \mathbf{P}_k^{-1} \underline{\rho}_k)}}{\sigma}\right) , & \text{strong interferences} \end{cases}$$

For the case of two users ($K = 2$), the error probability expression in (10) can be simplified as,

$$\begin{aligned} P_e(1) = \ & \frac{1}{2}\left\{ Q\left(\frac{\sqrt{a_1} - R_1(-1,+1)}{\sigma_1} \right) + \right. \\ & \left. Q\left(\frac{\sqrt{a_1} - R_1(-1,-1)}{\sigma_1} \right) \right\} , \end{aligned} \tag{11}$$

with

$$R_1(-1,+1) = \sigma^2 \frac{(1 + \sigma^2/a_2)/\sqrt{a_1} + \rho/\sqrt{a_2}}{(1 + \sigma^2/a_1)(1 + \sigma^2/a_2) - \rho^2} ,$$

$$R_1(-1,-1) = \sigma^2 \frac{(1 + \sigma^2/a_2)/\sqrt{a_1} - \rho/\sqrt{a_2}}{(1 + \sigma^2/a_1)(1 + \sigma^2/a_2) - \rho^2} ,$$

$$\sigma_1 = \sigma \frac{\sqrt{\left((1 + \sigma^2/a_2) - \rho^2\right)^2 + \rho^2 (1 - \rho^2)}}{(1 + \sigma^2/a_1)(1 + \sigma^2/a_2) - \rho^2} .$$

We can also easily verify the following limit results,

$$\lim_{\frac{\sigma^2}{a_2} \ll 1} P_e(1) = Q\left(\frac{\sqrt{a_1}}{\sigma}\right) , \qquad \text{BPSK limit}$$

$$\lim_{\frac{\sigma^2}{a_2} \gg 1} P_e(1) = Q\left(\frac{\sqrt{a_1\,(1 - \rho^2)}}{\sigma}\right) , \quad \text{decorrelating detector}$$

4. Examples and Implementation Issues

Numerical examples and computer simulation results further confirm our above derivations. Note that implementation of the LMMSE detector needs knowledge of matrix P and SNR matrix, but the adaptive bootstrap multiuser detector developed by BarNess et. al. [2] can achieve the same performance as that of the LMMSE detector without these knowledge. In Figure 1(a)(b)(c), we show the performance comparison of the decorrelating detector, the LMMSE detector, and the adaptive bootstrap multiuser detector. We also plot the BPSK limit as a reference lower bound on the performance. It can be seen that there is an equivalence between the LMMSE detector and the adaptive bootstrap multiuser detector. The improved performances of the LMMSE and the adaptive bootstrap multiuser detectors are obtained by trading a little bias for less noise variance, which finally results in an overall less mean squared error (MSE). And the adaptive bootstrap multiuser detector provides a practical implementation of the LMMSE detector for CDMA communication applications.

References

[1] U. Madhow, and M. L. Honig, "MMSE Interference Suppression for Direct-Sequence Spread-Spectrum CDMA," *IEEE Trans. on Communications*, vol. 42, no. 12, Dec. 1994.

[2] Y. Bar-Ness, and J. Punt, "Adaptive Bootstrap CDMA Multiuser Detector," to appear in special issue of *Signal Separation and Interference Cancellation for Personal, Indoor and Mobile Radio Communication*, Wireless Personal Communications, An International Journal, Kluwer Academic Publishers.

[3] S. Verdu, "Minimum Probability of Error for Asynchronous Gaussian Multiple-Access Channel," *IEEE Trans. on Information Theory*, vol. IT-32, no. 1, pp. 85--96, Jan. 1986.

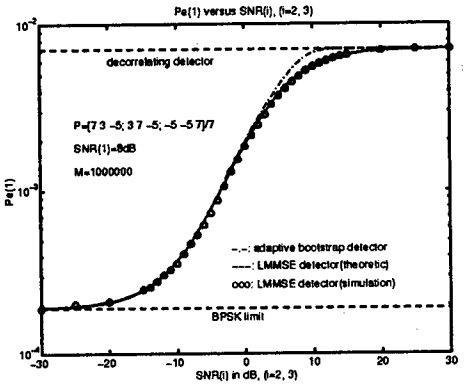
[4] R. Lupas and S. Verdu, "Linear Multiuser Detector for synchronous code division multiple access channels," *IEEE Trans. on Information Theory*, vol. IT-35, no. 1, pp. 123-136, Jan. 1989.

[5] M. K. Varanasi and B. Aazhang, "Near-Optimum Detection in Synchronous Code-Division Multiple-Access Systems," *IEEE Trans. on Commun.*, vol. 39, no. 5, pp. 725--736, May. 1991.

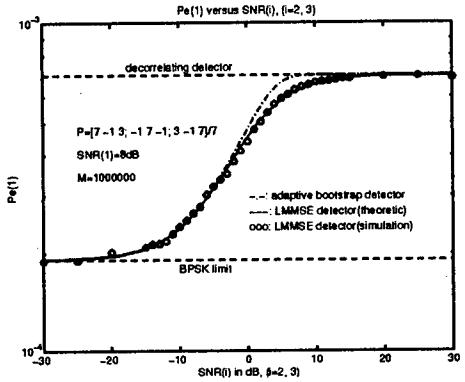
[6] H. Ge, and Y. Bar-Ness, "Bayesian Approach of Multiuser Separation and Interference Suppression in CDMA Communication Systems," *Proc. of 29th Asilomar Conf. on Signals, Systems and Computers*, Oct. 1995.

[7] P. J. Bickel, and K. J. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Prentice Hall, Englewood Cliffs, New Jersey, 1977.

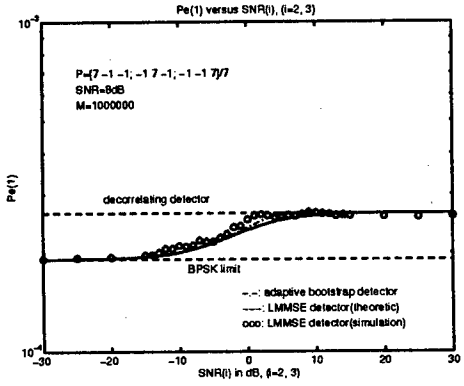
[8] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley Publishing Company, 1990.



(a). High capacity case



(b). Medium capacity case



(c). Low capacity case.

Figure 1: Performance of the LMMSE detector and the adaptive bootstrap detector in multi-user CDMA system. Also shown in figure are the results of the decorrelating detector and the BPSK limit. Parameters used: $SNR_1 = 8\,dB$, $K = 3$, $M = 1000000$ independent trials.

Reduced Complexity Blind 2D RAKE Receiver for CDMA †

Javier Ramos and Michael D. Zoltowski
School of Electrical Engineering
Purdue University
West Lafayette, IN 47907-1285
e-mail: mikedz@ecn.purdue.edu , ramos@ecn.purdue.edu

ABSTRACT

We previously presented [1] a blind 2D RAKE receiver for CDMA that cancels strong multi-user access interference (MAI) and optimally combines multipath. After passing the output of each antenna through a matched filter based on the spreading waveform of the desired user, one estimates the signal plus interference spatio-frequency correlation matrix during that portion of the bit interval where the fingers of the RAKE occur, and the interference alone spatio-frequency correlation matrix during that portion of the bit interval away from the fingers. A reduced complexity scheme that outperforms the previous algorithm in a MAI dominant environment is presented based on a data adaptive transformation to a beamspace of dimension equal to the effective number of spatial degrees of freedom taken up by the desired user's multipath.

1. INTRODUCTION

In [1] we presented a blind space-time processing scheme for a Direct Sequence Spread Spectrum based CDMA PCS/cellular communications system that cancels co-channel interference while simultaneously combining multipath in an optimal "RAKE-like" fashion. After passing the output of each antenna through a matched filter based on the spreading waveform of the desired user, one estimates the signal plus interference spatio-temporal correlation matrix during that portion of the bit interval where the fingers of the RAKE occur, and the interference alone spatio-temporal correlation matrix during that portion of the bit interval away from the fingers. It was shown that the weight vector yielding the optimum signal to interference plus noise ratio for bit decisions is the "largest" generalized eigenvector of the resulting matrix pencil.

Taking a cue from either the IS-95 standard or the coarse acquisition code embedded in the GPS signal, suppose the chip duration is 1 microsecond. Experimental measurements in an urban cellular environment reveal that the worst case time delay spread, τ_{max}, due to multipath is on the order of 10 microseconds [2]. Sampling two times per chip, the resulting space-time correlation matrix would be of dimension $20N \times 20N$, where N is the number of antennas: 120×120 in the case of $N = 6$ antennas. The large dimensionality of the spatio-temporal correlation matrix pencil prompted an investigation into a frequency domain implementation of a RAKE receiver in [1].

The primary advantage of a frequency domain implementation of a RAKE receiver is that we can select only frequency values within the mainlobe of the spectrum of the autocorrelation function of the spreading waveform and the number of such values required for "good" performance can be substantially less than the number of time samples recorded during the multipath time delay spread. The result is that the spatio-frequency correlation matrix is of significantly smaller dimension than the space-time correlation matrix with no degradation in performance. Another possible advantage is that the frequency domain implementation may allow for a lower sampling rate than space-time processing since we can allow aliasing in the sidelobes of the autocorrelation function of the spreading waveform as no frequency samples from there are used.

Extensive simulations have revealed that taking roughly 10 frequency samples equi-spaced between minus half the chip rate to plus half the chip rate provides "good" performance. Thus, the spatio-frequency correlation matrices are roughly of dimension $10N \times 10N$ regardless of the value of τ_{max}: 60×60 in the case of $N = 6$ antennas. The motivation for this paper is twofold. First, the dimension of the spatio-frequency snapshot is quite large for a reasonable number of antennas. Second, a spatial null is required to cancel each strong MAI since each MAI is a broadband interferer. Thus, the number of MAI's that can be canceled is limited by the number of antennas, N, and not the dimension of the spatio-frequency snapshot vectors, $10N$. This paper theoretically analyzes a reduced complexity scheme, originally proposed in [3], based on a data adaptive transformation to a beamspace of dimension equal to the effective number of spatial degrees of freedom taken up by the desired user's multipath. Low dimension spatio-frequency correlation matrices are formed in the reduced dimension beamspace. We begin the development with the space-time data model.

2. SPACE-TIME DATA MODEL

The $N \times 1$ array snapshot vector $\mathbf{x}(n)$ containing the outputs of each of the N antennas comprising the array at discrete time n is modeled as

$$\mathbf{x}(t) = \sum_{k=1}^{K} \rho_k \sum_{n=0}^{N_b-1} \mathbf{a}(\theta_k^d) D(n) c(t - nT_b - \tau_k) + \sum_{i=1}^{J} \sum_{n=0}^{N_b-1} \mathbf{a}(\theta_i) \sigma_i D_i(n) c_i(t - nT_b) + \mathbf{n}_w(t) \quad (1)$$

where $\mathbf{a}(\theta)$ is the spatial response of the array. For the sake of notational simplicity, we here assume that the spatial response vector depends on a single directional parameter, θ, the direction of arrival (DOA) of a given source. However, no model is assumed for $\mathbf{a}(\theta)$ in the algorithm to be presented; the algorithm works for any array geometry. $1/T_b$ is the symbol rate. K is the number of different paths the Signal of Interest (SOI) arrives from, θ_k^d denotes the directions

†This research was supported by the Air Force Office of Scientific Research under grant no. F49620-95-1-0367, the National Science Foundation under under grant no. MIPS-9320890, and the Army Research Office's Focused Research Initiative under grant number DAAHO4-95-1-0246.

associated with the k-th path, and τ_k is the corresponding relative delay of the k-th path. ρ_k is the complex amplitude of the k-th multipath arrival for the SOI at the reference element. $D(n)$ and $D_i(n)$ are the digital information sequences for the SOI and MAI sources, respectively. J broadband interferers (MAI) impinge upon the array. σ_i is the complex amplitude of the i-th interferer at the reference element of the array. $c(t)$ and $c_i(t)$ are the spreading waveforms for the SOI and i-th MAI, respectively. The vector $\mathbf{n}_w(t)$ contains white noise. N_b is the number of bits over which all parameters characterizing the model in (1) are assumed to be constant. N_b might be quite small in cases of rapidly evolving dynamics.

The spreading waveform for the i-th MAI is modeled as

$$c_i(t) = \sum_{m=0}^{N_c-1} d_i(m) p_c(t - mT_c) \tag{2}$$

where $1/T_c$ is the chip rate, $d_i(n)$ is a pseudo-noise (PN) sequence [1], $p_c(t)$ is the chip waveform assumed common to all sources, N_c is the number of chips per bit common to all MAI's. The spreading waveform for the desired source, $c(t)$, is defined similarly but with a different PN sequence.

The received signal at each antenna is sampled at a rate $f_s = L_c/T_c$, where L_c is the number of samples per chip. The sampled output of each antenna is passed through a filter with impulse response $h[n] = c[-n]$, where $c[n] = c(nT_c/L_c)$. Let $\mathbf{X}_F(n)$ denote the $N \times N_c L_c$ matrix whose j-th row contains $N_s = N_c L_c$ samples of the output of the j-th antenna after the matched filter for the n-th (i.e., one) bit period. Given the model for $\mathbf{x}(t)$ in (1), $\mathbf{X}_F$ may be expressed as

$$\mathbf{X}_F(n) = D(n)\mathbf{A}\mathbf{T} + \mathbf{A}_I \mathbf{\Sigma}_I \mathbf{B}(n)\mathbf{P} + \mathbf{N}(n) \tag{3}$$

$\mathbf{A}$ is an $N \times K$ matrix whose K columns are are $\mathbf{a}(\theta_k^d)$, $k = 1, \cdots, K$. $\mathbf{T}$ is the $K \times N_s$ (recall $N_s = N_c L_c$) matrix given by

$$\mathbf{T} = \begin{pmatrix} r_{cc}(-\tau_1) & r_{cc}\left(\frac{T_c}{L_c}-\tau_1\right) & \cdots & r_{cc}\left((N_s\text{-}1)\frac{T_c}{L_c}-\tau_1\right) \\ r_{cc}(-\tau_2) & r_{cc}\left(\frac{T_c}{L_c}-\tau_2\right) & \cdots & r_{cc}\left((N_s\text{-}1)\frac{T_c}{L_c}-\tau_2\right) \\ \vdots & \vdots & \vdots & \vdots \\ r_{cc}(-\tau_K) & r_{cc}\left(\frac{T_c}{L_c}-\tau_K\right) & \cdots & r_{cc}\left((N_s\text{-}1)\frac{T_c}{L_c}-\tau_K\right) \end{pmatrix} \tag{4}$$

where $r_{cc}(\tau)$ is the autocorrelation function for the SOI's spreading waveform, $c(t)$. In the case where the chip waveform, $p_c(t)$, is rectangular and the processing gain is large, $r_{cc}(\tau)$ has the following triangular shape:

$$r_{cc}(\tau) = \begin{cases} 1 - \frac{|\tau|}{T_c} & \text{if} \quad |t| \le T_c \\ 0 & \text{if} \quad |t| > T_c \end{cases} \tag{5}$$

$\mathbf{\Sigma}_I$ is a $J \times J$ diagonal matrix containing the amplitudes (at the reference element), σ_i, $i = 1, ..., J$, for each of the MAI's. $\mathbf{B}(n)$ is a $J \times J$ diagonal matrix containing the bit values for the J MAI's. The columns of the $N \times J$ matrix $\mathbf{A}_I$ are $\mathbf{a}(\theta_i)$, $i = 1, \cdots, J$. $\mathbf{P}$ is the $J \times N_s$ matrix

$$\mathbf{P} = \begin{pmatrix} c_1[n] * h[n] \\ \vdots \\ c_J[n] * h[n] \end{pmatrix} \tag{6}$$

where $c_i[n] = c_i(nT_c/L_c)$. where $*$ is the linear convolution operator truncated at N_s samples. $\mathbf{N}(n)$ is the noise contribution; the rows of $\mathbf{N}(n)$ are independent Gaussian process but the individual components of each row are correlated because of the matched filtering operation.

[1] Without loss of generality, real-valued spreading waveforms have been assumed for notational simplicity.

3. SPACE-FREQUENCY DATA MODEL

Define the $N_s \times N_w$ selection matrix $\mathbf{\Gamma}_i$ as

$$\mathbf{\Gamma}^i = \begin{pmatrix} \mathbf{0}_{(i-1)\times N_w} \\ \mathbf{I}_{N_w \times N_w} \\ \mathbf{0}_{(N_s - i - N_w + 1)\times N_w} \end{pmatrix} \quad i = 1, ..., N_s - N_w \tag{7}$$

where

$$N_w = \frac{L_c \tau_{max}}{T_c} \tag{8}$$

where τ_{max} is the worst case time delay spread due to multipath. Without loss of generality, we have chosen τ_{max} such that

$$N_m = \frac{\tau_{max}}{T_c} \tag{9}$$

is an integer. The important thing is that τ_{max} be an upper bound on the experimentally measured worst case multipath time delay. If $p_c(t)$ is one microsecond in duration, a reasonable number for N_m is 10 in an urban cellular environment.

Define the i-th spatio-frequency snapshot for the n-th bit as

$$\mathbf{y}^{(i)}(n) = \text{vec}\left(\mathbf{X}_F(n)\mathbf{\Gamma}^{(i)}\mathbf{W}\right) \tag{10}$$

The $N_w \times L$ matrix $\mathbf{W}$ is composed of $L \le N_w$ columns of the N_w point DFT matrix; the ℓ-th column of $\mathbf{W}$ has the form

$$\mathbf{w}_\ell = \left[1, e^{-j2\pi\frac{\ell}{N_w}} \cdots e^{-j2\pi\frac{\ell(N_w-1)}{N_w}}\right]^T. \tag{11}$$

Note that the analog frequency separation between adjacent spectral lines is

$$\Delta f = \frac{1}{\tau_{max}} = \frac{1}{N_m}\frac{1}{T_c} \tag{12}$$

Finally, vec$(\cdot)$ is the operator that maps an $N \times L$ matrix to an $NL \times 1$ vector by concatenating its columns. Summarizing the steps implied by (10), the procedure is to compute L spectral lines of the rows of $\mathbf{X}_F(n)$ over $N_w = N_m L_c$ samples starting at the i-th column, and then stack the resulting L $N \times 1$ vectors in an $NL \times 1$ vector.

Let's examine the structure of the $NL \times 1$ spatio-frequency snapshot vector formed by substituting the form of $\mathbf{X}_F(n)$ in (3) into (10). Assuming the processing gain, i.e., the number of chips per bit, N_c, to be large, and that we have approximate bit synchronization for the desired user, the term corresponding to the SOI is only nonzero during the first N_m chips associated with index $i = 1$. Assume $r_{cc}(\tau)$ to have the triangular shaped described by (5) corresponding to a rectangular chip waveform. In this case, each row of $\mathbf{T}$ is simply a sampled version of a time-delayed replica of (5), delayed by τ_i where i is the row index. Using simple properties of the Fourier Transform it follows that

$$\mathbf{y}_s(n) \triangleq \text{vec}\left(D(n)\mathbf{A}\mathbf{T}\mathbf{\Gamma}^{(i)}\mathbf{W}\right) = D(n)\text{vec}(\mathbf{A}\mathbf{\Phi}\mathbf{F}) \tag{13}$$

where $\mathbf{F}$ is an $L \times L$ diagonal matrix: the ℓ-th diagonal entry of $\mathbf{F}$ is of the form

$$T(\ell) = \left(\frac{\sin\left(\frac{\pi\ell}{N_m}\right)}{L_c \sin\left(\frac{\pi\ell}{N_w}\right)}\right)^2 \tag{14}$$

The ℓ-th column of the $J \times L$ matrix $\mathbf{\Phi}$ is of the form

$$\boldsymbol{\phi}_\ell = \left[e^{-j2\pi\ell\frac{\tau_1}{\tau_{max}}}, e^{-j2\pi\ell\frac{\tau_2}{\tau_{max}}}, \cdots, e^{-j2\pi\ell\frac{\tau_J}{\tau_{max}}}\right]^T \tag{15}$$

It is instructive to examine the asymptotic structure of the $LN \times LN$ spatio-frequency correlation matrix of $\mathbf{y}_s(n)$:

$\mathbf{K}_S = E[\mathbf{y}_s(n)\mathbf{y}_s^H(n)]$. Since the only time varying quantity is $D(n)$, it follows that

$$\mathbf{K}_S = \sigma_b^2 \mathbf{d}\mathbf{d}^H, \quad \text{where: } \mathbf{d} = (\mathbf{F} \otimes \mathbf{A}\boldsymbol{\Sigma}_S)\text{vec}(\boldsymbol{\Phi}) \tag{16}$$

where $\sigma_b^2 = E[D^2(n)]$, $\boldsymbol{\Sigma}_S$ is a $K \times K$ diagonal matrix containing the complex amplitudes of the K multipaths for the SOI. The result in (16) was obtained by invoking the property $\text{vec}(\mathbf{ADB}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{D})$. Note that $\mathbf{K}_S$ is a rank one matrix.

The energy contribution of the MAI's to a given bit period is approximately evenly spread across the entire bit period. Define

$$\mathbf{y}_I^{(i)}(n) = \text{vec}\left(\mathbf{A}_I\boldsymbol{\Sigma}_I\mathbf{B}(n)\mathbf{P}\boldsymbol{\Gamma}^{(i)}\mathbf{W}\right) \tag{17}$$

It can be shown that the MAI spatio-frequency correlation matrix $\mathbf{K}_I = E[\mathbf{y}_I^{(i)}(n)\mathbf{y}_I^{(i)H}(n)]$ may be expressed as

$$\begin{aligned}\mathbf{K}_I &= E[\text{vec}(\mathbf{A}_I\boldsymbol{\Sigma}_I\mathbf{B}(n)\mathbf{P}\boldsymbol{\Gamma}^{(i)}\mathbf{W})\text{vec}^H(\mathbf{A}_I\boldsymbol{\Sigma}_I\mathbf{B}(n)\mathbf{P}\boldsymbol{\Gamma}^{(i)}\mathbf{W})] \\ &= \mathbf{R}_F \otimes \mathbf{A}_I\boldsymbol{\Sigma}_I^2\mathbf{A}_I^H \end{aligned} \tag{18}$$

where we have exploited the following four properties: (i) $\text{vec}(\mathbf{ADB}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{D})$, (ii) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$, and (iii) $E\{\mathbf{B}(n)\mathbf{B}^H(n)\} = \sigma_b^2\mathbf{I}_{J\times J}$ (the data from different sources are assumed to be uncorrelated,) and (iv) $E[d_i(m)d_j(\ell)] = \delta_{ij}\delta_{m\ell}$, the chip values comprising each PN sequence are modeled as independent and identically distributed. $\mathbf{R}_F = \mathbf{W}^T\boldsymbol{\Upsilon}_I\mathbf{W}^*$, where $\boldsymbol{\Upsilon}_I$ is a Toeplitz-symmetric matrix whose first column is $r_{cc}^2(mT_c/L_c)$, $m = 0, ..., N_w - 1$. Note that $\mathbf{R}_F$ is full rank and $\mathbf{A}_I$ is rank J. It follows that $\mathbf{K}_I$ is rank LJ.

Regarding the noise contribution, a similar development reveals that the spatio-frequency correlation matrix for $\mathbf{y}_N^{(i)}(n) = \text{vec}\left(\mathbf{N}(n)\boldsymbol{\Gamma}^{(i)}\mathbf{W}\right)$ has the asymptotic form

$$\mathbf{K}_N = \boldsymbol{\Psi} \otimes \mathbf{I}_N \tag{19}$$

where $\boldsymbol{\Psi}$ is is an $L \times L$ matrix. $\boldsymbol{\Psi}$ may be expressed as $\boldsymbol{\Psi} = \mathbf{W}^T\boldsymbol{\Upsilon}_N\mathbf{W}^*$, where $\boldsymbol{\Upsilon}_N$ is a Toeplitz-symmetric matrix whose first column is $r_{cc}(mT_c/L_c)$, $m = 0, ..., N_w - 1$.

4. BLIND ADAPTIVE 2D RAKE RECEIVER

The signal plus interference plus noise spatio-frequency correlation matrix is estimated as

$$\hat{\mathbf{K}}_{S+I+N} = \frac{1}{N_b}\sum_{n=0}^{N_b-1}\mathbf{y}^{(1)}(n)\mathbf{y}^{(1)H}(n) \tag{20}$$

Observe that only spatio-frequency snapshot is extracted from each bit in forming $\hat{\mathbf{K}}_{S+I+N}$. The interference plus noise spatio-frequency correlation matrix is estimated as

$$\hat{\mathbf{K}}_{I+N} = \frac{1}{N_b(N_s\text{-}1)}\sum_{n=0}^{N_b-1}\sum_{i=2}^{N_s-N_w}\mathbf{y}^{(i)}(n)\mathbf{y}^{(i)H}(n) \tag{21}$$

In the development below we show that the optimum set of spatio-frequency weights for weighting and summing the frequency samples computed in the vicinity of the fingers of the RAKE is the "largest" generalized eigenvector of the matrix pencil $\{\hat{\mathbf{K}}_{S+I+N}, \hat{\mathbf{K}}_{I+N}\}$.

We here show that the asymptotic spatio-frequency weight vector $\mathbf{w}$ obtained as the "largest" generalized eigenvector of the matrix pencil $\{\mathbf{K}_{S+I+N}, \mathbf{K}_{I+N}\}$ corresponds to the solution to the Minimum Variance Distortionless Response (MVDR) problem

$$\min_{\mathbf{w}} \ \mathbf{w}^H\mathbf{K}_{I+N}\mathbf{w}$$

$$\text{subject to: } \mathbf{w}^H\mathbf{d} = 1 \tag{22}$$

where $\mathbf{d}$ is the $NL \times 1$ vector defined in (16).

The "largest" generalized eigenvector of the pencil $\{\mathbf{K}_{S+I+N}, \mathbf{K}_{I+N}\}$ is the solution to the unconstrained optimization problem

$$\max_{\mathbf{w}}\frac{\mathbf{w}^H\mathbf{K}_{S+I+N}\mathbf{w}}{\mathbf{w}^H\mathbf{K}_{I+N}\mathbf{w}} = \max_{\mathbf{w}}\frac{\mathbf{w}^H(\mathbf{K}_{I+N} + \sigma_s^2\mathbf{d}\mathbf{d}^H)\mathbf{w}}{\mathbf{w}^H\mathbf{K}_{I+N}\mathbf{w}} \tag{23}$$

where we have invoked the fact that $\mathbf{K}_S$ is rank one. It follows that

$$\max_{\mathbf{w}}\left\{1 + \frac{\mathbf{w}^H\sigma_s^2\mathbf{d}\mathbf{d}^H\mathbf{w}}{\mathbf{w}^H\mathbf{K}_{I+N}\mathbf{w}}\right\} = \min_{\mathbf{w}}\frac{\mathbf{w}^H\mathbf{K}_{I+N}\mathbf{w}}{\mathbf{w}^H\sigma_s^2\mathbf{d}\mathbf{d}^H\mathbf{w}} \tag{24}$$

which is equivalent to the constrained minimization problem in (22). The "inversion" of the maximization problem resulting in the equivalent minimization problem is possible since $\mathbf{K}_{I+N}$ is a positive-definite matrix.

We now prove each sub-vector of the optimum weight vector corresponding to a particular frequency bin is orthogonal to each column of $\mathbf{A}_I$. Equation (18) reveals that the spatio-frequency correlation matrix of the MAI is a Kronecker product of the $L \times L$ matrix $\mathbf{R}_F$ with the $N \times N$ matrix $\mathbf{A}_I\boldsymbol{\Sigma}_I^2\mathbf{A}_I^H$. Let $\mathbf{E}_F$ be an $L \times L$ matrix whose columns are the eigenvectors of $\mathbf{R}_F$. Since $\mathbf{R}_F$ is a full rank Hermitian matrix, it follows that $\mathbf{E}_F^H\mathbf{E}_F = \mathbf{E}_F\mathbf{E}_F^H = \mathbf{I}_L$. However, $\mathbf{A}_I\boldsymbol{\Sigma}_I^2\mathbf{A}_I^H$ is only rank J. Let Let $\mathbf{E}_S$ be an $N \times J$ matrix whose columns are the eigenvectors of $\mathbf{A}_I\boldsymbol{\Sigma}_I^2\mathbf{A}_I^H$ associated with the J nonzero eigenvalues. From signal subspace theory, it follows that $\mathbf{E}_S = \mathbf{A}_I\mathbf{T}$ where $\mathbf{T}$ is a $J \times J$ full rank matrix. It was noted previously that $\mathbf{K}_I$ is of rank LJ. Let $\mathbf{E}_I$ be an $LN \times LJ$ matrix whose columns are the eigenvectors of $\mathbf{K}_I$ associated with the LJ nonzero eigenvalues. It follows from the theory of Kronecker products, that $\mathbf{E}_I = \mathbf{E}_F \otimes \mathbf{E}_S$.

Post-multiplying $\mathbf{E}_I$ by a full rank $LJ \times LJ$ matrix yields a matrix whose range space is the same as that of $\mathbf{E}_I$. Post-multiplying $\mathbf{E}_I$ by the full rank matrix $\mathbf{E}_F^H \otimes \mathbf{T}^{-1}$ yields

$$\mathbf{G}_I = \mathbf{E}_I(\mathbf{E}_F^H \otimes \mathbf{T}^{-1}) = (\mathbf{E}_F \otimes \mathbf{E}_S)(\mathbf{E}_F^H \otimes \mathbf{T}^{-1}) \tag{25}$$

$$= \mathbf{E}_F\mathbf{E}_F^H \otimes \mathbf{E}_S\mathbf{T}^{-1} = \mathbf{I}_L \otimes \mathbf{A}_I \tag{26}$$

where we have used the Kronecker product property $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$.

The projection operator $\mathbf{P}_I$ onto the range space of $\mathbf{K}_I$ may be expressed as

$$\mathbf{P}_I = \mathbf{E}_I\mathbf{E}_I^H = \mathbf{G}_I(\mathbf{G}_I^H\mathbf{G}_I)^{-1}\mathbf{G}_I^H = \mathbf{I}_L \otimes \mathbf{P}_{A_I} \tag{27}$$

where $\mathbf{P}_{A_I}$ is the projection operator $(N \times N)$ into the subspace generated by the columns of $\mathbf{A}_I$.

The optimum weight vector is given by the Wiener solution as $\mathbf{w}^{opt} = \mathbf{K}_{I+N}^{-1}\mathbf{d}$. When the power of the MAI's is much stronger than than the receiver generated noise, the optimum weight vector is well approximated by $\mathbf{w}^{opt} \approx \mathbf{d} - \mathbf{P}_I\mathbf{d}$. Substituting the expression for $\mathbf{P}_I$ in (27) in $\mathbf{d} - \mathbf{P}_I\mathbf{d}$ yields the following asymptotic expression for the optimum weight vector:

$$\mathbf{w}^{opt} = \begin{pmatrix} \mathbf{d}_1 - \mathbf{P}_{AI}\mathbf{d}_1 \\ \vdots \\ \mathbf{d}_L - \mathbf{P}_{AI}\mathbf{d}_L \end{pmatrix} \tag{28}$$

where $\mathbf{d}_i$ is the i-th $N \times 1$ sub-vector of the $NL \times 1$ vector $\mathbf{d}$ defined in (16). It follows that the i-th $N \times 1$ sub-vector of $\mathbf{w}^{opt}$ is orthogonal to each column of $\mathbf{A}_I$, i.e., the array pattern obtain at *each* frequency bin exhibits a spatial null in the direction of *each* and every MAI.

5. BLIND SPATIAL PRE-PROCESSING

The spatio-frequency MVDR processor should accomplish two tasks: (i) it should null the MAI's and (ii) it should optimally combine the fingers corresponding to the different paths of the SOI. The optimum weight vector is the largest "generalized" eigenvector of an $NL \times NL$ spatio-frequency correlation matrix pencil. This large dimensionality translates into a corresponding large computational burden, detracting from the real-time applicability of the scheme and slowing up the time to convergence as well.

We here restrict our attention to a scenario where the MAI's are the primary source of interference. Since canceling each broadband MAI consumes L degrees of freedom – so that J spatial nulls are formed towards the MAI's at each of the L frequency bins – the interference rejection capability of the algorithm is not diminished if it is divided into two stages: spatial-only pre-processing with N degrees of freedom to cancel the MAI's followed by blind MVDR spatio-frequency in the reduced dimension beamspace.

In this scheme, we first transform to a $p < N$ dimensional beamspace using the p "largest" generalized eigenvectors of the $N \times N$ spatial correlation matrix pencil $\{\hat{\mathbf{R}}_{S+I+N}, \hat{\mathbf{R}}_{I+N}\}$. Here p is the number of dominant multipath for the SOI that are resolvable in space. $\hat{\mathbf{R}}_{S+I+N}$ is formed from snapshots measured in the vicinity of the "fingers," while $\hat{\mathbf{R}}_{I+N}$ is formed from snapshots measured away from the "fingers."

6. SIMULATION

A simulations was conducted employing a six element uniformly-spaced linear array with half-wavelength spacing. Both the desired source and the interferer were DS-CDMA signals with different Gold codes and 127 chips per bit; the duration of a chip was one microsecond. The modulation overlay was BPSK. A simple two-ray multipath model was used for the desired source wherein the direct path arrived at an elevation angle of 0° relative to broadside with an SNR of 10 dB per element. The second ray arrived at an angle of 10° with a relative delay of 2 chips and an SNR 6 dB below that of the direct path and phase shifted by 45° at the array center. The interferer was modeled as arriving at a single discrete angle, 30° elevation, with an SNR of 30 dB per element. There were two samples per chip.

The beam pattern obtained with the weight vectors of the first stage computed as the " two largest" generalized eigenvectors of the matrix pencil $\{\hat{\mathbf{R}}_{S+I+N}, \hat{\mathbf{R}}_{I+N}\}$ are plotted in Figure 1. Both patterns are observed to peak near the respective angular directions of the multipath arrivals for the desired user, and have a deep null in the direction of the interferer. The two "largest" generalized eigenvectors were employed to transform to a 2-dimensional beamspace. Applying the spatio-frequency processing scheme to the outputs of the $p = 2$ respective beams yields the signal constellation plotted in Figure 2.

Comparing the computational load of this two-stage procedure with spatio-frequency processing in the original element space, the latter requires the computation of the "largest" generalized eigenvector of a 60×60 matrix pencil. In contrast, the former requires computation of the two "largest" generalized eigenvectors of a 6×6 spatial matrix pencil followed by the computation of the "largest" generalized eigenvector of a 20×20 spatio-frequency correlation matrix pencil in beamspace. Moreover, relative to performance, the two-stage procedure offers faster convergence as there are many more spatial snapshots per bit for forming $\hat{\mathbf{R}}_{S+I+N}$ and $\hat{\mathbf{R}}_{I+N}$ than spatio-frequency snapshots per bit for forming $\hat{\mathbf{K}}_{S+I+N}$ and $\hat{\mathbf{K}}_{I+N}$. Coupled with the fact that $\hat{\mathbf{R}}_{S+I+N}$ and $\hat{\mathbf{R}}_{I+N}$, as well as $\hat{\mathcal{K}}_{S+I+N}$ and $\hat{\mathcal{K}}_{I+N}$, are of much smaller dimension than $\hat{\mathbf{K}}_{S+I+N}$ and $\hat{\mathbf{K}}_{I+N}$, there is a substantial decrease in the error of the estimate of the optimum space- frequency weight vector in beamspace relative to that in element space.

7. REFERENCES

1. M. D. Zoltowski and J. Ramos, "Blind Adaptive Beamforming for CDMA Based PCS/Cellular," *Conference Record of the 29th Asilomar IEEE Conference on Signals, Systems, and Computers,* 30 Nov. 1995.
2. T.S. Rappaport, S. Y. Seidel, and S. Yoshida, "900 MHz Multipath Propagation Measurements for U.S. Digital Cellular Radiotelephone," *IEEE Globecom '89,* pp. 84-89, 1989.
3. M. D. Zoltowski and J. Ramos, "Blind Multi-User Access Interference Cancellation for CDMA Based PCS/Cellular Using Antenna Arrays," *Proc. of 1996 IEEE ICASSP,* May 1996.

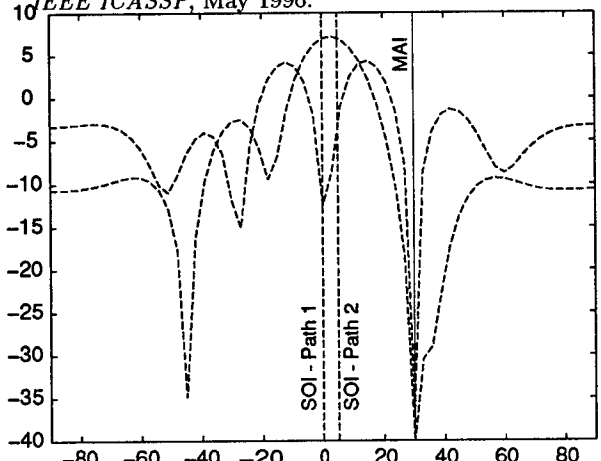


Figure 1. Array patterns obtained with space-only blind processing only

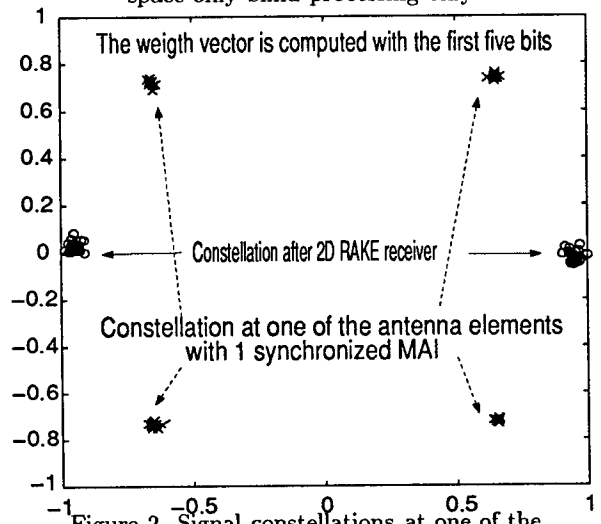


Figure 2. Signal constellations at one of the Antennas and after 2D RAKE Receiver

Null Placement Configurations for Radio Network Capacity Enhancement

Leandros Tassiulas
Dept. of Electrical Engineering and
Institute for Systems Research
University of Maryland
College Park, MD 20742
leandros@eng.umd.edu

Abstract

The use of array antennas in the radio nodes of a wireless network will increase the capacity since by appropriate null placement, the cochannel interference can be suppressed and the channel reuse will be enhanced. A radio network with arbitrary topology and array antennas in the nodes is considered in this paper. Each radio node may act either as a receiver or as a transmitter and can place a number of nulls in its radiation pattern. A collection of links may share the same channel if the interfering transmitters within the range of each receiver are cancelled out by null placement either at the receiver or at the transmitter side. The following problem is considered. Given a collection of links, identify a null placement configuration such that the links share the same channel without interference. It is shown that such a null placement configuration can be found by solving a maximum flow problem in an appropriately defined capacitated network.

1. Introduction

Array antennas with linear combining of the element outputs have been studied extensively over the past decades. By controlling the weights in the linear combination of the element outputs, the radiation pattern of the antenna can be designed to have certain desirable characteristics like a high gain narrow beam in the direction of the intended communicator and nulls in the directions of jammers/interceptors. Efficient signal processing algorithms for direction of arrival estimation, adaptive beam and null steering and multiple source-location estimation have been proposed [11, 10, 1, 2]. Most of the existing work on the subject is for the case of a single node employing an array antenna which communicates with another transceiver in the presence of multiple interferers.

It is apparent that the capability of an array antenna to shape its radiation pattern can be utilized in a wireless networking environment with multiple interacting radio links. The beamforming capabilities of an array antenna can be used both in the receiver and the transmitter end of a wireless link to improve the quality. The receiver antenna places the main lobe towards the direction of its transmitter in order to increase the link gain, while the reception nulls are placed towards the directions of interfering transmitters in order to reduce interference. Similarly the transmitter places its main lobe towards the direction of its receiver while the nulls should be placed such that the receivers of other links with high interference levels are not affected. In this way the signal to interference ratio of cochannel links can be enhanced and a large number of cochannel connections can be accommodated in each channel. Techniques that utilize the beamforming capabilities of array antennas to increase the capacity of wireless networks are currently under investigation by several researchers and they are termed *Spatial Division Multiple Access (SDMA)*. In this paper we consider the problem of increasing network capacity by using the null placement capability of array antennas, to facilitate the coexistence of large numbers of cochannel links.

2 Radio Network Model

Consider N radio nodes with one transceiver per node. The antenna of node v has e_v elements. Neighboring transmissions using the same channel interfere unless the transmitters and receivers place the nulls of their radiation patterns appropriately. We assume that the transmission of node v interferes the reception of node w from node k if w is in the neighborhood of v, and neither v nor w place a null in the direction from v to w. The proximity among the nodes and its impact to the signal and interference levels among them is captured by the connectivity graph $G = (X, E)$ where X is the set of radio nodes and edge (v, w) belongs to E if and only if nodes v, w are within range one from

0-8186-7576-4/96 $5.00

the other. Hence edge (v, w) implies both that the nodes can talk among themselves as well as that they interfere one another. A communication link between nodes v and w can be established only if (v, w) belongs to E. Furthermore the reception at some node v from some node w is interfered by the cochannel transmission of some other node u, if (u, v) belongs to E and neither v nor u place a reception or transmission null respectively towards each other.

In multihop radio networks, the same channel is reused by multiple communication links in order to increase the traffic capacity. Assume that the nodes employ omnidirectional antennas. Then a set of links $S = \{(v_1, w_1), ..., (v_M, w_M)\}$ can use the same channel if and only if the receiver w_i of each link i is within the transmission range of its own transmitter v_i and out of range of any other transmitter in the set S. That is, $(v_i, w_i) \in E$ for $i = 1, ..., M$ and $(v_j, w_i) \notin E$ for all $i, j = 1, ..., M$, $i \neq j$. For increasing network capacity, it is desirable to have sets of cochannel links with large cardinality. The larger the cochannel sets are, the larger the number of simultaneous transmissions that can be accommodated with a fixed number of channels. The problem of link scheduling to alleviate interference and increase the capacity of packet radio networks was studied extensively in the past [5, 3, 4, 9, 6, 8].

If the nodes possess array antennas, then the neighboring links can share the same channel since cochannel interference can be alleviated by null placement. Therefore, larger sets of links can share the same channel and the network capacity may increase.

3 Null Placement

Node v has an array with $e_v - 1$ elements and therefore can place $e_v - 1$ nulls, either when it acts as a receiver or as a transmitter. Since a node v can interfere or be interfered by other nodes which are within its transmission range, the nulls are placed towards those directions. Let's denote by $N(v)$ the set of nodes w which are within range of node v, that is $(v, w) \in E$. Node v places nulls towards the directions of at most $e_v - 1$ nodes of those in $N(v)$. The null placement of node v is described by the set $U(v)$ of the nodes which are nulled.

Consider a set of links $S = \{(v_1, w_1), ..., (v_M, w_M)\}$ where the nodes v_i act as transmitters and the nodes w_i act as receivers. Let $V = \{v_i : i = 1, ..., M\}$ and $W = \{w_i : i = 1, ..., M\}$ be the sets of transmitters and receivers respectively. The links in S may constitute a cochannel set if and only if for any two nodes $v_i \in V, w_j \in W, i \neq j$ which are within the transmission range one of the other $((v_i, w_j) \in E)$, either v_i or w_j place a null towards the other. The beam pattern for the set of links U is specified by the collection $U = \{U(v), v \in V \cup W\}$, that is called null placement configuration in the following. Hence, a set of links is an eligible cochannel set under a null placement configuration U if and only if

A. Nodes v_i, w_i have their main beams directed towards each other, therefore $v_i \notin U(w_i)$, $w_i \notin U(v_i)$.

B. When (v_i, w_j) belongs to $E, i \neq j$, then either $w_j \in U(v_i)$ of $v_i \in U(w_j)$. (Interference cancelation)

In determining cochannel sets with large cardinality, the next problem becomes of interest:

P: Given a set of links S and array antennas with a certain number of elements at each node, find a null placement configuration under which the set of links S is an eligible cochannel set.

Let's partition the collection of sets U representing the null placement configuration for the set of links S, in two subcollections U_v and U_w; $U_v = \{U(v_i) : i = 1, ..., M\}$, is the null placement of the transmitters and $U_w = \{U(w_i) : i = 1, ..., M\}$, is the null placement of the receivers. Note that for a specific null placement configuration U_v in the transmitters, it is simple to determine whether there is a receiver null placement configuration U_w such that S is an eligible cochannel set for the configuration $U = U_v \cup U_w$. It is enough that for every node w_i, the number of transmitters $v_j, i \neq j$ which are within range of w_i and have not placed a null towards the direction of w_i is less than or equal to $e_{w_i} - 1$. In this case the receiver w_i may cancel the interfering nodes by placing its own nulls towards their directions. Hence problem P can be rephrased as follows:

P1: Find a transmitter null placement configuration U_v such that for each receiver $w_i \in W$, the number of transmitters in V still interfering w_i after the null placement U_v is less than or equal to $e_{w_i} - 1$.

In the following we show that a feasible null placement configuration can be obtained by the solution of a maxflow problem.

Consider a directed bipartite graph G_B with sets of nodes V and W and set of edges E_B that consists of all the edges of G connecting nodes between V and W with an imposed direction from V to W. The edges $(v_i, w_i), i = 1, ...M$ are excluded from E_B. Construct a flow network G_F by augmenting G_B as follows. Augment the set of nodes $V \cup W$ by two additional nodes s (source) and d (destination). Augment the set of links by $|V|$ links directed from s to each one of the nodes in V and $|W|$ links directed from each node in W to d. Define a capacity function on the set of links as follows. Each link in E_B has capacity equal to one. A link from s to node $v_i \in V$ has capacity C_{sv_i} equal to $(d(v_i) - e_{v_i} + 1)^+$, where $d(v_i)$ is the number of links emanating from node v_i in the graph G_B, and a link from a node $w_i \in W$ to d has capacity $e_{w_i} - 1$.

The null placement problem P1 is equivalent to an integer maximum flow problem in G_F as it is argued in the following. Consider feasible integer flow vectors in G_F, that is nonnegative integer vectors f, with one component f_{vw} for each link (v, w), which satisfy the link capacity constraints and the flow conservation equations. The flow transfer of a flow vector is the sum of the flows of all links emanating from node s. The maxflow problem is to identify the flow vector with the maximum flow transfer. For more details on the maxflow problem, the reader is referred to [7]. The problem P1 and the maxflow problem in G_F are equivalent in the following sense.

There is a null placement configuration for which S is feasible cochannel set if and only if the maximum flow transfer in G_F is equal to $\sum_{v_i \in V} C_{sv_i}$.

This claim is justified in the following. Consider a flow vector f° that achieves the maximum flow transfer. For that flow vector, the flow through each link (s, v_i) will be equal to its capacity, $f^\circ_{sv_i} = C_{sv_i}$. Note that the flow of each link in E_B will be equal to 1 or 0 and because of the flow conservation equation for v_i, exactly $f^\circ_{sv_i}$ links emanating from node v_i will have flow equal to 1 and the rest equal to 0. That is, exactly $(d(v_i) - e_{v_i} + 1)^+$ links emanating from v_i have flow equal to 1. For each node w, there is a number of links with capacity equal to 1 terminating in w_i and only one link with capacity $e_{w_i} - 1$ originating from w_i to d. Because of the flow conservation equations in node w_i, at most $e_{w_i} - 1$ links with flow equal to 1 may terminate in w_i from some node in V. Consider a null placement configuration where each transmitter v_j sets the nulls towards the directions of the links that carry zero flow and each receiver w_i sets the nulls towards the directions of the links carrying flow equal to 1. This null placement is feasible since each node $v_j(w_i)$ needs to place only up to $e_{v_j} - 1$ ($e_{w_i} - 1$) nulls. Furthermore for each interfering pair $(v_j, w_i) \in E_B$, $j \neq i$, the interference is cancelled either by a null from v_j if $F_{v_j w_i} = 0$ or by a null from w_i if $F_{v_j w_i} = 1$. Hence it is shown that given a feasible flow vector with flow transfer equal to $\sum_{v_j \in V} C_{sv_j}$, a null placement configuration U for which S is an eligible cochannel set can be obtained. It can be argued similarly that if there exist a null placement configuration for which S is a feasible cochannel set, then the maximum flow transfer in G_F is equal to $\sum_{v_j \in V} C_{sv_j}$.

4 Discussion

A model for a radio network with array antennas was considered and it was shown how the radiation nulls should be placed such that a certain collection of links to constitute an eligible (interference free) cochannel set.

There are several directions for further investigation, including: more accurate modeling of the interference using path losses and exact values of the antenna radiation gain towards each direction; adaptive null placement without the knowledge of the interferers directions; consideration of the effect of the signaling schemes on the beamforming. Most of these problems have already been considered for a single receiver and it remains to be addressed in a network context.

References

[1] Y. Bresler and A. Macovski. On the number of signals resolvable by a uniform linear array. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 34:1361–1375, December 1986.

[2] Y. Bresler and A. Macovski. Optimum beamforming for coherent signals and interferences. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 36:833–843, June 1988.

[3] A. Ephremides and T. Truong. A distributed algorithm for efficient and interference-free broadcasting in radio network. *IEEE Trans. on Commun.*, COM-38(3), March 1990.

[4] B. Hajek and G. Sasaki. Link scheduling in polynomial time. *IEEE Trans. on Information Theory*, 34:910–917, September 1988.

[5] I.Chlamtac and A.Lerner. Fair algorithm for maximal link activation in multihop packet radio networks. *IEEE Trans. on Commun.*, COM-35(7):739–746, July 1987.

[6] L.Kleinrock and J. Silvester. Spatial reuse in multihop packet radio networks. *Proc. IEEE*, 75(1), January 1987.

[7] C. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice Hall, 1982.

[8] M. Post, P. Sarachic, and A. S. Kershenbaum. A biased-greedy algorithm for scheduling multi-hop radio networks. pages 564–572, March 1985.

[9] I. Proceedings. Special issue on packet radio networks. *Proc. IEEE*, 75(1), January 1987.

[10] B. Widrow, K. Duvall, R. Gooch, and W. Newman. Signal cancelation phenomena in adaptive antennas: Causes and cures. *IEEE Trans. on Antennas and Propagation*, 30:469–478, May 1982.

[11] B. Widrow, P. Mantey, L. Griffiths, and B. Goode. Adaptive antenna systems. *Proceedings of IEEE*, 57:2143–2159, December 1967.

Optimum Multiple Antenna Quantization for Reception of Fading Signals in Noise

Rick S. Blum *
Electrical Engineering and Computer Science Department
Lehigh University
Bethlehem, PA 18015

Abstract

A multiple antenna diversity scheme is investigated for digital wireless communications. Antenna observations are immediately quantized and sent to a fusion center to decide which symbol was transmitted. The optimum reception scheme is described for the case where frequency shift keying is employed and where slow Rayleigh fading and Gaussian additive noise are present. Two cases are studied. In the first case an accurate estimate of the signal-to-noise ratio is available at each receiver. In the second case estimates are not available. Results indicate that two or three bit quantizations may be most appropriate.

1 Introduction

There is significant interest in using wireless communication systems in environments where severe multipath fading is present, which can limit system performance [1]. To mitigate the effects of multipath fading, diversity techniques using multiple antennas have been proposed [2, 3] and it has been found that the performance improvements obtained by using these schemes can be significant. There appears to be a trend towards increasing the portion of wireless receivers that are implemented using digital technology in many applications. Recent improvements in electronic technology indicate that all-digital receivers are becoming practical at many frequencies of interest and further improvements in the speed of analog-to-digital converters are expected to continue this trend. These facts indicate that multiple antenna diversity schemes that combine quantized samples should be considered.

Consider a multipath fading environment where noncoherent binary frequency shift keying (FSK) is to be employed [1]. Assume that N receivers, each with an associated antenna, are to be employed to achieve a diversity gain. A nonselective fading channel is considered where the fading is assumed to be slow enough so that it can be assumed constant over several bit periods. In our explicit examples, Rayleigh fading is assumed. The observations at each receiver are assumed to include additive zero-mean Gaussian noise.

Each of the receivers will generate a multiple bit decision and a single final decision will be made by fusing the decisions from the individual receivers. Assume that synchronization between the individual receiver decisions has been achieved, so that each set of receiver decisions correspond to the same transmitted digit. We consider two cases. One case where an accurate estimate of the signal-to-noise ratio is available for the observations made at each receiver and a second case where no such estimate is available.

2 Optimum Combining

The optimum scheme (minimum probability of error) for fusing the decisions from the individual receivers is to form the likelihood ratio for the set of individual receiver decisions [5] and to compare this to a threshold. Denote the decision at the j^{th} receiver by U_j which can take on any of the values $1, \ldots, M_j$. Then the optimum final decision U_0 is to decide for a "1" sent ($U_0 = 1$) if (ones and zeros equally likely [2])

$$\sum_{j=1}^{N} \sum_{k=1}^{M_j} I(u_j = k) W_{j,k} > 0 \qquad (1)$$

where

$$W_{j,k} = \ln\left(\frac{Prob(U_j = k | 1 \text{ sent})}{Prob(U_j = k | 0 \text{ sent})}\right), \qquad (2)$$

u_j is the observed value of the random variable U_j, and $I(u_j = k)$ is an indicator function which is unity if $u_j = k$

*This material is based upon work supported by the National Science Foundation under Grant No. MIP-9211298

[1] The analysis given here is applicable to spread spectrum signaling as described in [4].

[2] The extension to cases where ones and zeros were not equally likely is straightforward. The zero on the left hand side of (1) becomes $\ln(Prob(0 \text{ sent})/Prob(1 \text{ sent}))$.

0-8186-7576-4/96 $5.00

and zero otherwise. If the left hand side of (1) is less than zero, then the final decision decides a "0" was sent. Note that we can decide "0" or "1" for the event where the left hand side of (1) equals zero without affecting performance. The form of the fusion rule given in (1) is valid in either of the two cases we consider. The calculations of $Prob(U_j = k|1 \text{ sent})$ and $Prob(U_j = k|0 \text{ sent})$ are different in each case, since these calculations depend on the schemes used by the individual receivers to generate their multibit decisions.

One special case of interest is that where receiver signal-to-noise ratio (SNR) estimates are available, and where the sensor SNRs change so slowly that the estimates can be sent to the fusion center with infinite precision. Since the update rates necessary are so slow this communication is ignored. This case was considered in [6] for individual receivers that make binary decisions. In this special case the weights are given by

$$W_{j,k} = \ln\left(\frac{Prob(U_j = k|\gamma_j, 1 \text{ sent})}{Prob(U_j = k|\gamma_j, 0 \text{ sent})}\right). \quad (3)$$

where γ_j is the SNR estimate at receiver j.

3 Optimum Receiver Quantizers

Each of the individual receivers consists of two bandpass matched filters, each matched to a sinusoid (over the bit period) with a different frequency. A sinusoid with one of these frequencies corresponds to a "1" being sent, while a sinusoid with the other frequency corresponds to a "0" being sent. The outputs of the matched filters are sampled at the end of the bit interval and then envelope detected to produce the random variables R_{0j} (large for "0" sent) and R_{1j} (large for "1" sent) at the jth individual receiver. The relative sizes of R_{0j} and R_{1j} determine the likelihood of that a "1" or "0" was sent. Thus an important quantity is the observed value of the random variable $V_j = R_{1j} - R_{0j}$ which has probability density function (pdf) $f_{V_j}(v_j|\ell \text{ sent})$ given the symbol $\ell = 0$ or $\ell = 1$ was sent.

First consider the case where no estimates of the receiver SNRs are available. The best decision scheme at the j^{th} individual receiver should perform a quantization of the likelihood ratio of the receiver observations [7]. Thus the j^{th} individual receiver should decide $U_j = k$ if $v_j \in A_{j,k}$ where

$$A_{j,k} = \left\{v_j : t_{j,k-1} \le \ln\left(\frac{f_{V_j}(v_j|1 \text{ sent})}{f_{V_j}(v_j|0 \text{ sent})}\right) < t_{j,k}\right\} \quad (4)$$

and

$$\frac{f_{V_j}(v_j|1 \text{ sent})}{f_{V_j}(v_j|0 \text{ sent})} = \frac{\int_{\gamma_j=0}^{\infty} f_{V_j|\Gamma_j}(v_j|1 \text{ sent}, \gamma_j) f_{\Gamma_j}(\gamma_j) d\gamma_j}{\int_{\gamma_j=0}^{\infty} f_{V_j}(v_j|0 \text{ sent}, \gamma_j) f_{\Gamma_j}(\gamma_j) d\gamma_j} \quad (5)$$

In (5), $f_{V_j|\Gamma_j}(v_j|\gamma_j, \ell \text{ sent})$ denotes the conditional pdf of V_j which is (for unit variance noise) [8]

$$f_{V_j}(v_j|0 \text{ sent}, \gamma_j) = \int_{r=max(0,-v_j)}^{\infty} r \exp\left(-\frac{r^2}{2}\right) \exp(-\gamma_j) I_0\left(r\sqrt{2\gamma_j}\right)(r + v_j) \exp\left(-\frac{(r+v_j)^2}{2}\right) dr \quad (6)$$

and under the appropriate symmetry conditions

$$f_{V_j}(v_j|1 \text{ sent}, \gamma_j) = f_{V_j}(-v_j|0 \text{ sent}, \gamma_j) \quad (7)$$

In (5) $f_{\Gamma_j}(\gamma_j)$ is the pdf of the signal-to-noise ratio at the jth receiver. For example, assuming Rayleigh fading gives a specific form for $f_{\Gamma_j}(\gamma_j)$ which is

$$f_{\Gamma_j}(\gamma_j) = \frac{1}{\mu_j} \exp\left(-\frac{\gamma_j}{\mu_j}\right) u(\gamma_j) \quad (8)$$

where μ_j is the average signal-to-noise ratio at the jth receiver and $u(x) = 1$ for $x \ge 0$ and is zero otherwise. Using the regions in (4) allows us to compute the required probabilities needed to calculate (1) as

$$Prob(U_j = k|\ell \text{ sent}) = \int_{v_j \in A_{j,k}} f_{V_j}(v_j|\ell \text{ sent}) dv_j, \quad (9)$$

$\ell = 0, 1$.

Now assume that an estimate of the signal-to-noise ratio of the observations at each receiver is available, which we take to be equal to the true SNR γ_j. In this case, the decisions from the jth individual receiver should be based on V_j and γ_j. The best decision scheme at the j^{th} individual receiver should perform a quantization of the likelihood ratio of (V_j, γ_j). Thus, the j^{th} receiver should decide $U_j = k$ if $(v_j, \gamma_j) \in A^e_{j,k}$ where $A^e_{j,k} =$

$$\left\{(v_j, \gamma_j) : t_{j,k-1} \le \ln\left(\frac{f_{V_j}(v_j|1 \text{ sent}, \gamma_j)}{f_{V_j}(v_j|0 \text{ sent}, \gamma_j)}\right) < t_{j,k}\right\} \quad (10)$$

with

$$\frac{f_{V_j,\Gamma_j}(v_j, \gamma_j|1 \text{ sent})}{f_{V_j,\Gamma_j}(v_j, \gamma_j|0 \text{ sent})} = \frac{f_{V_j}(v_j|1 \text{ sent}, \gamma_j)}{f_{V_j}(v_j|0 \text{ sent}, \gamma_j)} \quad (11)$$

and the required probabilities needed to calculate (1) are computed as

$$Prob(U_j = k|\ell \text{ sent}) = \int_{(v_j,\gamma_j) \in A^e_{j,k}} f_{\Gamma_j}(\gamma_j) f_{V_j}(v_j|\ell \text{ sent}, \gamma_j) dv_j d\gamma_j \ \ell = 0, 1. \quad (12)$$

4 Optimum Thresholds

For a given set of thresholds $t_{j,k}, j = 1,\ldots,N, k = 1,\ldots,M_j$ and fading statistics, the reception scheme is now well defined. The receiver thresholds are chosen to minimize the probability of error which is

$$P_e = Prob(0 \text{ sent})Prob(\text{ error } |0 \text{ sent}) + Prob(1 \text{ sent})Prob(\text{ error } |1 \text{ sent}) \quad (13)$$

where

$$Prob(\text{ error } |0 \text{ sent}) = \sum_{u_1=1}^{M_1} \cdots \sum_{u_N=1}^{M_N} Prob(U_0 = 1|U_1 = u_1,\ldots,U_N = u_N) Prob(U_1 = u_1|0 \text{ sent}) \cdots Prob(U_N = u_N|0 \text{ sent}), \quad (14)$$

$Prob(U_0 = 1|U_1 = u_1,\ldots,U_N = u_N)$ is specified by the fusion rule in (1). The quantities like $Prob(U_1 = u_1|0 \text{ sent})$ in (14) can be calculated using (12). An expression similar to (14) exists for $Prob(\text{ error } |1 \text{ sent})$ as given by

$$Prob(\text{ error } |1 \text{ sent}) = \sum_{u_1=1}^{M_1} \cdots \sum_{u_N=1}^{M_N} Prob(U_0 = 1|U_1 = u_1,\ldots,U_N = u_N) Prob(U_1 = u_1|1 \text{ sent}) \cdots Prob(U_N = u_N|1 \text{ sent}). \quad (15)$$

We have searched for the thresholds which minimize P_e in (13) by using a numerical gradient descent based technique. While it is difficult to guarantee that an absolute minimum has been found, this technique is relatively simple to apply and solutions which give good performance can be obtained easily provided only a small number of individual receivers and quantization levels are involved. As a specific example, consider a case with two individual receivers with $\mu_1 = \mu_2 = 10$ db SNR, and unit variance noise. To simplify matters assume even symmetric receiver thresholds [8]. To further simplify matters, consider the case where the set of thresholds at each receiver are constrained to be identical. Table 1 gives the best schemes we found for cases with receiver SNR estimates available and $M_1 = M_2 = 2,4,6$. The results in Table 1 are for the case where the receiver SNR estimates must be encoded in the same bits as the receiver decisions are encoded. If the SNR of the receiver observations is changing very slowly then one might assume these estimates can be sent to the fusion center without any overhead. This is the case considered in [6].

The other possibility is where SNR estimates are not available. Table 2 gives the best schemes we found for cases with no receiver SNR estimates available and $M_1 = M_2 = 2,4,6$. These results give an indication of the performance that can be obtained. The results in both

$M_1 = M_2$	P_e	$t_{j,1}$	$t_{j,2}$	$t_{j,3}$	$t_{j,4}$
2	0.0833				
4	0.0251	2.40			
6	0.0202	1.02	3.70		

Table 1. Best solutions (SNR estimate available) with SNR = 10 dB. Other receiver assumed identical ($t_{j,0} = 0$ and $t_{j,M_j/2} = \infty$ for $j = 1,2$). Rest of thresholds at $-t_{j,1}, -t_{j,2}, -t_{j,3}, -t_{j,M_j/2-1}$.

$M_1 = M_2$	P_e	$t_{j,1}$	$t_{j,2}$	$t_{j,3}$	$t_{j,4}$
2	0.0833				
4	0.0305	1.47			
6	0.0245	0.87	2.00		

Table 2. Best solutions (no SNR estimate) with SNR = 10 dB. Other receiver assumed identical ($t_{j,0} = 0$ and $t_{j,M_j} = \infty$ for $j = 1,2$). Rest of thresholds at $-t_{j,1}, -t_{j,2}, -t_{j,3}, -t_{j,M_j/2-1}$.

Table 1 and Table 2 indicate that there is a distinguishable improvement in performance when using two-bit decisions over the performance that can be obtained when using single bit decisions. For increases beyond two bits the performance gains occur more gradually (results for $M_1 = M_2 > 6$, not shown in Table 1 and Table 2, show even more gradually improvement). This suggests that two or three bit decisions may be adequate in many cases. Similar results have been obtained at other signal-to-noise ratios. These results are consistent for those obtained other quantized reception problems, [9] which are sometimes called distributed detection problems.

5 Discussion

It is not surprising that, typically, cases with SNR estimates yield better performance then cases without estimates. The best centralized scheme (no quantization) without SNR estimates is a noncoherent detection scheme which has received significant attention. If the average SNRs are identical at each individual receiver then the optimum centralized scheme is square-law combining. Even if the average SNRs are identical at each individual receiver then the optimum centralized scheme for the case where the SNR estimates are available is not square-law combining. The optimum centralized scheme is discussed in [10]. This is a case which is intermediate to that of pure coherent and pure noncoherent detection. It is useful to note that the performance of this scheme is bounded by the performance of the coherent and

noncoherent schemes, since it appears difficult to develop an analytical expression for this performance. Note that the performance of the optimum centralized scheme allows us to compute the performance of the distributed scheme as M_1 and M_2 approach ∞.

The binary receiver decision case is quite interesting, since in this case the performance for the cases given in Table 1 and Table 2 are exactly the same. This is reasonable since the best receiver thresholds, even without the symmetry or identical receiver threshold assumption, are at zero. In the non-binary cases with SNR estimates available at the receivers, the thresholds used at the individual receivers are essentially chosen to be different for each different SNR. In the binary case this does not occur, the best thresholds are always zero. Thus the SNR estimate is not actually used. In fact it is easy to show that in either case probability of error is exactly equal to that for the single individual receiver case with unknown SNR with an average value $\mu = \mu_1 = \mu_2$ which is $P_e = P_s = 1/(2+\mu)$. This means that using the SNR estimate does not improve performance and that using two rather than one individual receiver also does not improve performance.

Now consider a case with N individual receivers. In these cases the best fusion rule reduces to a majority rule which will randomly choose $U_0 = 0$ or $U_0 = 1$ if the half the receivers decide a zero was sent and half decide a one was sent (under the constraint of like sensors). The overall error probability with N individual receivers is

$$P_e = \left(\lfloor\frac{N}{2}\rfloor - \frac{N-1}{2}\right)\binom{N}{N/2} P_s^{N/2}(1-P_s)^{N/2} + \sum_{k=\lfloor N/2\rfloor+1}^{N}\binom{N}{k} P_s^k(1-P_s)^{N-k} \quad (16)$$

For $N = 1$ or $N = 2$ we see that $P_e = P_s$. In fact if N is any odd integer, (16) shows that there is no improvement due to increasing N by one. The problem is the random decision which is made if the half the receivers decide a zero was sent and half decide a one was sent. For $N > 2$ we generally find $P_e < P_s$.

For the special case where the receiver SNRs are changing slowly, so that exact SNR estimates can be sent to the fusion center without overhead, the results are different. The performance in this case must be as good or better than the other two cases we consider. In this case the fusion center can combine the receiver decisions based on the true SNR of the observations used to make each decision. For the case of two individual receivers it is easy to show that the receiver decision with the highest SNR will determine the final decision. Due to this the performance is equivalent to that for selection diversity which is $P_e = 1/(2+2\mu+\mu^2/2) < 1/(2+\mu)$ since $2+2\mu+\mu^2/2 > 2+\mu$. Thus in this case there is an improvement over the single individual receiver case.

References

[1] N. Seshadri, C. W. Sundberg and V. Weerackody, "Advanced Techniques for Modulation, Error Correction, Channel Equalization, and Diversity," *AT & T Technical Journal*, vol. 72, No. 4, pp. 48-63, Jul. 1993.

[2] J. H. Winters and J. Saltz, "The impact of Antenna Diversity on the Capacity of Wireless Communications," *IEEE Transactions on Communications*, vol. COM-42, No. 2, pp. 1740-1751, Feb. 1994.

[3] A. F. Naguib, A. Paulraj and T. Kailath, "Capacity Improvement with Base-Station Antenna Arrays in Cellular CDMA," *IEEE Transactions on Vehicular Technology*, vol. VT-43, No. 3, pp. 691-698, Aug. 1994.

[4] Y. Chau and J. Sun, "Optimal distributed detection diversity for direct-sequence CDMA over a shadowed Rician fading land-mobile satellite channel," *IEEE Vehicular Technology Conference*, Vol. 2, pp. 947-951, June 1994.

[5] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, AES-22, pp. 98-101, Jan. 1986.

[6] A. D. Kot and C. Leung, "Optimal Partial Decision Combining in Diversity Systems," *IEEE Transactions on Communications*, pp. 981-991, July 1990.

[7] I. Y. Hoballah and P. K. Varshney, "Distributed Bayesian signal detection," *IEEE Transactions on Information Theory*, vol. 35, No. 5, pp. 995-1000, Sept. 1989.

[8] R. S. Blum, "Distributed Reception of Fading Signals in Noise," *28th Annual Conference on Information Sciences and Systems*, Johns Hopkins University, Baltimore, MD, pp. 490-495, MD, March 1995.

[9] C. C. Lee and J. J. Chao, "Optimal local decision space partitioning for distributed detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, No. 4, pp. 536-543, July 1989.

[10] H. S. Abdel-Ghaffar and S. Pasupathy, "performance of the noncoherent biquadratic and GLRT receivers over two-path channels with known amplitudes and Rayleigh fading," *IEEE Transactions on Vehicular Technology*, vol. 44, No. 4, pp. 715-728, Nov. 1995.

A Multi-element Blind Adaptive Multiuser Detector *

Vafa Ghazi-Moghadam Mostafa Kaveh Laurie B. Nelson
Department of Electrical Engineering
University of Minnesota
Minneapolis, MN 55455

Abstract

Multiuser detection techniques provide attractive performance characteristics for CDMA systems. A recently proposed blind adaptive multiuser detector provides near-far resistance without requiring any more information than a conventional detector. In this paper, an optimum and two suboptimum multi-element blind adaptive receivers are proposed. These receivers exploit the spatial distribution of the users in a multiple access environment. The steady state performance of these detectors is analyzed and their complexity and the level of information required by each of them is compared.

1 Introduction

A major limiting factor in the performance of the conventional(matched filter) receiver for CDMA systems is the *near-far problem*, where a strong interferer may prevent the reliable detection of the desired user. Multiuser detection techniques provide alternatives to the conventional detector, by exploiting various levels of knowledge about the interfering signals to effect near-far resistance [4].

The optimum multiuser detector attains the performance of a single-user detector by assuming the knowledge of the signature waveform, the timing and the received amplitude of each of the users. This nonlinear detector has superior performance to the conventional detector, but is exponentially complex in the number of users. Several suboptimum multiuser detectors have been proposed which require less knowledge of the interfering signals and/or have lower complexity, but maintain near-far resistance. An example is the *decorrelating detector*, which performs a linear transformation on the outputs of the matched filter, cancelling out the effect of multiple access interference on each user . When the interfering users are weak compared to background noise level, the performance of the decorrelating detector may become worse than a conventional detector. Linear MMSE detectors solve this problem by incorporating the knowledge of the users' energies [2]. These detectors perform like a decorrelator in the presence of strong multiple access interference, and like a conventional detector, when the background noise dominates. The chief advantage of the MMSE detector, however, is in its ability to be easily implemented in an adaptive fashion. This eliminates the need for the knowledge of the signature waveforms of the interfering users. A training sequence has to be retransmitted, every time there is a severe change in the received signal, which can become cumbersome in rapidly changing environments.

A blind adaptive multiuser detector has been proposed, which overcomes this problem [1].This receiver, which only requires the knowledge of the signature sequence of the desired user and its timing (same as the conventional detector), uses as cost function, the output energy of the receiver. The receiving filter is derived by minimizing the output energy, subject to constant response to the signature waveform of the desired user. This detector, which is similar to the *generalized sidelobe canceller* in array processing, consists of two orthogonal branches, where the filter in one branch is the desired user's signature sequence, while the other filter is adapted to minimize multiple access interference. It can be shown that the mean-square-error and the output energy differ by a constant, and therefore it is possible to achieve the MMSE detector performance, e.g near-far resistance, without requiring a training sequence.

In order to reduce interference further, we present a multi-element blind adaptive multiuser detector. An optimum multi-element detector, along with two suboptimum detectors are derived, analyzed and compared. Specifically, the interaction of the spatial and

*This work was supported in part by the National Science Foundation under Grant MIP-9202081.

0-8186-7576-4/96 $5.00

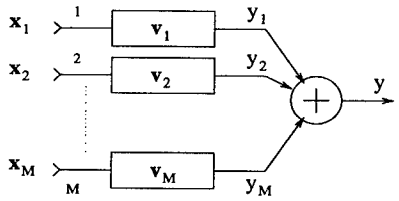


Figure 1. Optimum multi-element receiver

temporal processing stages of the multi-element blind receiver are discussed. A blind adaptive multi-element CDMA receiver which uses a conventional detector as the temporal processing stage was presented in [3]. Part of the work presented here is a combination of that blind array with the blind adaptive detector in [1].

2. Signal Model

Let K users transmit simultaneously over a pass-band channel. These transmissions are received by an array of M antennas. The propagation delay between antenna elements is assumed to be small relative to the inverse of the transmission bandwidth, i.e. the received signals at the M baseband array outputs are identical to within a complex constant. The vector of received signals is then given by

$$\mathbf{x}(t) = \sum_{k=1}^{K} \mathbf{a}_k e_k b_k s_k(t) + \sigma \mathbf{n}(t), \qquad (1)$$

where $\mathbf{x}(t) = [x_1(t) \cdots x_M(t)]^T$, $\mathbf{a}_k = [a_{1k} \cdots a_{Mk}]^T$ is the array response vector for user k, e_k^2 is the energy, $s_k(t)$ is the normalized signature waveform over symbol interval T, and $\mathbf{n}(t)$ is the vector of the additive white Gaussian noise. The $\{s_k(t)\}$ are real, linearly independent with chip rate $T_c = T/N$. Using vector representation for signature waveforms, we can rewrite the received signal as an $M \times N$ matrix:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^H \\ \vdots \\ \mathbf{x}_M^H \end{bmatrix} = \sum_{k=1}^{K} \mathbf{a}_k e_k b_k \mathbf{s}_k^T + \sigma \mathbf{N}. \qquad (2)$$

3 Optimum multi-element receiver

Figure 1 shows the block diagram of an optimum multi-element blind adaptive receiver. This can be thought of as the the natural extension of the single-element blind adaptive multiuser detector[1]. The received signal at each antenna can be written as :

$$\mathbf{x}_m = \sum_{k=1}^{K} a_{mk}^* e_k b_k \mathbf{s}_k + \sigma \mathbf{n}_m = \sum_{k=1}^{K} e_k b_k \mathbf{s}_{mk} + \sigma \mathbf{n}_m, \quad (3)$$

where a_{mk}^*, is assumed to be known. The output of the m^{th} linear filter is :

$$y_m = \sum_{k=1}^{K} e_k b_k < \mathbf{s}_{mk}, \mathbf{v}_m > + \sigma < \mathbf{n}_m, \mathbf{v}_m > . \quad (4)$$

For convenience, we assume the desired user to be k=1. As mentioned before, each linear filter consists of two orthogonal branches $\mathbf{v}_m = \mathbf{s}_{m1} + \mathbf{h}_m$, where $< \mathbf{s}_{m1}, \mathbf{h}_m >= 0$. Orthogonality of the two filters ensures that no component of desired user is passed through the adaptive branch, thereby avoiding the cancellation of the desired signal at the output of the detector. Combining the outputs of the filters results in :

$$y = \sum_{k=1}^{K} e_k b_k (\sum_{m=1}^{M} < \mathbf{s}_{mk}, \mathbf{v}_m >) + \sigma \sum_{m=1}^{M} < \mathbf{n}_m, \mathbf{v}_m > \qquad (5)$$

To adapt this receiver to detect the desired user in a blind fashion, the linear filters are chosen to minimize the output energy, while maintaining the response of the receiver to the desired user at a constant level. The optimization problem can be formulated as :

$$\min_{\mathbf{v}_1, \cdots, \mathbf{v}_M} E[yy^*] \qquad \text{subject to} \qquad \sum_{m=1}^{M} < \mathbf{s}_{m1}, \mathbf{v}_m >= 1 \qquad (6)$$

where

$$E[yy^*] = \sum_{m,m'=1}^{M} \mathbf{v}_m^H \mathbf{R}_{m,m'} \mathbf{v}_{m'}, \qquad (7)$$

$$\mathbf{R}_{m,m'} = \mathbf{S}_m \mathbf{E}^2 \mathbf{S}_{m'}^H + \sigma^2 \delta_{m,m'} \mathbf{I}_N,$$
$$\mathbf{S}_m = [\mathbf{s}_{m1}, \cdots, \mathbf{s}_{mK}],$$
$$\mathbf{E} = \mathrm{diag}(e_1, \cdots, e_K). \qquad (8)$$

Arranging $\mathbf{R}_{m,m'}$ for all $m, m' = 1, \cdots, M$ into a $MN \times MN$ matrix $\underline{\underline{\mathbf{R}}} = \{\mathbf{R}_{m,m'}\}$, and forming $\underline{\mathbf{v}} = [\mathbf{v}_1^T, \cdots, \mathbf{v}_M^T]^T$ and $\underline{\mathbf{s}_1} = [\mathbf{s}_{m1}^T, \cdots, \mathbf{s}_{M1}^T]^T$, we can rewrite the optimization problem as :

$$\min_{\underline{\mathbf{v}}} \underline{\mathbf{v}}^H \underline{\underline{\mathbf{R}}}\, \underline{\mathbf{v}} \qquad \text{subject to} \quad \underline{\mathbf{v}}^H \underline{\mathbf{s}_1} = 1 \qquad (9)$$

The solution to this problem is known to be

$$\underline{\mathbf{v}} = \frac{\underline{\underline{\mathbf{R}}}^{-1} \underline{\mathbf{s}_1}}{\underline{\mathbf{s}_1}^H \underline{\underline{\mathbf{R}}}^{-1} \underline{\mathbf{s}_1}} \qquad (10)$$

4 Suboptimum multi-element receiver

The block diagram of the suboptimum multi-element blind adaptive multiuser detector is shown in

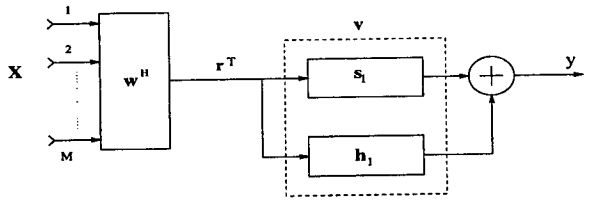


Figure 2. Suboptimum multi-element receiver

Figure 2. The first stage of this receiver is a beamformer, where the received signals at each antenna are weighted and combined. The beamformer output is sent through an adaptive linear filter. Consider $\mathbf{w} = [w_1 \cdots w_M]^T$ to be the weight vector, then the output of the beamformer is given by :

$$\mathbf{r}^H = \mathbf{w}^H\mathbf{X} = \sum_{k=1}^{K}(\mathbf{w}^H\mathbf{a}_k)e_k b_k \mathbf{s}_k^T + \sigma(\mathbf{w}^H\mathbf{N}) \quad (11)$$

If $\mathbf{v}$ is the impulse response of the linear filter, the output of this receiver can be written as:

$$y = \mathbf{w}^H\mathbf{X}\mathbf{v} = \sum_{k=1}^{K}(\mathbf{w}^H\mathbf{a}_k)e_k b_k < \mathbf{s}_k, \mathbf{v} > +\sigma(\mathbf{w}^H\mathbf{N}\mathbf{v}) \quad (12)$$

The beamformer in the first stage translates the multi-dimensional problem into a single dimension problem, the only difference is that the output of the receiver is now a function of both $\mathbf{v}$ and $\mathbf{w}$, so a performance improvement due to added spatial discrimination is expected. We call this receiver *suboptimum*, mainly because the total number of adaptive filter taps is $M+N$, compared to MN taps for the optimum receiver. It has reduced complexity , but inferior performance to the optimum receiver.

4.1 Known array response vector

In this section, we assume that the suboptimum receiver knows the array response vector of the desired user, and develop the adaptation rule for the beamformer weights, as well as the taps of linear filter. The cost function to optimize is the output energy of the receiver. To detect the desired user, $\mathbf{w}$ and $\mathbf{v}$ are varied to minimize the output energy, subject to the constraints of having constant spatial gain in the direction of the desired user, and constant temporal gain for the signature waveform of the desired user. In other words,

$$\min_{\mathbf{v},\mathbf{w}} E[yy^*] \quad \text{subject to} \quad \begin{aligned} < \mathbf{s}_1, \mathbf{v} > &= 1 \\ \mathbf{w}^H\mathbf{a}_1 &= 1. \end{aligned} \quad (13)$$

Using Eq. 12, the output energy can be written as :

$$E[yy^*] = \sum_{k=1}^{K} e_k^2 < \mathbf{v}, \mathbf{s}_k >^2 |\mathbf{w}^H\mathbf{a}_k|^2 + \sigma^2\mathbf{w}^H\mathbf{w} < \mathbf{v}, \mathbf{v} > \quad (14)$$

The solution to the optimization problem is :

$$\mathbf{v} = \frac{\mathbf{R}_s^{-1}\mathbf{s}_1}{\mathbf{s}_1^T\mathbf{R}_s^{-1}\mathbf{s}_1}; \qquad \mathbf{w} = \frac{\mathbf{R}_a^{-1}\mathbf{a}_1}{\mathbf{a}_1^H\mathbf{R}_a^{-1}\mathbf{a}_1}, \quad (15)$$

where

$$\begin{aligned} \mathbf{R}_s &= \mathbf{S}\mathbf{A}_w\mathbf{S}^T + \sigma^2(\mathbf{w}^H\mathbf{w})\mathbf{I}_N, \\ \mathbf{R}_a &= \mathbf{A}\mathbf{S}_v\mathbf{A}^T + \sigma^2 < \mathbf{v}, \mathbf{v} > \mathbf{I}_M, \\ \mathbf{S} &= [\mathbf{s}_1, \cdots, \mathbf{s}_K,] \\ \mathbf{A} &= [\mathbf{a}_1, \cdots, \mathbf{a}_K], \\ \mathbf{A}_w &= \mathrm{diag}(|\mathbf{w}^H\mathbf{a}_1|^2, \cdots, |\mathbf{w}^H\mathbf{a}_K|^2), \\ \mathbf{S}_v &= \mathrm{diag}(< \mathbf{v}, \mathbf{s}_1 >^2, \cdots, < \mathbf{v}, \mathbf{s}_k >^2). \end{aligned} \quad (16)$$

Since $E[yy^*]$ is a function of both $\mathbf{w}$ and $\mathbf{v}$ the solutions for $\mathbf{w}$ and $\mathbf{v}$ are not independent, making it very difficult to analyze the system.

4.2 Unknown array response vector

In this part we consider a situation, where no information about the array response vector of the user of interest is available. This is mostly the case in practice, where incoherent modulation and multipath propagation make it impossible to have a correct estimate of the array response vector at the receiver. The proposed receiver uses the knowledge of the signature sequence of the desired user to blindly adapt the beamformer weights, so that the desired user is passed, while the energy of the interfering signals is minimized at the output of the beamformer. The output energy of the beamformer is used as the cost function; the beamformer weights are chosen, so that this energy is minimized. A constraint must be used, so that the trivial solution is avoided. As explained before, the linear filter maintains a constant response to the desired user, while it adapts itself to minimize the interference. Thus, we can use for the constraint, a constant energy at the output of the receiver in Figure 2. So the optimization rule for the beamformer weights can be written as follows :

$$\min_{\mathbf{w}} E[\mathbf{r}^H\mathbf{r}] \quad \text{subject to} \quad E[yy^*] = 1, \quad (17)$$

where,

$$\begin{aligned} E[\mathbf{r}^H\mathbf{r}] &= \mathbf{w}^H\mathbf{R}\mathbf{w}, \\ \mathbf{R} &= \mathbf{A}\mathbf{E}^2\mathbf{A}^H + N\sigma^2\mathbf{I}_M, \\ E[yy^*] &= \mathbf{w}^H\mathbf{R}_a\mathbf{w}. \end{aligned} \quad (18)$$

User	1	2	3	4
ρ_{1k}	1	3/7	-1/7	-3/7
AOA	0°	5°	−25°	55°
SNR(dB)	10	14	20	14

Table 1. Angle of arrivals and SNRs

Receiver	$M=1$		$M=2$		
	Conv	Blind	Opt	Sub	
				$\mathbf{a}_1$	–
SNIR(dB)	−13.7	6.7	12.1	11	8

Table 2. SNIRs for various receivers

The solution to this problem is given by the generalized eigenvector corresponding to the minimum eigenvalue of the matrix pencil $(\mathbf{R}, \mathbf{R}_a)$. The optimization rule for the linear filter is same as previous case. Because of stronger constraint used for the beamformer, the first receiver has better performance, but the advantage of the second one is that it does not require any more knowledge than the conventional detector and the single-element blind detector.

5 Performance analysis

In this section, we analyze and compare the steady state performance of the proposed detectors with the single-element blind detector and the conventional detector. As the performance measure, we use the signal to interference plus noise ratio. We also compare the beampatterns of the two suboptimum detectors, which employ a beamformer as their first stage. We consider a system with $K = 4$ users, processing gain of $N = 7$ and $M = 2$ antennas.

Figure 3 shows the spatial distribution of the four users, along with the beampatterns corresponding to the two suboptimum beamformers discussed in the previous section. Table 1 lists the angle of arrivals, the crosscorrelations with user 1 and SNRs for the desired user and the interfering users.

It can be seen that adding an antenna improves the signal to interference ratio of the receiver. The spatial gain depends on the algorithm used. An interesting fact about the suboptimum multi-element detectors is the way the spatial and temporal processing stages interact. We can see that although the detector with no knowledge of $\mathbf{a}_1$ puts a stronger null in the direction of the strongest interferer, the overall performance of the other detector is better. This is because the level of interference rejection of the linear temporal filter is a *nonlinear* function of the strength of interference at its

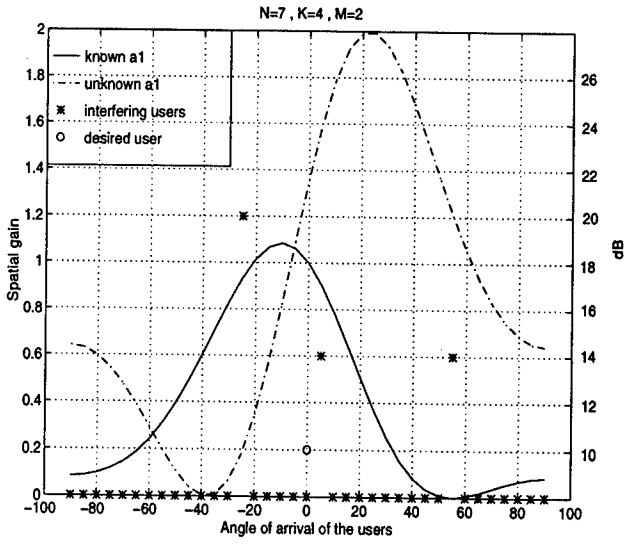


Figure 3. Beampatterns of the suboptimum receivers

input. The interplay between interference rejection by temporal and spatial processing is under investigation.

6 Conclusion

The steady state performances of three multi-element blind adaptive receivers were studied. It can be seen that all these receivers can take advantage of the spatial distribution of the users to reduce the level of interference. More study needs to be done to investigate the robustness of these receivers, when there is no spatial discrimination between the users. Future work will also include the analysis of the adaptive algorithm for these receivers, and the study of their performance in an asynchronous multipath environment.

Acknowledgment The authors appreciate discussions with Prof. M. Honig of Northwestern University.

References

[1] M. Honig, U. Madhow, and S. Verdú. Blind adaptive multiuser detection. *IEEE Trans. on Info. Theory*, 41, July 1995.

[2] U. Madhow and M. Honig. MMSE interference suppression for direct-sequence spread-spectrum CDMA. *IEEE Trans. on Comm.*, 42, Dec. 1994.

[3] C. Vaidyanathan, K. Buckley, and S. Hosur. A blind adaptive antenna array for cdma systems. In *Proc. of the Asilomar Conf. on Sig., Sys. and Comp.*, Nov. 1995.

[4] S. Verdú. Adaptive multiuser detection. In *Proc. of the Third IEEE Int. Symp. on Spread Spectrum Techniques and Appl.*, Oulu, Finland, July 1994.

New Bounds On Rake Structures for DS-CDMA Over Frequency-Selective Rayleigh Fading Channels

Sergio Calvo, Gregori Vazquez
Polytechnic University of Catalonia. Dpt. of Signal Theory and Communications
c/ Gran Capità s/n, 08034 Barcelona, SPAIN. Tel: 34-3-4016440 Fax: 34-3-4016447
e-mail: gregori@gps.tsc.upc.es, sergio@gps.tsc.upc.es

Abstract

An upper bound is derived for the probability of error in an asynchronous binary direct-sequence spread-spectrum multiple-access communications system operating over frequency selective Rayleigh fading channels. A coherent RAKE receiver with predetection selective diversity combining is considered. The performance of a multipath-combinig receiver is determined for the case of multiple interfering transmitters. Furthermore, the performance of the system is determined in terms of parameters of the signature sequences. These parameters can be used as guides in selecting sequences for the system. The bounds agree with the exponential portion of a normal distribution in which the interfering interference components subtract from the signal amplitude. The results obtained are verified by simulation.

I. Introduction

In this paper, we consider the performance analysis of the multipath-combinig receiver also called RAKE receiver. The analysis applies to systems that use binary phase-shift-keyed (BPSK) modulation. We consider a multipath-combinig receiver and determine the performance of the system for the case of multiple interfering transmitters that use different PN sequences but having a small factor of correlation.

The system designer use to assume that the limiting corrupting signal has a Gaussian distribution. This assumption can no longer be justified, and the interference in multiple-access schemes will not show a Gaussian characteristics. Although a small thermal noise power will be present, it is necessary to consider the joint effect of the Gaussian-distributed thermal noise and the non-Gaussian distributed interference.

The analysis will be based on the well known Chernoff bound. This analysis requires only the evaluation or bounding of the moment-generating function of the additive interference. The bound is expressed in terms of a parameter that is the unique solution of an equation containing the derivative of the moment generating function of the interference. The bound is tight for high SNRs, which is the region of interest in most mobile communications systems.

This work has been partially supported by the Spanish Research Council (CYCIT) under grant TIC95-1022-C05-01

II. A System and Channel Model

The system of communications proposed for this study is shown in figure (1).

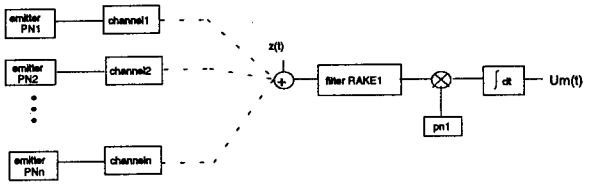


FIG 1. Communications system with several simultaneously active users

The k-th transmitted signal for a binary DS CDMA system with BPSK modulation and arbitrary chip waveform can be expressed as

$$s_k(t) = \sqrt{2P_k} \sum_i b_k(i)\psi(t)u_k(t - iT) \qquad (1)$$

where P_k is the power in each K-th transmitted signal, $u_k(t)$ is the signature waveform and $b_k(i)$ the i-th symbol of the k-th user. An equal power assumption ,P_k, is made for convenience in the analysis. We assume that there are N code chips in each data symbol ($T=NT_c$) and the period of the signature sequence ($u_k(t)$) is N. The received signal from a typical transmitter consists of a random number of paths of the transmitted signal. The delay, amplitude, and phase associated with each path are also random.

$$r_k(t) = \sum_{n=1}^{L} \alpha_n(t)\exp\left(j\Phi_n(t)\right)s_k\left[t - \tau_n(t)\right] + z(t) \qquad (2)$$

In (2) the random variable L represents the number of paths of the k-th user. The random variables $\alpha_n(t)$, $\tau_n(t)$, $\Phi_n(t)$, represent the amplitude, delay, and phase associated with the nth path of a signal from transmitter k. We

0-8186-7576-4/96 $5.00

consider the delay $\tau_n(t)$ as an integer of Tc. The term $z_k(t)$ represents the additive white Gaussian noise (AWGN) with complex spectral density No Watt/Hz.

We consider the fact that in a real communication system the links from each of the K active transmitters to the listening receiver are mutually independents. The random variables variables $\alpha_n(t)$, $\tau_n(t)$, $\Phi_n(t)$, and the data symbols, are assumed to form a set of mutually independent random variables. Also, each phase $\Phi_n(t)$ is assumed to be uniformly distributed on the interval $[0,2\pi)$. We model the arrival of signal paths at receiver, $\tau_n(t)$, by a nonhomogeneous Poisson process with the arrival rate $P_d(x)$. The amplitudes of the signal paths, $\alpha_n(t)$, exhibit a Rayleigh distribution. The structure of the receiver is shown in figure (2).

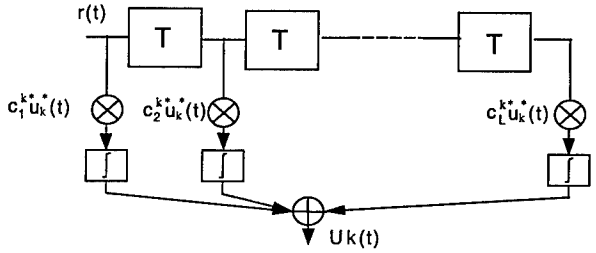


fig 2. Rake Structure

It employs a single tap delay line trough which is passed the received signal $r_k(t)$. The signal at each tap is correlated with $(c_n^k)^*(t)u_k^*(t)$, n = 1, 2, ..., L, where $u_k(t)$ is the PN sequence of the k-th desired user, $c_n^k = \alpha_n^k \exp\left(j\Phi_n^k\right)$ the impulse response of the channel associated to the desired user and (*) denotes complex conjugate.

III. Evaluation of Error Probability Bound

We shall now evaluate the performance of the RAKE receiver with simultaneous active users under the condition that the fading is sufficiently slow to allow us to estimate $c_n^k(t)$ perfectly (without noise). Thus the decision variables may be expressed in the form

$$U_k(i) = \Re e \sum_{q=1}^{L} (c_q^k)^* \int_T r(t)u_k^*(t - qTc)dt \qquad (3)$$

Stationarity of the amplitudes paths of arrivals is assumed. The received signal can be expressed as follows

$$r(t) = \sum_i b_k(i) \sum_{n=1}^{L} c_n^k u_k(t - nTc - iT) + \sum_{nf=1,\, nf\neq k}^{Nf} \sum_i b_{nf}(i) \sum_{p=1}^{L} c_p^{nf} u_{nf}(t - pTc - iT) + z(t) \qquad (4)$$

where $u_{nf}(t)$ are the PN sequences and $b_{nf}(i)$ the symbols of the Nf users that share the same transmission band. Eq (3) can be expressed as:

$$U_k = \Re e \sum_{q=1}^{L} \left|c_q^k\right|^2 b_k \int_T u_k(t - qTc)u_k^*(t - qTc)dt + \Re e \sum_{nf=1,\, nf=k,q\neq n}^{Nf} \sum_{q=1n=1}^{L}\sum^{L} \alpha_q^k \alpha_n^{nf} \exp j(\Phi_n^{nf} - \Phi_q^k) b_{nf} \int_T u_{nf}(t - nTc)u_k^*(t - qTc)dt + \Re e \sum_{q=1}^{L} \alpha_q^k \exp(-j\Phi_q^k) \int_T z(t)u_k^*(t - qTc)dt \qquad (5)$$

When the signals are antipodal, a single decision variable suffices. Then, if consider the maximum cross-correlation value between the set of signature waveforms, Eq (5) can be simplifies to the following:

$$U_k \approx \Re e \left(b_k E \sum_{q=1}^{L} (\alpha_q^k)^2 + \gamma E \sum_{nf=1,\, nf=k,q\neq n}^{Nf} \sum_{n=1}^{L}\sum_{q=1}^{L} \alpha_q^k \alpha_n^{nf} e^{j\Psi_{q,n}^{Nf}} + \sum_{q=1}^{L} \alpha_q^k N_q^k \right) \qquad (6)$$

with: $N_q^k = e^{-j\Phi_q^k} \int_T z(t)u_k^*(t - qTc)$ and $e^{j\Psi_{q,n}^{nf}} = e^{j(\Phi_n^{nf} - \Phi_q^k)} b_{nf}$

$$\gamma = \frac{2^{(m+1)/2} + 1}{2^m - 1} \geq \left| \frac{\int_T u_k(t - mTc)u_q(t - pTc)dt}{\int_T u_k(t)u_k(t)dt} \right|_{k = q,\, p \neq m}$$

where $E = \int_T u_k(t)u_k^*(t)dt$ is the energy of the k-th signature waveform and the normalized cross-correlation bound holds for m-length Gold sequences. The interference will be modeled as follows:

$$\eta = \gamma E \sum_{i=1}^{NfL'} \xi_i \cos\Theta_i \qquad (7)$$

being ξ_i a random variable that results from the product of two independent Rayleigh distributed random variables, Θ_i is assumed to be uniformly distributed in the interval $[0,2\pi)$ and L'=L(L-1). We define the total interference z as $z = \eta + n$, that is, the multi-user interference plus de thermal noise term. As far as z is the sum of two independent, zero-mean random variables, the Chernoff bound applies:

$$\Pr(z > x) \leq \exp^{-\lambda x} E\left[\exp \lambda n\right] E\left[\exp \lambda \eta\right] \quad \text{for all } \lambda > 0 \qquad (8)$$

after some manipulations (8) becomes:

$$\Pr(z > x) \leq \exp^{-\lambda x} \exp^{-\frac{1}{2}\lambda^2\sigma_n^2} \prod_{i=1}^{NfL'} I_0\left(\lambda\xi_i\right) \quad \text{for all } \lambda > 0 \qquad (9)$$

Equation (10) can be simplified by using exponential upper bounds to the order zero modified Bessel function

$$\Pr(z > x) \leq \exp^{-\lambda x} \exp^{\frac{1}{2}\lambda^2\sigma_n^2} \exp^{\sum_{i=1}^{NfL'} \lambda\xi_i} \text{ for all } \lambda > 0 \qquad (10)$$

The optimal value of λ which minimizes the right side of (10) is obtained by setting the derivative of the expression with respect to λ to zero:

$$\Pr(z > x) \leq \exp - \frac{\left(E\sum_{i=1}^{L'} \alpha_i^2 - E\gamma \sum_{i=1}^{L'Nf} \xi_i\right)^2}{\sigma_n^2} \qquad (11)$$

where $\sigma_n^2 = E\,No \sum_{i=1}^{L'} \alpha_i^2$. We define γ_b as:

$$\gamma_b = \frac{E\sum_{i=1}^{L'} \alpha_i^2 - E\gamma \sum_{i=1}^{L'Nf} \xi_i}{\sqrt{\sigma_n^2}} \qquad (12)$$

The final step in this derivation is to average the conditional error probability given in (11) over the fading channel statistics. Thus we evaluate the integral:

$$Pe = \int_0^\infty Pr(\gamma_b) f(\gamma_b) d\gamma_b \qquad (13)$$

It is difficult to find the distribution function of (12), and with the resultant expression is not possible to obtain a closed expression of (13). Thus we consider the average of σ_n^2 instead the exact expression. Defining $E[\alpha_n^2] = 2\sigma^2$,the expression (12) reduces to:

$$\gamma_b = \sqrt{\frac{E}{2N_o\sigma^2 L'}}\left(\sum_{i=1}^{L'} \alpha_i^2 - \gamma \sum_{i=1}^{NfL'} \xi_i\right) \qquad (14)$$

It can be shown that:

$$f(\gamma_b) = \frac{e^{-\frac{\gamma_b}{2\sigma^2\Omega}}}{\Omega(L'-1)!(2M-1)!(\sigma^2)^{L'2M}\,2^{L'}} \sum_{j=0}^{L'-1} \binom{L'-1}{j} \left(\frac{\gamma_b}{\Omega}\right)^{L'-1-j} \Gamma(j+2M) \frac{1}{\beta^{j+2M}}$$

with M= L'Nf and $\beta = \frac{\gamma+2}{2\gamma\sigma^2}$ when the variance of the amplitudes of the paths of arrival belonging to different users are equal. We can define Ω as:

$$\Omega = \sqrt{\frac{E}{2NoL'\sigma^2}} \qquad (15)$$

Then, the error probability (13) results in

$$Pe \le A \sum_{j=0}^{L'-1} Bj \exp\left(\frac{1}{4\sigma^2\Omega}\right)^2 \frac{\sqrt{\pi}}{2}(L'-1-j)! \sum_{k=0}^{\infty} \frac{(-1)^k \left(\frac{1}{4\sigma^2\Omega}\right)^k}{2^{L'-1-j-k}\,k!\,\Gamma\left(1+\frac{L'-1-j-k}{2}\right)}$$

where the terms in this series correponding to k=L'-j+1, L'-j+3, L'-j+5, ... are understood to be zero.

$$A = \frac{1}{\Omega(L'-1)!(2L'Nf)!(\sigma^2)^{2L'M}\,2^{L'}\,\gamma^{2L'Nf}}$$

$$Bj = \binom{L'-1}{j} \Gamma(j+2L'Nf) \frac{1}{\beta^{j+2L'Nf}} \frac{1}{\Omega^{L'-1-j}}$$

IV. Numerical Results

The format of the signal has the following parameters: bandwidth W = $(100nsec)^{-1}$ = 10 Mhz ; spread factor TW = 127 ; transmission rate, 1/T = 78.7 Kbits/sec. A maximum of 71 chips delay spread has been considered and thus the intersimbolic interference will be negligible. The transmission scheme of figure (1), in which up to 6 pseudorandom sequences of maximum period theoretically mutually orthogonal, have been simulated, corresponding to five possible interfering users. Each user's sequence is multiplied by the symbol to be transmitted (BPSK modulated). Once the signals are modulated, they go through a selective frequency time variant channel. A different channel has been considered for each user. Value L' is computed as L(L-1) where L is taken as the mean of the paths generated at each impulse responses of the different users and it results a value of 15 paths. Considering that the variance of the arriving paths has been assumed to be the same for each channel $E[\alpha k^2] = 0.12$, the results from figure (3) are obtained when only one interfering user is present. Figures (4) and (5) show the probability of error and its upper bound for three and five interfering users respectively and finally, figures (6) and (7) show only the upper bound for one, three and five interfering users for different signal to noise ratios.

V. Conclusions

Several interesting facts are obtained from the near-far effect in an optimum receiver under multipath operation conditions. These include the effect of the increase in the factor γ and the number of interfering users on the BER, and, on the other hand, the limitation of the system capacity as function of the number users.

It is important to observe that small increases in the factor γ can affect the BER in the event of multiple users access. In figure (8) for one interfering user, it can be observed than an increase of γ can affect considerably the BER of the system. It should be taken into account that for the optimum detection, it is in a certain way desirable for the multipath problem, as the receiver, adapted to the channel of the mobile of interested, enhance the appropriate paths and discriminate the paths arriving from other PN sequences (diversity).

VI. Bibliography

[1] M. Abramovitz, Handbook of Mathematical Functions, Dover publications, 1970

[2] B.R.Saltzberg, " Intersymbol Interference Error Bounds with application to Ideal Bandlimited Signaling", IEEE Transactions on Information Theory, vol IT-14, n° 4, July 1968

[3] D.Borth and M.Pursley, " Analysis of Direct-Sequence Spread-Spectrum Multiple-Access Communications Over Rician Fading Channels ", IEEE Transactions on communications, vol. com-27, n° 10, October 1979

[4] E.Geraniotis, " Direct-Sequence Spread-Spectrum Multiple-Access Communications Over Nonselective and Frequency-Selective Rician Fading Channels", IEEE Transactions on communications, vol. com-34, n° 8, August 1986.

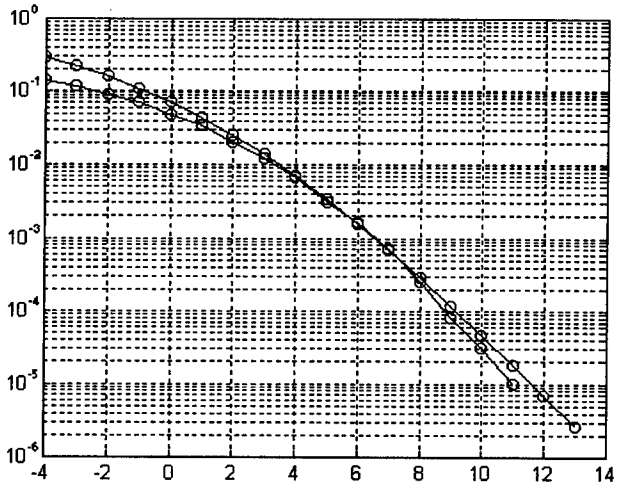


fig 3. Probability of error with one interfering user

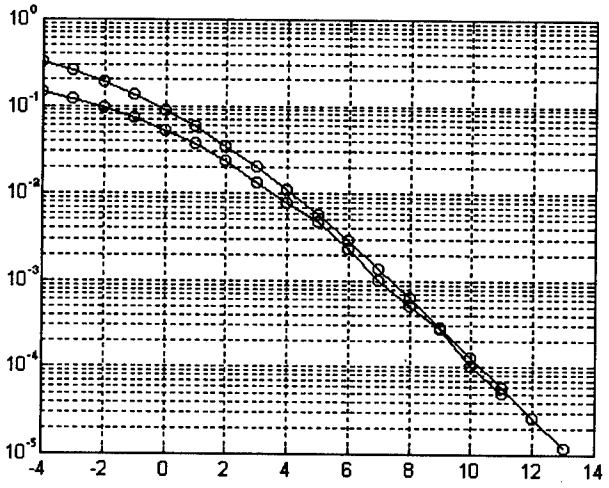


fig 4. Probability of error with three interfering users

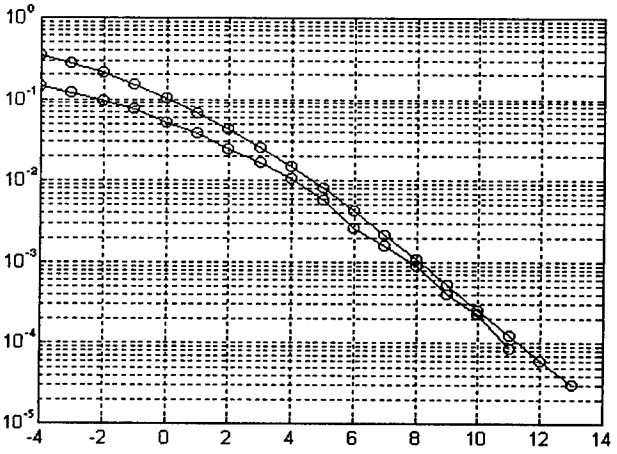


fig 5. Probability of error with five interfering users

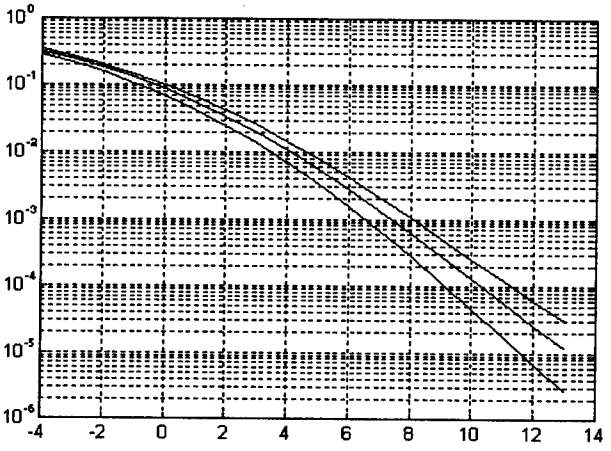


fig 6. Theoretic results

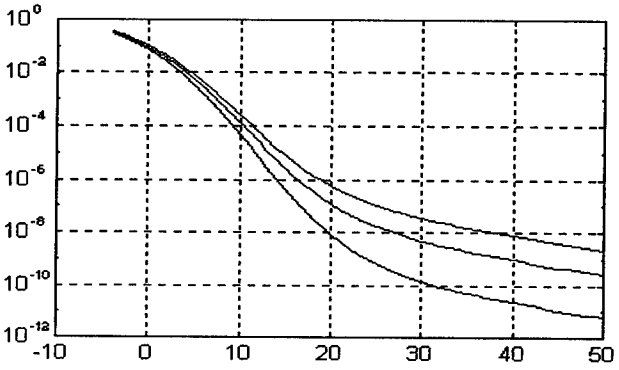


fig 7. Probability of error with 1,3 and 5 interfering users

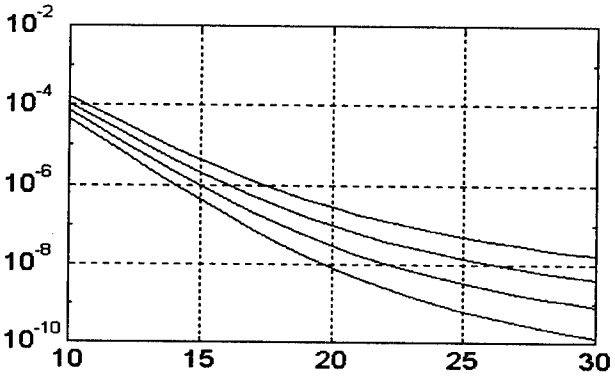


fig 8. Probability of error for one interfering user with $\gamma = 0.14, 0.16, 0.18, 0.20$ and L'=15

[5] J.Lehnert and M.B.Pursley, "Multipath Diversity Reception of Spread-Spectrum Multiple-Access communications", IEEE Transactions on communications, vol. com-35, nª 11, November 1967

[6] B.Glance and L.J.Greenstein, " Frequency-Selective Fading Effects in Digital Mobile Radio with Diversity Combining", IEEE Transactions on communications, vol. com-31, nª 9, September 1983.

ROBUST DATA DETECTION IN ASYNCHRONOUS DS-CDMA IN THE PRESENCE OF TIMING UNCERTAINTY

Jaume Riba, Jason Goldberg, and Gregori Vázquez

Department of Signal Theory and Communications, E.T.S.E. Telecomunicació, Campus Nord UPC.
Edifici D-5, c/. Gran Capità. s/n, 08034 Barcelona, Spain, Email: {jriba,jason,gregori}@gps.tsc.upc.es

ABSTRACT

The decorrelating and minimum mean squared error data detectors for direct sequence code division multiple access (DS-CDMA) communications systems are known to exhibit low vulnerability to the near-far problem. Nevertheless, the performance of these algorithms is highly sensitive to accurate knowledge of the user propagation delays as well as inter-symbol and/or inter-chip interference such as that produced by frequency-selective fading channels. In this paper, a new sub-optimum symbol-by-symbol detector is presented which is robust in the presence of these two effects.

1. INTRODUCTION

Direct-sequence code division multiple access (DS-CDMA) communications systems have recently received increased attention as a promising candidate for emerging mobile digital radio networks. For this reason, much work has been reported on the problem of multi-user detection in DS-CDMA. For asynchronous systems, the standard matched filter-bank detector is known to fail for users of widely disparate power (the so-called "near-far problem"). Moreover, the unrealistically high computational complexity of the optimum (i.e., minimum probability of error) detector [1] has motivated research on sub-optimum multi-user detectors [2]-[3].

Such systems rely on exact knowledge of additional parameters such as carrier phase, signal strength, and propagation delay for each user. However, such receivers can exhibit high sensitivity to errors in estimates of these parameters, especially propagation delay, as was shown to be the case for the decorrelating detector in [4]. Similar problems would be observed for frequency-selective fading channels. This issue has been addressed in [5] where a detection scheme relying on a multipath ray model was proposed. The technique requires estimation of the propagation delays of the individual rays of each user, and data detection is subsequently performed by forming a linear combination of symbol estimates associated with *each* ray.

This paper presents a simple, direct approach to robust data detection in the presence of uncertain propagation delay estimates and/or frequency selective fading. The technique is based on a generalization of a maximum signal to interference plus noise ratio (MSINR) symbol-by-symbol detector (the generalization to a block approach is straightforward). The multipath/propagation time uncertainty is taken into account by a simple statistical model. The resulting detectors are shown to be near-far resistant and insensitive to "small" errors in propagation delay estimation and/or frequency selective multipath channels with "small" delay spread. The technique may also be useful in cases where accurate timing estimates are available but can be updated with relatively low frequency. In this case, the receiver is robust is the presence of small changes in timing which take place over the propagation time up-date interval. As with other sub-optimum approaches computational complexity is linear in the number of users.

This work has been supported in part by EEC Contract HCM/CHRXCT-930405, PRONTIC/CICYT TIC95-1022-C05-01 and CIRIT/Generalitat de Catalunya GRQ93-3021.

2. PROBLEM FORMULATION

Consider a K user asynchronous DS-CDMA system nominally operating over a channel with additive white Gaussian noise (AWGN). Binary Phase Shift Keying (BPSK) modulation is used. Using the notation of [4, 5], the symbol interval will be denoted as T and the chip interval as $T_c = T/N$, where N is the number of chips per symbol. The kth user's code waveform is of unit amplitude and is denoted by $b_k(t)$. It can be expressed as a pulse amplitude modulation of $\{c_k(n)\}_{n=0}^{N-1}$, the pseudo-noise (PN) sequence associated with the kth user:

$$b_k(t) = \sum_{n=0}^{N-1} c_k(n) p(t - nT_c) \tag{1}$$

where $p(t)$ is a pulse whose duration, in general, exceeds the chip interval, T_c. The data sequence for the kth user, $d_k(m) \in \{-1, +1\}$, is pulse amplitude modulated by a single period of the corresponding code waveform resulting in a baseband signal written as:

$$s_k(t) = \sum_{m=-\infty}^{\infty} d_k(m) b_k(t - mT). \tag{2}$$

The transmitted signal is the product of the baseband signal and the carrier: $\sqrt{2\gamma_k}\cos(\omega_c t + \theta'_k)$ where ω_c, γ_k, and θ'_k respectively denote carrier frequency, kth user power and carrier phase.

In general, the channel associated with the kth user can be modeled as a linear time-varying system $h_k(t, \tau)$ which denotes the channel response at time t to an impulse applied τ seconds in the past. The received signal is:

$$r'(t) = n'(t) + \sum_{k=1}^{K} \mathrm{Re}\left\{\sqrt{2\gamma_k}\int_{-\infty}^{\infty} h_k(t,\tau) s_k(t-\tau) e^{j(\omega_c t - \omega_c \tau + \theta'_k)} d\tau\right\} \tag{3}$$

where $n'(t)$ is AWGN of two-sided power spectral density level, $N_o/2$. The equivalent complex baseband representa-

0-8186-7576-4/96 $5.00

tion of the signal is given as:

$$r(t) = n(t) + \sum_{k=1}^{K} \sqrt{2\gamma_k} \int_{-\infty}^{\infty} h_k(t,\tau) s_k(t-\tau) e^{j(-\omega_c \tau + \theta_k')} d\tau \tag{4}$$

If the channel is modeled as a simple constant delay (possibly not precisely known) then its impulse response can be written as $h_k(t,\tau) = \delta(t-\tau_k)$ where it will be assumed that $\tau_k \in [-T/2, T/2)$. In this case the received signal is:

$$r(t) = n(t) + \sum_{k=1}^{K} \sqrt{2\gamma_k} e^{j\theta_k} s_k(t-\tau_k) \tag{5}$$

where $\theta_k = -\omega_c \tau_k + \theta_k'$. It is this model that will be used throughout the paper.

Next define the received signal vector as $\mathbf{r}(m) \in \mathbb{C}^{QN}$ as the sum of a signal vector $\mathbf{s}(m)$ and noise vector $\mathbf{n}(m)$:

$$\begin{aligned} \mathbf{r}(m) &= \left[r(mT), r\left(mT + \frac{T}{QN}\right), \cdots, r\left([m+1]T - \frac{T}{QN}\right) \right]^T \\ &= \mathbf{s}(m) + \mathbf{n}(m) \end{aligned} \tag{6}$$

where $(\cdot)^T$ denotes transpose. It is not difficult to verify that the signal vector can be expressed as:

$$\mathbf{s}(m) = \mathbf{J} \sum_{k=1}^{K} \sum_{i=-1}^{1} \mathbf{b}_k(\tau_k + iT)\, d_k(m+i) \tag{7}$$

$$\mathbf{J} = [\mathbf{0}_{QN}, \mathbf{I}_{QN}, \mathbf{0}_{QN}] \in \mathbb{R}^{QNx3QN} \tag{8}$$

$$[\mathbf{b}_k(\tau_k + iT)]_n = \sqrt{2} b_k([n-1]T/QN - [i+1]T - \tau_k) \tag{9}$$

$$n \in \{1, 2, \cdots, 3QN\}, i \in \{-1, 0, 1\}$$

where the subscripts on the square zero and identity matrices in (8) denote the dimension of these matrices, and $[\cdot]_n$ denotes the nth element of a vector. The above expression can be written more compactly in matrix form:

$$\begin{aligned} \mathbf{s}(m) &= \mathbf{J} \sum_{i=-1}^{1} \mathbf{B}(\tau + i\mathbf{T})\, \mathbf{d}(m+i) \\ \mathbf{B}(\tau + i\mathbf{T}) &= [\mathbf{b}_1(\tau_1 + iT), \cdots, \mathbf{b}_K(\tau_K + iT)] \\ \mathbf{d}(m+i) &= \left[\sqrt{\gamma_1} e^{j\theta_1} d_1(m+i), \cdots, \sqrt{\gamma_K} e^{j\theta_K} d_K(m+i)\right]^T \\ \tau &= [\tau_1, \cdots, \tau_K], \quad \mathbf{T} = T\underbrace{[1, \cdots, 1]}_{K \text{ elements}} \end{aligned} \tag{10}$$

The correlation matrix of the received vector is:

$$\begin{aligned} \mathbf{R} &= E\left[\mathbf{r}(m)\mathbf{r}^H(m)\right] \\ &= \sum_{i=-1}^{1} \mathbf{J}\mathbf{B}(\tau + i\mathbf{T})\, \mathbf{\Gamma} \mathbf{B}^H(\tau + i\mathbf{T})\, \mathbf{J}^H + \sigma^2 \mathbf{I} \end{aligned} \tag{11}$$

$$\mathbf{\Gamma} = E\left[\mathbf{d}(m+i)\, \mathbf{d}^H(m+i)\right] \tag{12}$$

where $E[\cdot]$ is the expectation operator. It is assumed that each user's symbols are uncorrelated with those of other users (implying that $\mathbf{\Gamma}$ is diagonal with diagonal elements equal to user powers).

Let us define the signal-to-interference-and-noise-ratio (SINR) and the mean-squared-error (MSE) as:

$$\begin{aligned} \mathrm{SINR}(\mathbf{w}_k) &= \frac{\mathbf{w}_k^H \mathbf{R}_{s_k} \mathbf{w}_k}{\mathbf{w}_k^H \mathbf{R}_{[i+n]_k} \mathbf{w}_k} \\ \mathrm{MSE}(\mathbf{w}_k) &= E\left[\left|d_k(m) - \mathbf{w}_k^H \mathbf{r}(m)\right|^2\right] \end{aligned} \tag{13}$$

The "signal" correlation matrix $\mathbf{R}_{s_k}$ is defined as the correlation matrix of the signal vector in the presence only of the mth symbol of the kth user:

$$\mathbf{R}_{s_k} = \gamma_k \mathbf{J}\mathbf{b}_k(0)\, \mathbf{b}_k^H(0)\, \mathbf{J}^H \tag{14}$$

where it is noted that, without loss of generality, the associated propagation delay is set to zero: $\tau_k = 0$. Conversely, the interference-plus-noise correlation matrix $\mathbf{R}_{[i+n]_k}$ is defined as the correlation matrix of the received vector in the presence only of the noise, the remaining $K-1$ users and the $(m-1)$th and $(m+1)$th symbols of the kth user:

$$\mathbf{R}_{[i+n]_k} = \mathbf{R} - \mathbf{R}_{s_k}. \tag{15}$$

Consider now the maximum SINR (MSINR) symbol-by-symbol receiver for user k. It is well known that this problem can be solved using generalized eigenanalysis. The general solution of the MSINR receiver can then be written as:

$$\begin{aligned} \mathbf{w}_{o_k(MSINR)} &= \arg \max_{\mathbf{w}_k \in \mathbb{C}^{QN}} \mathrm{SINR}(\mathbf{w}_k) \\ &= \alpha \mathbf{e}_{max}\left[\mathbf{R}_{s_k}, \mathbf{R}_{[i+n]_k}\right] \end{aligned} \tag{16}$$

where $\mathbf{e}_{max}$ denotes the generalized eigenvector associated with the maximum eigenvalue of the above matrix pencil, and α is an arbitrary (non-zero) constant. We can also consider a minimum mean squared error (MMSE) symbol-by-symbol receiver for user k similar to [6]:

$$\begin{aligned} \mathbf{w}_{o_k(MMSE)} &= \arg \min_{\mathbf{w}_k \in \mathbb{C}^{QN}} MSE(\mathbf{w}_k) \\ &= \beta \mathbf{R}^{-1} \mathbf{p}_k \qquad \mathbf{p}_k = E[\mathbf{r}(m) d_k^*(m)] \end{aligned} \tag{17}$$

where β is a constant. In the particular case of a rank-one matrix $\mathbf{R}_{s_k}$, it is well known $\mathbf{w}_{o_k(MSINR)} = \mathbf{w}_{o_k(MMSE)}$. It is also well known that if the noise power σ^2 is very low compared to powers of the interferering users, the MMSE receiver acts as a decorrelator completely nulling the effect of the interferering users. That is to say, the magnitude of receiver output for the kth user will be approximately zero for each of the interfering users at the times for which the output provides an estimate for the kth user's symbols. However, under these low noise conditions, it has recently been shown that inaccurate timing estimates for the users can drastically reduce performance resulting in high receiver sensitivity to near-far effects [4]. The problem addressed in this paper is the design of near-far resistant receivers that are robust in the presence of such timing errors.

3. EFFECT OF TIMING ERRORS

In practice, since τ is not directly available, *estimates* of the propagation delays, $\hat{\tau} = [\hat{\tau}_1, \cdots, \hat{\tau}_K]^T$, are used in the design

of the receivers. In this section, the effects of timing error on output SINR and MSE as defined in (13) are quantified.

It will be assumed that the timing estimates, $\hat{\tau}_k$, can be expressed as a sum of the true propagation delay, τ_k, and a zero mean Gaussian random variable, ϵ_k of variance $\sigma^2_{\epsilon_k}$.

$$\hat{\tau}_k = \tau_k + \epsilon_k, \qquad E_\epsilon[\epsilon_k] = 0, \quad E_\epsilon[\epsilon_k \epsilon_{k'}] = \sigma^2_{\epsilon_k}\delta(k-k') \tag{18}$$

where E_ϵ is the expectation over the K propagation delay errors. The above model also accounts for the multipath effects appearing in frequency selective fading channels (with small delay-spread). Then, in presence of timing errors, a corresponding average output SINR can be defined as:

$$\overline{\mathrm{SINR}}(\mathbf{w}_k) = \frac{\mathbf{w}_k^H \overline{\mathbf{R}}_{s_k} \mathbf{w}_k}{\mathbf{w}_k^H \overline{\mathbf{R}}_{[i+n]_k} \mathbf{w}_k} \tag{19}$$

$$\overline{\mathbf{R}}_{s_k} = E_\epsilon[\mathbf{R}_{s_k}] \qquad \overline{\mathbf{R}}_{[i+n]_k} = E_\epsilon[\mathbf{R}_{[i+n]_k}] \tag{20}$$

Similarly, a corresponding average MSE can be defined as:

$$\overline{\mathrm{MSE}}(\mathbf{w}_k) = E_\epsilon E\left[\left|d_k(m) - \mathbf{w}_k^H \mathbf{r}(m)\right|^2\right] \tag{21}$$

We propose the definitions (19) and (21) as measures of performance in presence of timing uncertainty.

4. ROBUST DETECTOR

In the absence of timing errors, the decorrelating detector (and the MMSE detector as $\sigma^2 \to 0$) completely null out the influence of all interfering signature waveforms at their specified timings (as defined by their propagation delays). In the presence of timing errors, such nulling is not guaranteed. Even for small errors the reduction in performance as measured by decrease in average SINR (19) or by increase in average MMSE (21) can be very significant, especially for high near-far scenarios. The effect of the timing errors as seen by averaging over the propagation delays corresponds to a sort of temporal smearing of each of the K user's signals. A detector which is designed to be robust in the presence of timing errors should take this smearing effect into account in order to create *broad* temporal nulls for the interfering users. To this end, we can define the robust MSINR and robust MMSE receivers as:

$$\mathbf{w}_{r_k(MSINR)} = \arg \max_{\mathbf{w}_k \in \mathbf{C}^{QN}} \overline{\mathrm{SINR}}(\mathbf{w}_k) \tag{22}$$

$$= \mathbf{e}_{max}\left[\overline{\mathbf{R}}_{s_k}, \overline{\mathbf{R}}_{[i+n]_k}\right]$$

$$\mathbf{w}_{r_k(MMSE)} = \arg \min_{\mathbf{w}_k \in \mathbf{C}^{QN}} \overline{\mathrm{MSE}}(\mathbf{w}_k) \tag{23}$$

$$= \beta \overline{\mathbf{R}}^{-1} \overline{\mathbf{p}}_k \qquad \overline{\mathbf{p}}_k = E_\epsilon E[\mathbf{r}(m) d_k^*(m)]$$

Due to the expectation over the kth propagation delay error, ϵ_k, the "signal" correlation matrix, $\overline{\mathbf{R}}_{s_k}$, has lost the rank-one property and, therefore, the two solutions are not equal in this case: $\mathbf{w}_{r_k(MSINR)} \neq \mathbf{w}_{r_k(MMSE)}$.

Finally, transformation to the Fourier domain, where a time-shift corresponds to a linear phase, will aid in formulation of simple closed form expressions for the average correlation matrices (20). In particular, we define a Discrete Fourier Transform (DFT) matrix as:

$$\mathbf{F} = \frac{1}{\sqrt{3QN}} \begin{bmatrix} 1 & e^{j2\pi\frac{M}{3QN}} & \cdots & e^{j2\pi\frac{M(3QN-1)}{3QN}} \\ 1 & e^{j2\pi\frac{M-1}{3QN}} & \cdots & e^{j2\pi\frac{(M-1)(3QN-1)}{3QN}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j2\pi\frac{M}{3QN}} & \cdots & e^{-j2\pi\frac{M(3QN-1)}{3QN}} \end{bmatrix}$$

$$M = (3QN-1)/2.$$

Now, resorting to the time-shift property, the DFT of the users' code waveforms can be writen as:

$$\mathbf{A}(\tau+\epsilon) = \mathbf{F}\mathbf{B}(\tau+\epsilon) = \mathbf{A}(\tau) \odot \mathbf{V}(\epsilon)$$
$$\mathbf{A}(\tau) = \mathbf{A}(0) \odot \mathbf{V}(\tau) \tag{24}$$

where $\odot$ denotes the element-wise Schur product, and the linear-phase matrix $\mathbf{V}(\mathbf{x})$ and vector $\mathbf{v}(x)$ are defined respectively as:

$$\mathbf{V}(\mathbf{x}) = [\mathbf{v}(x_1), \mathbf{v}(x_2), \cdots, \mathbf{v}(x_K)] \tag{25}$$
$$\mathbf{v}(x) = \left[e^{-j2\pi Mx/3T}, e^{-j2\pi(M-1)x/3T}, \cdots, e^{j2\pi Mx/3T}\right]^H$$

The unitary property of $\mathbf{F}$ and the linearity of the expectation operator imply:

$$\overline{\mathbf{R}} = \mathbf{J}\mathbf{F}^H\left[\left(\mathbf{A}(\tau)\,\Gamma\,\mathbf{A}^H(\tau)\right) \odot \mathbf{S} \odot \mathbf{Q}\left(\sigma^2_{\epsilon_{max}}\right)\right]\mathbf{F}\mathbf{J}^H + \sigma^2\mathbf{I}$$

$$\mathbf{S} = \sum_{i=-1}^{1} \mathbf{v}(iT)\,\mathbf{v}^H(iT) \tag{26}$$

$$\overline{\mathbf{R}}_{s_k} = \gamma_k \mathbf{J}\mathbf{F}^H\left[\left(\mathbf{a}_k(0)\,\mathbf{a}_k^H(0)\right) \odot \mathbf{Q}\left(\sigma^2_{\epsilon_{max}}\right)\right]\mathbf{F}\mathbf{J}^H \tag{27}$$

$$\overline{\mathbf{R}}_{[i+n]_k} = \overline{\mathbf{R}} - \overline{\mathbf{R}}_{s_k} \tag{28}$$

where as in [7]:

$$[\mathbf{Q}(\sigma^2_{\epsilon_{max}})]_{pq} = \left[E_\epsilon\left[\mathbf{v}(\epsilon)\,\mathbf{v}^H(\epsilon)\right]\right]_{pq} = e^{-2\sigma^2_{\epsilon_{max}}[\pi(p-q)]^2}$$

The problem of the exact propagation delays appearing in the parameterized computation of $\overline{\mathbf{R}}$ in (26) can be addressed by simply using the estimated delays, $\hat{\tau}_k$, in the arguments of the $\mathbf{V}(\cdot)$ and $\mathbf{v}(\cdot)$ in the computation of $\mathbf{A}(\tau)$ using (24) and (25).

Thus, in summary, the new robust MSINR receiver filter is formed by using (27) and (28) with estimated propagation delays in (22). The robust receiver (22) offers, as verified in the next section, the following compromise with respect to the optimum receiver defined in (16):

$$\mathrm{SINR}(\mathbf{w}_{o_k}) > \mathrm{SINR}(\mathbf{w}_{r_k}) \approx \overline{\mathrm{SINR}}(\mathbf{w}_{r_k}) >> \overline{\mathrm{SINR}}(\mathbf{w}_{o_k})$$

5. RESULTS

In this section, results of computer simulations of the performance of the new technique (22) and comparison with that of the ordinary MMSE/MSINR receiver (17) are presented. Consider a $K = 3$ user with $Q = 2$ samples per chip and $N = 31$ chips per symbol. Nyquist pulses with roll-off 0.5 are used, and, for convenience $T = 1$. The performance of receiver's for user $k = 1$ are consider with $\tau_1 = 0$, $\tau_2 = T/3$, and $\tau_3 = -T/3$ and user powers $\gamma_1 = 1$, $\gamma_2 = 50$,

and $\gamma_3 = 50$. To gain insight into the effect of the new procedure, consider Fig. 1, the output of the conventional receiver (designed for the above scenario) for user one when only one symbol of user two is present at its input. The receiver succeeds in placing two sharp null at times $t = 0$ and $t = T$ over the interfering user. Now, consider the same experiment but with the robust receiver of (22) as shown in Fig. 2. This time two *broad* temporal nulls are placed at times $t = 0$ and $t = T$. These broad nulls are what provides the robustness to timing uncertainty.

Next, for the above user powers and noise power $\sigma^2 = 0.1$, Fig. 3 shows the output SINR as defined in (19) for the ordinary receiver (*assumed* timing error variance zero) and the robust receiver (*assumed* timing error variance, $\sigma^2_{\epsilon_{max}} = 0.003$) as the *true* timing error variance is varied. The curves indicate that, even for relatively low near-far and low timing error variance, the conventional receiver is highly sensitive to uncertainty in the propagation delays while the robust receiver offers nearly constant performance with timing error variance. Lastly, Fig. 4 shows performance for the conventional and robust receivers as a function of the ratio of interferer power, $\gamma_2 = \gamma_3$, to desired user power, γ_1. The true timing error variance as well as that used in the design of the the robust receiver are again set to $\sigma^2_{\epsilon_{max}} = 0.003$. The conventional receiver is far more sensitive to interferer power than the robust technique.

6. CONCLUSIONS

A new method for the design of multi-user detectors which are robust in the presence of propagation time estimation errors and/or frequency selective multipath (with delay spread on the order of a few chips) has been presented. The new detector is near-far resistant and offers greatly improved performance over the conventional detector for a variety of scenarios. The technique is also useful for robust detection in cases where highly accurate timing estimates are available but can only be updated relatively infrequently.

7. REFERENCES

[1] S. Verdú. "Minimum probability of error for asynchronous Gaussian multiple access channels". *IEEE Transactions on Information Theory*, vol. 32: pp. 85–96, January 1986.

[2] R. Lupas and S. Verdú. "Near-far resistance of multiuser detectors in asynchronous channels". *IEEE Transactions on Communications*, vol. 38: pp. 496–508, April 1986.

[3] Z. Xie, R.T. Short, and C.K. Rushforth. "A family of suboptimum detectors for coherent multiuser communication". *IEEE Journal on Selected Areas in Communications*, vol. 8: pp. 683–690, May 1990.

[4] S. Parkvall, E. Ström, and B. Ottersten. "The impact of timing errors on the performance of linear DS-CDMA receivers". submitted to *IEEE Journal on Selected Areas in Communications*, April 1995.

[5] S. Parkvall and E. Ström. Parameter estimation and detection of DS-CDMA signals subject to multipath propagation. In *Proceedings of IEEE/IEE Workshop on Signal Processing in Multipath Environments*, Glasgow, Scotland. IEEE/IEE, 1995.

[6] E. Ström and S. Parkvall. Joint parameter estimation and detection DS-CDMA signals in fading channels. In *Proceedings of IEEE Global Telecommunications Conference*, 1995.

[7] T. Trump and B. Ottersten. "Estimation of nominal direction of arrival and angular spread using an array of sensors". submitted to *Signal Processing*, 1994.

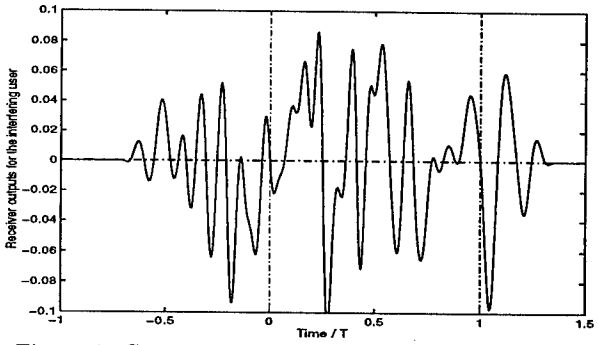


Figure 1: Conventional receiver one output vs. time

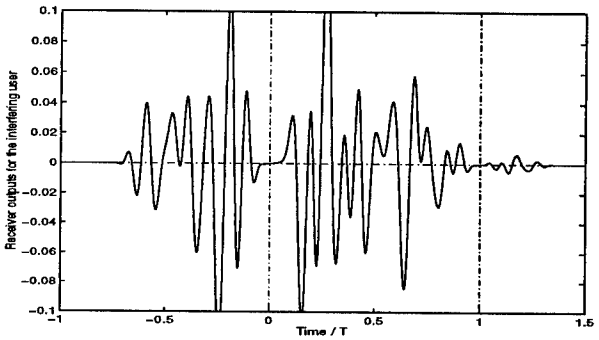


Figure 2: Robust receiver one output vs. time

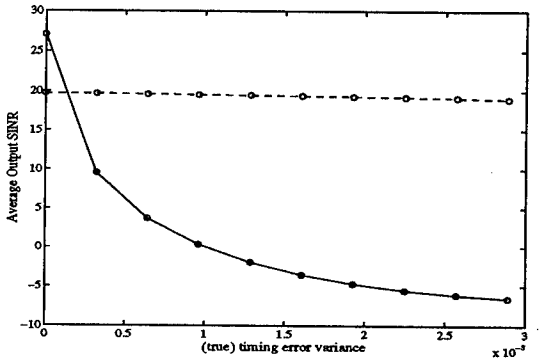


Figure 3: SINR vs. timing error var. (– conv., - - rob.)

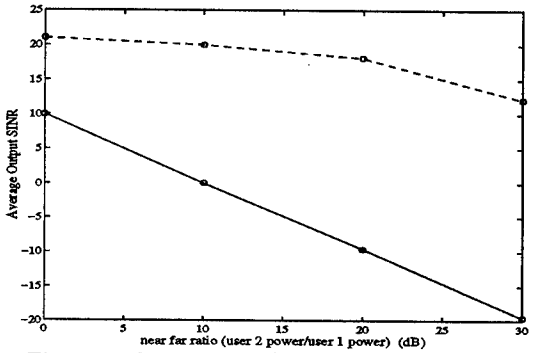


Figure 4: SINR vs. γ_2/γ_1 (– conv., - - rob.)

An LMS Array for CDMA Wireless Communication Systems

Frank Fruth and Evaggelos Geraniotis
Center for Hybrid and Satellite Communication Networks
Institute for Systems Research
University of Maryland College Park, MD 20742

Abstract

The use of adaptive arrays in a multi-rate multi-media CDMA network is analyzed, and simulation results are provided for a candidate two-rate system. Our proposed approach uses reference signal-based adaptation (LMS) for antenna weight control. Simulation results show adaptive arrays can significantly enhance the multi-media services that can be provided by a CDMA network. In addition, preliminary simulation results for RLS and LMS weight control algorithms are presented.

1. Introduction

Due to the growing user demands for wireless communication services, there has been much interest in finding alternative methods of increasing capacities of wireless systems beyond that achievable by today's systems. One emerging concept that has been receiving much attention is Space Division Multiple Access (SDMA). In SDMA, performance improvements are realized by exploiting the spatial distribution of users through spatial filtering provided by the use of antenna arrays at the base station. The application of SDMA techniques to wireless CDMA systems has been investigated in [3, 5, 7, 8], and these studies show that significant increases in cellular system capacities are realizable using antenna arrays. These papers, however, focus on cellular systems supporting a single user type, typically considered to be voice.

Future wireless systems, such as PCS, will be required to handle multi-media traffic types - voice, data, and video – which can have different data rates, as well as, different quality of service (BER) requirements. The flexibility of CDMA in accommodating multi-rate users makes it a promising technique for future wireless communication networks. But, as is the case for any conventional CDMA system, capacity is limited by practical considerations, available signal power and system bandwidth. In multi-media networks the capacity determines the quantity and diversity of traffic that can be supported by the network. In order to make multi-media system economically viable for large-scale implementations, methods of increasing capacities are needed. Therefore, in this paper we analyze the implementation of adaptive arrays in a multi-rate multi-media CDMA network and show their benefits in enhancing the multi-media services that can be provided by the network.

In addition, a critical issue associated with the implementation of adaptive arrays is the method for antenna weight control. Our proposed approach uses reference signal-based adaptation (LMS) for antenna weight control, as opposed to the "code filtering" approach introduced in [7, 4]. In this paper, we will take a preliminary look at the reference signal weight control algorithms, LMS and RLS, for the multi-media CDMA network.

In section 2, our system model is described. In section 3, the BER analysis for a multi-rate CDMA system with array processing is presented along with numerical results. In section 4, we present simulation results for LMS and RLS antenna weight control algorithms. And finally in section 5, we conclude with some final remarks.

2. System Model

In this analysis, we consider a single-cell fixed chip-rate multi-media CDMA system which provides service to S different user types with data rates given by $(R_1, R_2, \ldots, R_S)$. For simplicity, the k^{th} type-i user will be denoted as user (i,k). Since the system has a fixed chip-rate for all users, each user will have a processing gain dependent on its data rate, $N_i = R_c/R_i$, where R_c is the common chip rate.

The base station consists of a uniformly spaced linear array of M omnidirectional antenna elements. The antenna spacing is assumed to be half-wavelength, and time delays due to propagation across the array are modeled as phase shifts (i.e. narrowband array assumption). Figure 1 shows a block diagram of the base station receiver. The generation of the reference signal through the feedback loop was originally proposed by Compton [1], and will be discussed more in section 4.

0-8186-7576-4/96 $5.00

For this analysis, we will be interested in only the reverse link (mobile to base station) performance. We assume there exist direct paths between the mobiles and the base station. Multipath is not considered, but will be addressed in a future paper. The reverse link modulation is BPSK with coherent demodulation. All type i users are assumed power controlled, such that the average received signal power at each antenna element is P_i. A residual power control error for each user is included, and is modeled as a zero-mean log-normal random variable with variance σ_{dB}. The com-

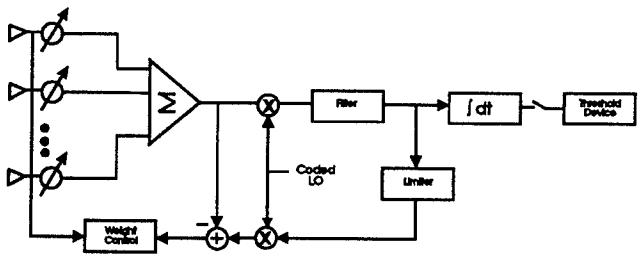


Figure 1. Base station receiver for desired user

plex representation of the received signal for user (i,k) is given as:

$$\tilde{\mathbf{x}}_{i,k}(t) = \sqrt{2P_i}10^{\varepsilon_{i,k}/20}\chi_i b_{i,k}(t-\tau_{i,k})c_{i,k}(t-\tau_{i,k})e^{j(\omega_c t-\theta_{i,k})}\underline{\mathbf{a}}_{i,k}$$

where $10^{\varepsilon_{i,k}/20}$ is the term representing the residual power control error, ω_c is the common carrier frequency, $\tau_{i,k}$ is the time delay with respect to the reference antenna element. We assume without loss of generality that $\tau_{i,k}$ is a random variable uniformly distributed over $[0, T_i = 1/R_i)$. $\theta_{i,k} = \phi_{i,k} + \tau_{i,k}\omega_c$ is the phase shift with respect to the reference antenna element, and is modeled as a random variable uniformly distributed over $[0, 2\pi)$. χ_i is a binary random variable representing the activity of the type i user, $Pr[\chi_i = 1] = \nu_i$. $b_{i,k}(t)$ represents the data waveform consisting of an i.i.d. sequence of rectangular pulses of amplitude ± 1 with duration T_i. $c_{i,k}(t)$ represents the code waveform consisting of a sequence of rectangular pulses of amplitude ± 1 with duration T_c. And $\underline{\mathbf{a}}_{\mathbf{i,k}}$ is the antenna response vector given by:

$$\underline{\mathbf{a}}_{i,k} = [1, e^{-j\gamma_{i,k}}, e^{-2j\gamma_{i,k}}, \ldots, e^{-(M-1)j\gamma_{i,k}}]^T$$

$$\gamma_{i,k} = \frac{2d\pi}{\lambda} sin\beta_{i,k}$$

where $\beta_{i,k}$ is the angular location of user (i,k) with respect to the broadside of the base station antenna array.

The total received signal at the antenna array is:

$$\overline{\mathbf{x}}(t) = \sum_{i=1}^{S}\sum_{k=1}^{K_s}\tilde{x}_{i,k}(t) + \overline{\mathbf{n}}(t)$$

where K_i is the number of type i users in the cell, and $\overline{\mathbf{n}}(t)$ is a complex gaussian process with

$$E[\overline{\mathbf{n}}^*(t)\overline{\mathbf{n}}^T(\tau)] = \sigma^2\mathbf{I}\delta(t-\tau)$$

3. Multi-rate CDMA BER Performance

3.1. BER Analysis

In this section we present our analytical approach for evaluating the average BER for user type-s. The combining of individual antenna element responses is optimal in the sense that it minimizes the squared error between the antenna array output and a given reference signal. To conduct our analysis, we consider an arbitrary scenario which consists of $(K_1, K_2, \ldots, K_S)$ users whose angular locations are given by $\underline{\beta} = (\beta_{1,1}, \beta_{1,2}, \ldots, \beta_{S,K_S})$. In addition, the instantaneous residual power control errors for each user is given by $\underline{\varepsilon} = (\varepsilon_{1,2}, \ldots, \varepsilon_{S,K_S})$, where $\varepsilon_{i,k}$ is the instantaneous residual power control error for user (i,k).

Our desired user is assumed to be (s,p) for which the optimal antenna weights are given by the Wiener solution:

$$\underline{\mathbf{w}}_{s,p}(\underline{\beta},\underline{\varepsilon}) = R_{xx}^{-1}\underline{\mathbf{a}}_{s,p}^* = (w_1 e^{j\psi_1}, \ldots, w_M e^{j\psi_M})$$

$$\mathrm{R_{xx}} = \mathrm{E}[\overline{\mathbf{x}}^*(\mathrm{t})\overline{\mathbf{x}}^{\mathrm{T}}(\mathrm{t})] = \sum_{\mathrm{i=1}}^{\mathrm{S}}\sum_{\mathrm{k=1}}^{\mathrm{K_S}} 2\mathrm{P_i}10^{\varepsilon_{\mathrm{i,k}}/20}\underline{\mathbf{a}}_{\mathrm{i,k}}^*\underline{\mathbf{a}}_{\mathrm{i,k}}^{\mathrm{T}} + \sigma^2\mathbf{I}$$

The output of the adaptive array for our desired user is:

$$\begin{aligned}\mathrm{y_{s,p}(t)} &= \mathrm{Re}[\underline{\mathbf{w}}_{\mathrm{s,p}}^{\mathrm{T}}\overline{\mathbf{x}}(\mathrm{t})] \\ &= \sum_{\mathrm{i=1}}^{\mathrm{S}}\sum_{\mathrm{k=1}}^{\mathrm{K_i}}\sqrt{2\mathrm{P_i}}10^{\varepsilon_{\mathrm{i,k}}/20}\chi_{\mathrm{i}}\mathrm{b_{i,k}(t-\tau_{i,k})c_{i,k}(t-\tau_{i,k})} \\ &\quad \sum_{\mathrm{m=1}}^{\mathrm{M}}\mathrm{w_m cos}(\omega_{\mathrm{c}}\mathrm{t}-\theta_{\mathrm{i,k}}-(\mathrm{m}-1)\gamma_{\mathrm{i,k}}+\psi_{\mathrm{m}}) + \mathrm{n'(t)}\end{aligned}$$

where $n'(t) = Re[\underline{w}_{s,p}^T\overline{\mathbf{n}}(t)]$ is a zero-mean gaussian random variable with variance $\frac{1}{2}\sigma^2\|\underline{w}_{s,p}\|^2$.

The array output is processed by a correlation receiver for the user desired user. Assuming $\tau_{s,p} = \theta_{s,p} = 0$, the output of the correlation receiver is given by:

$$\mathrm{z_{s,p}(t)} = \int_{\mathrm{T_s}} \mathrm{y_{s,p}(t)c_{s,p}(t)cos}\omega_{\mathrm{c}}\mathrm{tdt}$$

Using gaussian approximations for the other user interference, and using the results developed in [6], we compute the desired users SINR and BER for the given scenario $(\underline{\beta},\underline{\varepsilon})$ as

$$SINR_{s,p}(\underline{\beta},\underline{\varepsilon}) = \frac{E[z_{s,p}(nT_s)|\underline{\beta},\underline{\varepsilon}]}{\sqrt{Var[z_{s,p}(nT_s)|\underline{\beta},\underline{\varepsilon}]}}$$

$$BER_{s,p}(\underline{\beta},\underline{\varepsilon}) = Q(SINR(\underline{\beta},\underline{\varepsilon}))$$

And we can now obtain an average BER for the type s user by

$$\overline{BER}_s = \int_{\underline{\beta}} \int_{\underline{\varepsilon}} BER_{s,p}(\underline{\beta},\underline{\varepsilon}) f(\underline{\beta}) f(\underline{\varepsilon}) d\underline{\beta} d\underline{\varepsilon}$$

3.2. Numerical Results

The complexities associated with analyzing adaptive arrays does not permit a closed for expression for $\overline{BER}$, therefore in the this subsection we Monte Carlo simulations to evaluate average BER performance of a candidate two-rate CDMA system.

We consider a CDMA system providing service to two users types with data rates $R_1 = 9600$ bps and $R_2 = 19200$ bps, and associated processing gains $N_1 = 128$ and $N_2 = 64$. We assume a single 120° cell sector with 1, 3, 5, and 7 antenna elements. The variance of the residual power control error is fixed at .1 dB. The number of type 1 users, K_1, is fixed at 60, while the number of type 2 users, K_2, is varied from 5 to 25. The users are assumed uniformly distributed over the cell sector. The average BER for both user types is

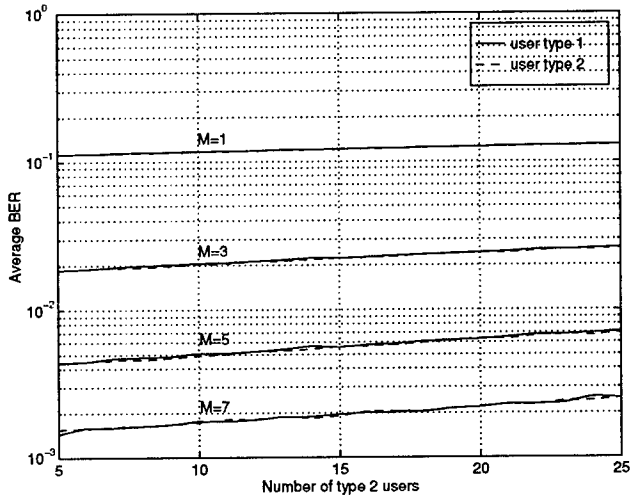


Figure 2. Average BER for user type 1 and 2 with $E_b^{(1)}/N_0 = E_b^{(2)}/N_0 = 0$ dB

analyzed under various power control approaches.

First, we consider the case of equal received bit energy-to-noise at the antenna elements, E_b/N_0 for both users. Figure 2 shows the average BER for user type 1 and 2 for $E_b^{(1)}/N_0 = E_b^{(2)}/N_0 = 0$ dB. This corresponds to $P_2 = 2P_1$. Also, both users are assumed to have an activity factor of 1. From these results we clearly see the improvements in average BER for both users as we go from 1 to 7 antenna elements at the base station. With one antenna performance is unacceptable, while with five and seven antenna elements we are able to maintain $\overline{BER} < 10^{-2}$ for all users. Also, note that each user experiences a graceful degradation in BER as we increase the number of type 2 users.

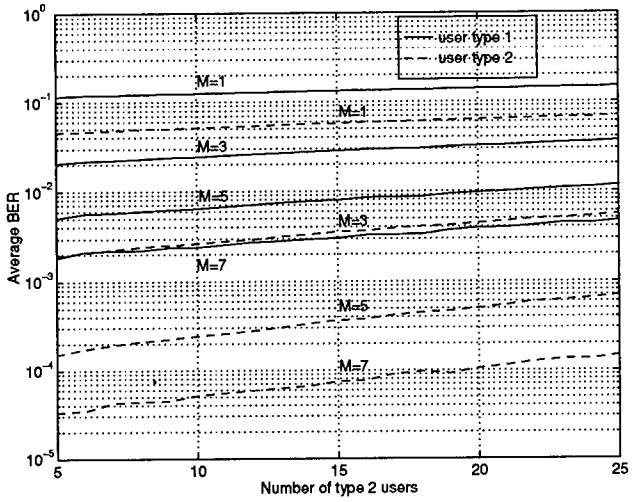


Figure 3. Average BER for user type 1 and 2 with $E_b^{(1)}/N_0 = 0$ dB and $E_b^{(2)}/N_0 = 3$ dB

Next, we consider the case when $E_b^{(1)}/N_0$ is kept at 0 dB, and $E_b^{(2)}/N_0$ is increased to 3 dB. This corresponds to $P_2 = 4P_1$. Figure 3 shows BER performance under this condition. Here, user type 1 BER performance has degraded slightly, and degrades more rapidly with the addition of type 2 users. However, user type 2 realizes over an order of magnitude improvement in average BER. So, multiple users with different qualities of service can be supported by appropriately adjusting user signal powers. If we increase the bit energies for both users, while maintaining the same ratio in their powers, we can realize still further improvements in BER performance. As we see, antenna arrays offer another "resource", in addition to user signal power and system bandwidth, which can be used to overcome limitations in capacities, and improve the multi-media services of CDMA networks.

4. Antenna Weight Control Algorithm

In the previous section we evaluated the performance gains of a multi-rate CDMA network when adaptive arrays are implemented at the base station. These performance gains were based on optimal combining of antenna responses assuming optimal antenna weights, and thus represent an

upper bound on that achievable in practice. The degree to which we approach this upper bound is dependent on the performance of weight control algorithms. In this paper, we propose an LMS array approach, whereby we assume a given reference signal and use the well-known LMS and RLS algorithms to generate the antenna weights recursively.

A method for generating a suitable reference signal for spread spectrum signals was introduced by Compton [2, 1]. This method is based on the premise that the combination of despreading, filtering, and respreading, which occurs in the feedback loop shown in Figure 1, provides a reference signal consisting of the desired user's signal, and an interference component which is uncorrelated with the received signal at the antenna elements.

For this simulation, we assume the feedback loop operates ideally, and use a reference signal consisting of a normalized version of the desired user's signal plus an interference component. The interference component is modeled as a zero-mean complex gaussian random variable with variance σ_r, and uncorrelated with the received signal at the antenna elements. Also, for this simulation, we assume both user types have an activity factor of one. In future work, we will extend this to the case of user types, such as voice, with activity factors less than one.

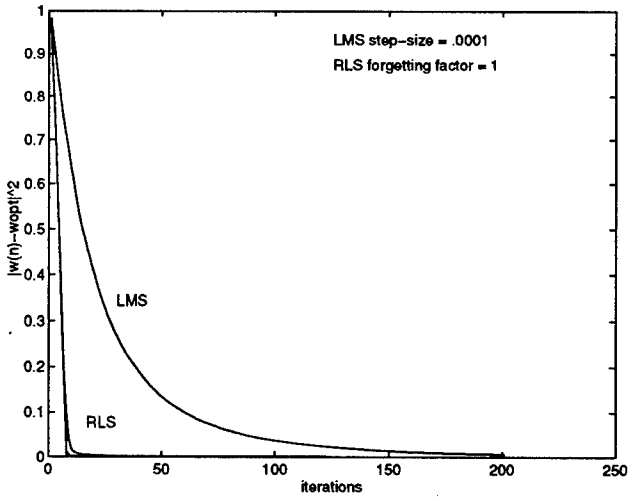


Figure 4. Convergence of RLS and LMS algorithms for user type 2

Figures 4 shows the results of antenna weight convergence for the LMS and RLS algorithms for user type 2 given $K_1 = 60$ and $K_2 = 20$. Similar results we observed for user type 1. The performance metric plotted is the norm-squared of the difference between the algorithm-computed antenna weights and the optimal antenna weights. Results shown are averaged over 100 independent trials. From the results, we observe that both algorithms converge to the optimal antenna weights. RLS converges within 10 data bits, whereas LMS requires on the order of 100 data bits. In addition, the results shown are for a reference signal with $\sigma_r = 0$ (ideal case), .1, and 1. Note, however, this had no effect on either algorithms performance. All cases are indistinguishable from the plots. This is expected, so long as the interference component of the reference signal is uncorrelated with the received signal at the antenna elements.

5. Conclusion

In this paper, we have shown that implementation of adaptive arrays are a promising technique for enhancing the multi-media services of a CDMA network. These enhancements include increased system capacities, as well as, improved robustness in handling multiple users with different quality-of-service requirements. Future work will focus on: (1) extending the BER analysis and weight control simulations to include multipath fading (IS-95), (2) further examination of the reference signal generation for a multi-rate CDMA environment, and (3) extending the weight control simulations to include users with non-unity activity factors.

References

[1] R. Compton. An adaptive array in a spread spectrum communication system. *Proceedings of IEEE*, 66, March 1978.

[2] R. Compton. *Adaptive Antennas Concepts and Performance*. Prentice Hall, 1980.

[3] J. C. Liberti and T. S. Rappaport. Analytical results for capacity improvements in CDMA. *IEEE Trans. Veh. Tech.*, VT-43(3), August 1994.

[4] A. Naguib and A. Paulraj. Recursive adaptive beamforming for wireless CDMA. *Proceedings ICC'95 (Seattle, Washington)*, June 1995.

[5] A. F. Naguib, A. Paulraj, and T. Kailath. Capacity improvements with base station antenna arrays in cellular CDMA. *IEEE Trans. Veh. Tech.*, VT-43(3), August 1994.

[6] T. Ottosson and A. Svensson. *Multi-Rate Performance in DS/CDMA Systems*. Technical Report no. 14, Department of Information Theory, Chalmus University of Technology, Goteborg, March 1995.

[7] B. Saurd, A. F. Naguib, G. Xu, and A. Paulraj. Performance of CDMA mobile communication systems using antenna arrays. *IEEE ICASSP'93*, Vol. VI, April 1993.

[8] G. V. Tsoulos, M. A. Beach, and S. C. Swales. Adaptive antennas for third generation DS-CDMA cellular systems.

WA-2: Array Processing II

JOINT ESTIMATION STRATEGY WITH APPLICATION TO EIGENSTRUCTURE METHODS

Alex B. Gershman *Johann F. Böhme*

Department of Electrical Engineering, Ruhr University
D-44780 Bochum, Germany

ABSTRACT

Numerous authors have attempted to improve the performance of eigenstructure methods, but all these approaches do not employ the additive information arising when several direction of arrival (DOA) estimation algorithms (referred to as *underlying estimators*) are used simultaneously. In this paper, we show that involving this information, one can achieve much better DOA estimation performance than that of each underlying estimator used separately. We introduce a *Joint Estimation Strategy* (JES) which represents a simple and effective way of extracting and combining such information. This strategy is then applied to the set of eigenstructure underlying DOA estimators including the MUSIC and Generalized Min-Norm (GMN) estimators.

1. INTRODUCTION

Eigenstructure techniques have proven to be an excellent tool for estimating DOA's of multiple narrowband sources in passive sensor arrays [1], [2]. At high signal to noise ratios (SNR's) and with a large number of snapshots, the eigenstructure techniques provide excellent estimation performance because the error variance is comparable to the Cramer-Rao bound (CRB). However, their performances become severely degraded at low SNR and when the number of data snapshots is small. The problem of improvement the performances of eigenstructure techniques has recently attracted significant attention. However, all known approaches to this problem do not employ additive information, arising when several underlying DOA estimation methods are used simultaneously. In this paper, we demonstrate that taking this information into account, one can achieve much better DOA estimation performance as compared to the conventional case, when each underlying estimator is used separately. We propose a Joint Estimation Strategy (JES) which represents a simple and effective way of extracting and combining such information. This strategy is then applied to the set of eigenstructure underlying DOA estimators, namely, to the MUSIC estimator [1] and the family of GMN techniques [3], [4].

This work was supported by the Alexander von Humboldt Foundation and the SASPARC Project of INTAS.

2. JOINT ESTIMATION STRATEGY

The central idea of JES is *resampling*. Several statistical resampling schemes are available, including the bootstrap scheme based on the resampling of initial data. In contrary to the bootstrap, our strategy can be interpreted as a resampling of *spatial spectrum*.

In order to explain the key idea of JES, let us consider a trial including the estimation of the covariance matrix using M data snapshots. The instantaneous performance of any estimation algorithm achieved in this single trial (without any statistical averaging) is hereafter referred to as a *local behavior* of this algorithm. Due to the fact that the underlying estimators are functionals of the sample covariance matrix (i.e., are different random functions of steering angle), their local behavior is different in each estimation trial. For example, considering two underlying estimators with comparable performance and increasing number of trials tested, one can always find some trials where the first estimator resolves the sources while the second estimator does not. In turn, some other trials always exist, which demonstrate the reverse local behavior where the first estimator does not resolve the sources while the second estimator does. This illustrates the intuitively clear fact that the probability of source resolution at least by one estimator among the simultaneously used estimators is always higher than that by each underlying estimator exploited separately. Evidently, for the extraction of the useful information arising when several estimators are used simultaneously, one should test the local behavior of each DOA estimator among the full set of underlying estimators and then to sort these estimators into two groups: the group of "successful" estimators resolving the sources in this concrete trial

0-8186-7576-4/96 $5.00

and the group of "unsuccessful" estimators which cannot resolve them. Then, only "successful" estimators should be used in such a way that allows to combine the results from each of them.

Let us consider the set of k underlying estimators with spectral functions $f_i(\theta)$, $i = 1, 2, \ldots, k$ of angle θ, which are calculated in parallel using the same data snapshots. Assume also that preliminary approximate estimates $\hat{q}$ and $\hat{\Theta}$ of the number of sources q and the angular sectors of source localization Θ are available. Then, testing the following hypothesis for each underlying estimator leads to the appropriate grouping of the "successful" DOA estimators:

$\mathcal{H}$: *The function* $f_i(\theta)$ *has more than* $\hat{q} - 1$ *separate spectral peaks localized in* $\hat{\Theta}$. □

Now, we are ready to formulate JES. It includes the following steps:

- Estimate the number of sources q using one of the existing signal detection techniques [5].

- Estimate the angular sectors of source localization Θ. One of the possible ways of estimating Θ using conventional beamformer is [3]:

$$\hat{\Theta} = [\vartheta_1^{\max} - \alpha\vartheta_1^{\text{left}}, \vartheta_1^{\max} + \alpha\vartheta_1^{\text{right}}]$$
$$\cup [\vartheta_2^{\max} - \alpha\vartheta_2^{\text{left}}, \vartheta_2^{\max} + \alpha\vartheta_2^{\text{right}}] \cup \cdots$$
$$\cdots \cup [\vartheta_p^{\max} - \alpha\vartheta_p^{\text{left}}, \vartheta_p^{\max} + \alpha\vartheta_p^{\text{right}}] \quad (1)$$

  where $\vartheta_l^{\max}$, $l = 1, 2, \ldots, p$ are the coordinates of the significant peaks of the conventional beamformer output, p is the total number of significant peaks, $\vartheta_l^{\text{right}}$ and $\vartheta_l^{\text{left}}$ are the right and left boundaries of each subinterval of estimated angular sectors Θ, and α is a positive coefficient close to 1. If the l-th peak has both right and left -3 dB decrease levels then $\vartheta_l^{\text{right}}$ and $\vartheta_l^{\text{left}}$ can be chosen as angular distances between the maximum of the l-th peak and the point of its -3 dB decrease, respectively. If the l-th peak has no right or left -3 dB decrease levels then $\vartheta_l^{\text{right}}$ and $\vartheta_l^{\text{left}}$ can be chosen as the angular distances between the maximum of the l-th peak and the corresponding right or left closest point in which the beamformer output transforms from the decreasing to increasing function.

- Test the hypothesis $\mathcal{H}$ for each DOA estimator from the total number of underlying estimators.

- If the hypothesis $\mathcal{H}$ is valid for $m > 0$ estimators $\tilde{f}_i(\theta)$, $i = 1, 2, \ldots, m$ from the total number of k estimators $f_i(\theta)$, $i = 1, 2, \ldots, k$, then estimate the l-th DOA θ_l as:

$$\hat{\theta}_l = \frac{1}{m} \sum_{i=1}^{m} \tilde{\theta}_l^{(i)} \quad (2)$$

  where $\tilde{\theta}_1^{(i)} < \tilde{\theta}_2^{(i)} < \cdots < \tilde{\theta}_{\hat{q}}^{(i)}$ is the ordered set of angles, corresponding to the $\hat{q}$ main maxima of the function $\tilde{f}_i(\theta)$.

  If the hypothesis $\mathcal{H}$ is wrong for all DOA estimators from the total number of underlying estimators, then estimate the l-th DOA θ_l as:

$$\hat{\theta}_l = \frac{1}{k} \sum_{i=1}^{k} \theta_l^{(i)} \quad (3)$$

  where $\theta_1^{(i)} < \theta_2^{(i)} < \cdots < \theta_{\hat{q}}^{(i)}$ is the ordered set of angles, corresponding to the $\hat{q}$ main maxima of the function $f_i(\theta)$. □

Equation (2) corresponds to the so-called *censored* averaging, using only "successful" estimators. In turn, (3) corresponds to the case when all estimators are "unsuccessful" and we have no reasons to prefer one estimator to another.

The presented strategy is a universal approach because it can be applied to any possible set of underlying estimators having angular spectral functions. One can involve into JES the root versions of estimators, too, using an appropriate transformation to the smooth spectral-type function [3]. This strategy allows to handle coherent source scenarios because it enables any preprocessing within each underlying estimator: for example, the well known spatial smoothing technique can be used. Moreover, one can choose the underlying estimators performing well in coherent source environments.

3. APPLICATION OF JES TO EIGENSTRUCTURE DOA ESTIMATORS

This section describes the underlying eigenstructure DOA estimators which can be successfully exploited in JES.

The i-th snapshot of the $n \times 1$ complex vector of n-element array outputs is given by

$$\boldsymbol{r}(i) = \boldsymbol{A}\boldsymbol{s}(i) + \boldsymbol{n}(i) \quad (4)$$

where $\boldsymbol{A} = [\boldsymbol{a}(\theta_1), \ldots, \boldsymbol{a}(\theta_q)]$ is the $n \times q$ matrix of the source wavefront vectors, $\boldsymbol{a}(\theta)$ is the $n \times 1$ wavefront vector corresponding to the direction θ, $\boldsymbol{s}(i)$ is the $q \times 1$ vector of random source waveforms, and $\boldsymbol{n}(i)$ is the

$n \times 1$ vector of sensor noise. The $n \times n$ spatial covariance matrix of array outputs can be expressed as [1], [2]:

$$\boldsymbol{R} = E[\boldsymbol{r}(i)\boldsymbol{r}^H(i)] = \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H + \sigma^2\boldsymbol{I} \qquad (5)$$

where $\boldsymbol{S}$ is the $q \times q$ covariance matrix of signal waveforms, $\boldsymbol{I}$ is the $n \times n$ identity matrix, $E[\cdot]$ and $\boldsymbol{H}$ denote the expectation operator and the Hermitian transpose, respectively. The sample covariance matrix is given by

$$\hat{\boldsymbol{R}} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{r}(i)\boldsymbol{r}^H(i) \qquad (6)$$

The eigendecomposition of the matrix (6) can be expressed as

$$\hat{\boldsymbol{R}} = \sum_{i=1}^{n} \hat{\lambda}_i \hat{\boldsymbol{u}}_i \hat{\boldsymbol{u}}_i^H \qquad (7)$$

where $\hat{\lambda}_i$ ($\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_n$) and $\hat{\boldsymbol{u}}_i$ are the i-th sample eigenvalue and i-th corresponding sample eigenvector, respectively. The popular MUSIC technique [1] estimates the DOA's as locations of $\hat{q}$ highest peaks of spectral function

$$f_{\text{MUSIC}}(\theta) = \left[\boldsymbol{a}^H(\theta)\hat{\boldsymbol{U}}_N\hat{\boldsymbol{U}}_N^H\boldsymbol{a}(\theta)\right]^{-1} \qquad (8)$$

where $\hat{\boldsymbol{U}}_N = [\hat{\boldsymbol{u}}_{\hat{q}+1}, \hat{\boldsymbol{u}}_{\hat{q}+2}, \ldots, \hat{\boldsymbol{u}}_n]$ is the $n \times (n-\hat{q})$ matrix constructed with the noise-subspace eigenvectors.

The GMN method [3], [4] represents the straightforward extension of the popular Kumaresan-Tufts MN technique [2] and estimates the DOA's as locations of $\hat{q}$ highest peaks of spectral function

$$f_{\text{GMN}}^{(i)}(\theta) = \left|\boldsymbol{a}^H(\theta)\boldsymbol{w}_i\right|^{-2} \qquad (9)$$

where the $n \times 1$ vector $\boldsymbol{w}_i$ is obtained by solving the following conditional minimization problem:

$$\min_{\boldsymbol{w}_i} \boldsymbol{w}_i^H\boldsymbol{w}_i, \qquad \hat{\boldsymbol{U}}_S^H\boldsymbol{w}_i = 0, \qquad \boldsymbol{w}_i^H\boldsymbol{e}_i = 1 \qquad (10)$$

Here $\hat{\boldsymbol{U}}_S = [\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2, \ldots, \hat{\boldsymbol{u}}_{\hat{q}}]$, and $\boldsymbol{e}_i$ is the vector with all zero elements except for the i-th one that is equal to 1. For $i = 1$ the GMN estimator coincides with the conventional MN estimator [2], i.e., $f_{\text{GMN}}^{(1)}(\theta) = f_{\text{MN}}(\theta)$. The solution of (10) can be written as [3]

$$f_{\text{GMN}}^{(i)}(\theta) = \left|\boldsymbol{a}^H(\theta)\hat{\boldsymbol{U}}_N\hat{\boldsymbol{U}}_N^H\boldsymbol{e}_i\right|^{-2} \qquad (11)$$

where the constant $[\boldsymbol{e}_i^H\hat{\boldsymbol{U}}_N\hat{\boldsymbol{U}}_N^H\boldsymbol{e}_i]^2$ is ignored in the numerator of (11). Eqn. (11) describes the family of GMN estimators, $f_{\text{GMN}}^{(i)}(\theta)$, $i = 1, 2, \ldots, n$.

Below, in simulations, we apply JES to the set of $n+1$ eigenstructure underlying DOA estimators, namely, to the MUSIC estimator (8) and n GMN estimators (11). The application of JES to these eigenstructure estimators only insignificantly increases the computational burden as compared with the MUSIC algorithm. For reduction the computational cost, the relationship between the MUSIC and GMN functions [3] as well as the fast algorithms [6], [7] can be employed.

4. SIMULATION RESULTS

In our simulations, we assume a uniform linear array of eight omnidirectional sensors with half-wavelength spacing, and two uncorrelated narrowband sources with equal power. A total of 100 statistically independent trials are used to obtain each simulated point of SNR. The number of snapshots taken in each trial is $M = 100$. Three scenarios with different source DOA separations are considered: 1). $\theta_1 = 0°, \theta_2 = 2°$, 2). $\theta_1 = 0°, \theta_2 = 4°$, and 3). $\theta_1 = 0°, \theta_2 = 8°$. Figures 1-3 show the experimental comparison of the JES-based, MUSIC and MN algorithms in terms of resolution probability for the scenarios 1-3, respectively. Figures 4-6 show the experimental DOA estimation RMS errors of these algorithms compared with CRB for the scenarios 1-3, respectively. It follows from simulations that the JES-based algorithm noticeably outperforms the MUSIC and MN (underlying) algorithms both in terms of resolution probability and RMS error.

REFERENCES

[1] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectral Estimation Workshop*, Rome, NY, pp. 234–258, 1979.

[2] R. Kumaresan, and D.W. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Trans. AES*, vol. 19, pp. 134–139, Jan. 1983.

[3] A.B. Gershman, and J.F. Böhme, "Joint estimation strategy: performance improvement by simultaneous use of several DOA estimators," submitted to *IEEE Trans. SP*, 1994.

[4] A.B. Gershman, "High-resolution direction finding using eigendecomposition based estimators and joint estimation strategy," *Electronics Letters*, vol. 27, 1991, pp. 2308–2309.

[5] M. Wax, and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. ASSP*, vol. 33, pp. 387–392, Apr. 1985.

[6] G. Xu and T. Kailath, "Fast subspace decomposition," *IEEE Trans. SP*, vol. 42, pp. 539–551, March 1994.

[7] V.T. Ermolaev, and A.B. Gershman "Fast algorithm for minimum-norm direction of arrival estimation," *IEEE Trans. SP*, vol. 42, pp. 2389–2394, Sept. 1994.

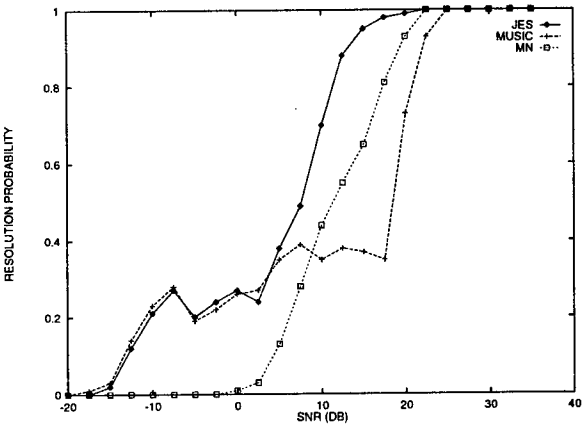


Figure 1: Experimental probabilities of source resolution versus SNR for the first scenario.

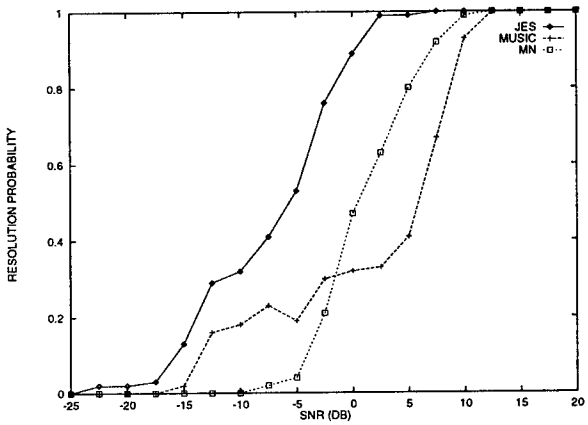


Figure 2: Experimental probabilities of source resolution versus SNR for the second scenario.

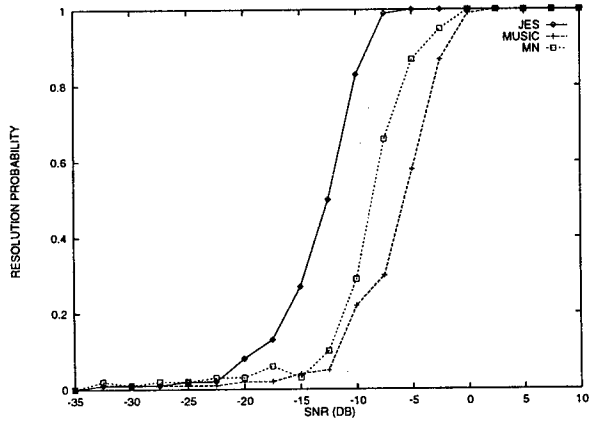


Figure 3: Experimental probabilities of source resolution versus SNR for the third scenario.

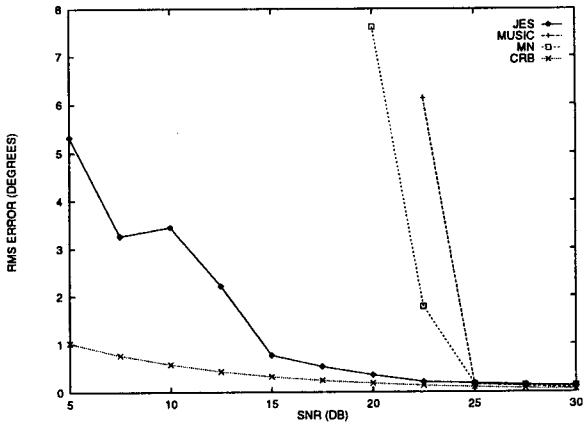


Figure 4: Experimental RMS error of DOA estimation versus SNR for the first scenario.

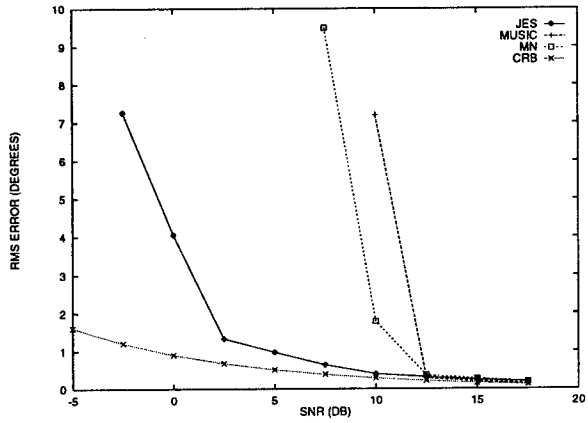


Figure 5: Experimental RMS error of DOA estimation versus SNR for the second scenario.

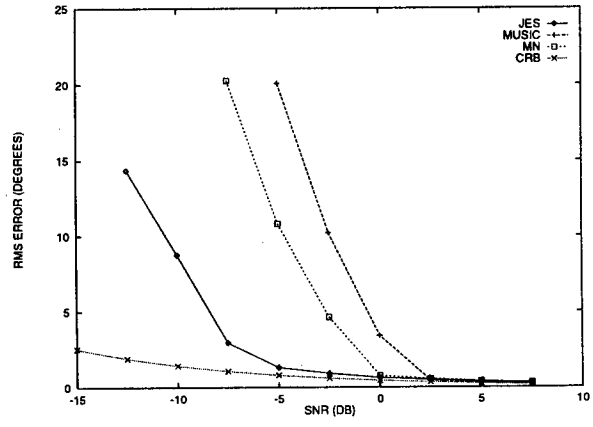


Figure 6: Experimental RMS error of DOA estimation versus SNR for the third scenario.

CRAMÉR-RAO BOUNDS FOR TARGET ANGLE AND DOPPLER ESTIMATION FOR AIRBORNE RADAR IN CAUCHY INTERFERENCE

R. Raspanti, P. Tsakalides, C. L. Nikias

Signal & Image Processing Institute
Dept. of Electrical Engineering – Systems
University of Southern California
Los Angeles, CA 90089–2564
e-mail: tsakalid@sipi.usc.edu

E. Del Re

Dept. of Electrical Engineering
University of Florence
Florence 50139
Italy
email: delre@ingfi1.ing.unifi.it

ABSTRACT

We describe new methods on the modeling of the amplitude statistics of airborne radar clutter by means of alpha-stable distributions. We develop target angle and Doppler, maximum likelihood-based estimation techniques from radar measurements retrieved in the presence of impulsive noise modeled as a multivariate isotropic alpha-stable random process. We derive the Cramér-Rao bounds for the additive Cauchy interference scenario to assess the best-case estimation accuracy which can be achieved. The results are of great importance in the study of space-time adaptive processing (STAP) for airborne pulse Doppler radar arrays operating in impulsive interference environments.

1. INTRODUCTION

Future advanced airborne radar systems must be able to detect, identify, and estimate the parameters of a target in severe interference backgrounds. As a result, the problem of clutter and jamming suppression has been the focus of considerable research in the radar engineering community. It is recognized that effective clutter suppression can be achieved only on the basis of appropriate statistical modeling. Recently, experimental results have been reported where clutter returns are impulsive in nature. In addition, a statistical model of impulsive interference has been proposed, which is based on the theory of symmetric alpha-stable ($S\alpha S$) random processes [1]. The model is of a statistical-physical nature and has been shown to arise under very general assumptions and to describe a broad class of impulsive interference.

Until recently much of the work reported for radar systems has concentrated mostly on target detection [2]. In this paper, we address the target parameter estimation problem through the use of radar array sensor data retrieved in the presence of impulsive interference. In particular, we derive Cramér-Rao bounds on angle and Doppler estimator accuracy for the case of additive sub-Gaussian noise. Initially, we consider the case of additive multivariate Cauchy noise, assuming knowledge of the *underlying matrix* of the distribution. The results obtained here can be viewed as generalizations of the work done in [3] to the 2-D frequency estimation problem in impulsive interference backgrounds. In Section 2, we present some necessary preliminaries on α-stable processes. In Section 3, we define the space-time adaptive processing (STAP) problem for airborne radar and we form the maximum likelihood function. In Section 4, we present the Cramér-Rao analysis and derive bounds on the variances of the spatial and temporal frequency estimates. Finally, in Section 5, we give some examples on the joint target angle and Doppler estimation performance.

2. SYMMETRIC ALPHA-STABLE DISTRIBUTIONS

In this section, we introduce the statistical model that will be used to describe the additive noise. The model is based on the class of *Complex Isotropic* $S\alpha S$ distributions which are well suited for describing signals that are impulsive in nature.

The symmetric α-stable ($S\alpha S$) distribution is best defined by its characteristic function

$$\varphi(\omega) = \exp(j\delta\omega - \gamma|\omega|^{\alpha}) \tag{1}$$

where α is the *characteristic exponent* restricted to the values $0 < \alpha \leq 2, \delta(-\infty < \delta < \infty)$ is the *location parameter*, and γ is the *dispersion* of the distribution. The dispersion plays a role analogous to the role that the variance plays for second-order processes. The characteristic exponent α is the most important parameter of the $S\alpha S$ distribution and it determines the shape of the distribution: the smaller the characteristic exponent α is, the heavier the tails of the $S\alpha S$ density.

$S\alpha S$ densities obey two important properties which further justify their role in data modeling: the *stability property* and the *generalized central limit theorem*. Unfortunately, no closed form expressions exist for the general $S\alpha S$ probability density functions (pdf) except for the Cauchy and the Gaussian case. However, power series expansions can be derived for the general pdf's [1]. Here, we are interested in the family of complex isotropic $S\alpha S$ random

The work of the second and third authors was supported by the Office of Naval Research under Contract N00014-92-J-1034.

0-8186-7576-4/96 $5.00

variables. A complex $S\alpha S$ random variable $X = X_1 + jX_2$ is *isotropic* if and only if the bivariate distribution (X_1, X_2) has uniform spectral measure. In this case, the characteristic function of X can be written as

$$\varphi(\omega) = \exp(j\Re[\omega X^*]) = \exp(-\gamma|\omega|^\alpha). \quad (2)$$

An important difference between the Gaussian and the other distributions of the $S\alpha S$ family is that only moments of order less than α exist for the non-Gaussian family members. If X follows the isotropic stable distribution with dispersion γ, the so called *fractional lower order moments* (FLOM) are given by

$$E|X|^p = C_2(p,\alpha)\gamma^{\frac{p}{\alpha}} \quad \text{for } 0 < p < \alpha, \quad (3)$$

where

$$C_2(p,\alpha) = \frac{2^{p+1}\Gamma(\frac{p+2}{2})\Gamma(-\frac{p}{\alpha})}{\alpha\Gamma(-\frac{p}{2})}. \quad (4)$$

3. STAP PROBLEM FORMULATION AND MAXIMUM LIKELIHOOD FUNCTION

Space-time adaptive processing (STAP) refers to multidimensional adaptive algorithms that simultaneously combine the signals from the elements of an array antenna and the multiple pulses of a coherent radar waveform, to suppress interference and provide target detection [4, 2, 5].

Consider a uniformly spaced linear array radar antenna consisting of N elements, which transmits a coherent burst of M pulses at a constant pulse repetition frequency (PRF) f_r and over a certain range of directions of interest. The pulses repetition interval is T_r. A space-time snapshot refers to the $MN \times 1$ vector of samples corresponding to a single range gate. Given a single snapshot containing target at angle ϕ and Doppler frequency f, the space-time snapshot can be written as [4]

$$\mathbf{x} = \beta\mathbf{v}(\phi, f) + \mathbf{n} \quad (5)$$

where β is the target's complex amplitude given by

$$\beta = x + jy. \quad (6)$$

The vector $\mathbf{v}$ is an $NM \times 1$ vector called the *space-time steering vector*. It may be expressed as

$$\mathbf{v}(\phi, f) = \mathbf{b}(f) \otimes \mathbf{a}(\phi) \quad (7)$$

where $\mathbf{a}(\phi)$ is the $N \times 1$ *spatial steering vector* containing the interelement phase shifts for a target at ϕ, and b(f) is the $M \times 1$ *temporal steering vector* that contains the interpulse phase shifts for a target with Doppler f. It is assumed that the functional form of $\mathbf{v}(\phi, f)$ is known. In addition, we can write

$$v_i(\phi, f) = b_{f(i)}(f) \cdot a_{g(i)}(\phi) \quad (8)$$

where $v_i(\phi, f)$ is the i-th element of the space-time steering vector $\mathbf{v}(\phi, f)$, $1 \le f(i) \le M$, and $1 \le g(i) \le N$.

The snapshot also contains a noise component $\mathbf{n}$. Here, the noise includes clutter, jamming, thermal noise, and any other undesired signals. As a first approximation to the problem, we assume that the noise present at the array is statistically independent both along the array sensors and along time, and is modeled as a complex isotropic Cauchy process with marginal pdf given by

$$\chi_\gamma(r) = \frac{\gamma}{2\pi(r^2 + \gamma^2)^{3/2}}, \quad (9)$$

Under the independence assumption it follows from (5) and (7) that the joint density function for the case of a single snapshot is given by [3]

$$f(\mathbf{n}) = \prod_{i=1}^{MN} f(n_i) = \frac{\gamma^{MN}}{(2\pi)^{MN}\prod_{i=1}^{MN}(\gamma^2 + |x_i - \beta v_i|^2)^{3/2}}. \quad (10)$$

In the following, it will be convenient to work with the normalized spatial and temporal frequency variables:

$$\psi = \frac{2\pi d}{\lambda_0}\sin\phi \quad , \quad \omega = 2\pi f T_r. \quad (11)$$

The estimation problem involves four real valued parameters. We arrange them to form a 4×1 parameter vector

$$\mathbf{\Theta} = [\theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4] = [\psi \quad \omega \quad x \quad y]. \quad (12)$$

Then, given a single snapshot $\mathbf{x}$, the likelihood function $L(\mathbf{\Theta})$, ignoring the constant terms, is given by

$$L(\mathbf{\Theta}) = -\frac{3}{2}\sum_{i=1}^{NM}\log(\gamma^2 + |x_i - \beta v_i(\psi,\omega)|^2). \quad (13)$$

4. CRAMÉR-RAO BOUND ANALYSIS

The Cramér-Rao bound for the error variance of an unbiased estimator $\hat{\mathbf{\Theta}}$ satisfies

$$\mathbf{C}_{\hat{\mathbf{\Theta}}} - \mathbf{J}(\mathbf{\Theta}) \ge 0 \quad (14)$$

where $\mathbf{C}_{\hat{\mathbf{\Theta}}}$ is the covariance matrix of $\hat{\mathbf{\Theta}}$ and ≥ 0 is interpreted as meaning that the matrix is semidefinite positive. The matrix $\mathbf{J}(\mathbf{\Theta})$ is the Fisher information matrix given by

$$\mathbf{J}(\mathbf{\Theta}) = E\{[\partial L(\mathbf{\Theta})/\partial\mathbf{\Theta}][\partial L(\mathbf{\Theta})/\partial\mathbf{\Theta}]^T\}. \quad (15)$$

First, we calculate the derivatives of the log-likelihood function given in (13) with respect to the components of $\mathbf{\Theta}$. We have that

$$\frac{\partial L}{\partial\psi} = 3\sum_{i=1}^{MN}\frac{\Re\{\beta^* b^*_{f(i)}(d^a_{g(i)})^* n_i\}}{\gamma^2 + |n_i|^2} \quad (16)$$

where $d^a_i = \partial a_i/\partial\psi, \quad i = 1, \cdots, N$. In addition

$$\frac{\partial L}{\partial\omega} = 3\sum_{i=1}^{MN}\frac{\Re\{\beta^* a^*_{g(i)}(d^b_{f(i)})^* n_i\}}{\gamma^2 + |n_i|^2} \quad (17)$$

where $d^b_i = \partial b_i/\partial\omega, \quad i = 1, \cdots, M$. Additionally,

$$\frac{\partial L}{\partial x} = 3\sum_{i=1}^{MN}\frac{\Re\{a^*_{g(i)} b^*_{f(i)} n_i\}}{\gamma^2 + |n_i|^2} \quad (18)$$

and

$$\frac{\partial L}{\partial y} = -3\sum_{i=1}^{MN} \frac{\Im\{a^*_{g(i)} b^*_{f(i)} n_i\}}{\gamma^2 + |n_i|^2}. \quad (19)$$

By performing the second derivatives and expectations in a similar way, the Fisher information matrix $\mathbf{J}(\mathbf{\Theta})$ is derived to be

$$\mathbf{J}(\mathbf{\Theta}) = \frac{3}{5\gamma^2} \cdot \begin{bmatrix} M|\beta|^2 \parallel \mathbf{d}_a \parallel^2 & |\beta|^2\rho & yM\delta_a & xM\delta_a \\ |\beta|^2\rho & N|\beta|^2 \parallel \mathbf{d}_b \parallel^2 & yN\delta_b & xN\delta_b \\ yM\delta_a & yN\delta_b & MN & 0 \\ xM\delta_a & xN\delta_b & 0 & MN \end{bmatrix},$$

where

$$\delta_a = \sum_{i=1}^{N} |d_i^a|, \quad \delta_b = \sum_{i=1}^{M} |d_i^b|, \quad \rho = \sum_{i=1}^{MN} |d^a_{g(i)}||d^b_{f(i)}|,$$

and $\mathbf{d}_a = [d_1^a \ \cdots \ d_n^a]$, $\mathbf{d}_b = [d_1^b \ \cdots \ d_n^b]$. Since target angle and Doppler are the two parameters of primary interest, we shall focus on the upper left 2×2 block of the Fisher information matrix $\mathbf{J}_{2\times 2}$. The inverse of matrix $\mathbf{J}_{2\times 2}$ is obtained by applying the partitioned matrix inversion lemma. The result is

$$\mathbf{J}^{-1}_{2\times 2}(\mathbf{\Theta}) = \frac{1}{\xi} \cdot \frac{5\gamma^2}{3|\beta|^2} \cdot \begin{bmatrix} N(\parallel \mathbf{d}_b \parallel^2 - \frac{1}{M}\delta_b^2) & \delta_a\delta_b - \rho \\ \delta_a\delta_b - \rho & M(\parallel \mathbf{d}_a \parallel^2 - \frac{1}{N}\delta_a^2) \end{bmatrix}, \quad (20)$$

where $\xi = (M \parallel \mathbf{d}_a \parallel^2 - \frac{M}{N}\delta_a^2)(N \parallel \mathbf{d}_b \parallel^2 - \frac{N}{M}\delta_b^2) - (\delta_a\delta_b - \rho)^2$. The Cramér-Rao bounds of the resulting spatial and temporal frequency estimates are obtained from (20) as

$$CRB(\psi) = \frac{\gamma^2}{|\beta|^2} \cdot \frac{5N(\parallel \mathbf{d}_b \parallel^2 - \delta_b^2/M)}{3\xi} \quad (21)$$

and

$$CRB(\omega) = \frac{\gamma^2}{|\beta|^2} \cdot \frac{5M(\parallel \mathbf{d}_a \parallel^2 - \delta_a^2/N)}{3\xi}. \quad (22)$$

Finally, by using (11), we get

$$CRB(\phi) = CRB(\psi) \cdot \frac{\lambda_0^2}{(2\pi d)^2} \frac{1}{\cos^2(\phi)} \quad (23)$$

and

$$CRB(f) = CRB(\omega) \cdot \frac{1}{(2\pi T_r)^2} \quad (24)$$

A useful insight on the CRB can be gained if we consider the case of linear array whose sensors are spaced a half-wavelength apart, and a waveform with an uniform pulse repetition interval. The spatial and temporal steering vectors for such system are:

$$\mathbf{a}(\psi) = \begin{bmatrix} 1 \\ e^{-j\psi} \\ \vdots \\ e^{-j(N-1)\psi} \end{bmatrix}, \quad \mathbf{b}(\omega) = \begin{bmatrix} 1 \\ e^{-j\omega} \\ \vdots \\ e^{-j(M-1)\omega} \end{bmatrix}. \quad (25)$$

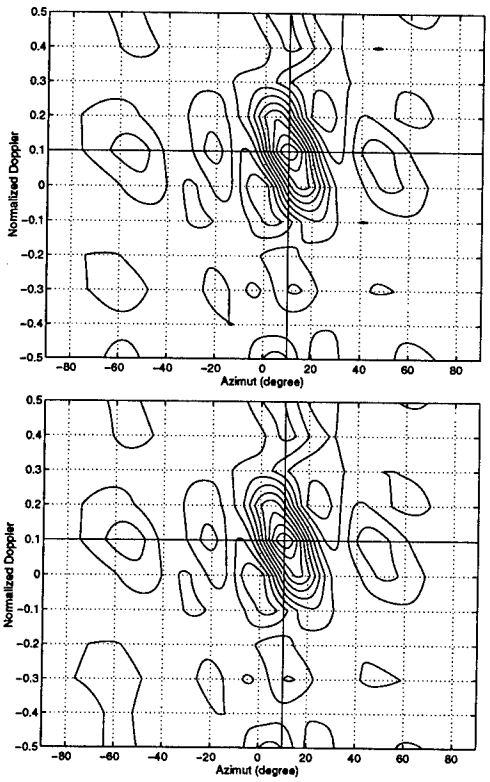


Figure 1: MLG (top) and MLC (bottom) angle-Doppler spectra ($N = 5$, $M = 4$, $\phi = -10°, fT_r = 0.1$). Additive Gaussian noise ($\alpha = 2$, $\gamma = 20, GSNR = 4$ dB).

In this case, it follows from (21) and (22) that

$$CRB(\psi) = \frac{\gamma^2}{|\beta|^2} \cdot \frac{20}{M^2N^2(N^2-1)} \quad (26)$$

and

$$CRB(\omega) = \frac{\gamma^2}{|\beta|^2} \cdot \frac{20}{M^2N^2(M^2-1)}. \quad (27)$$

5. SIMULATION RESULTS

In this simulation experiment, we test the robustness of the maximum likelihood estimator based on the Cauchy assumption (MLC). We assume a linear array with $N = 5$ elements that transmits a coherent burst of $M = 4$ pulses. We considered a single target located at $\phi = 10°$ and having Doppler such that $fT_r = \omega/2\pi = 0.1$. Since the alpha-stable family determines processes with infinite variance for $\alpha < 2$, we define an alternative signal-to-noise ratio (SNR). Namely, we define the *Generalized*-SNR (GSNR) to be the ratio of the signal power over the noise dispersion γ:

$$GSNR = 10\log\left(\frac{|\beta|^2}{\gamma}\right) \quad (28)$$

In Figures 1 and 2 we plot isosurfaces of space-time spectral estimates (likelihood functions) for the maximum likelihood estimator based on the Gaussian assumption (MLG)

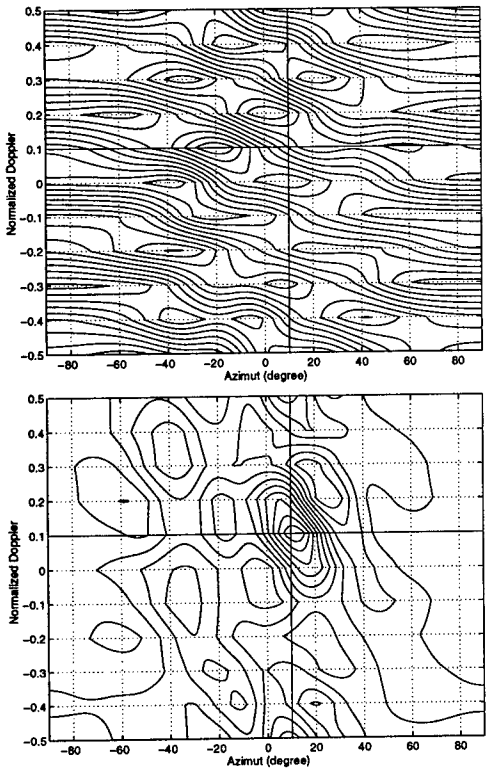


Figure 2: MLG (top) and MLC (bottom) angle-Doppler spectra ($N = 5$, $M = 4$, $\phi = -10°, fT_r = 0.1$). Additive stable noise ($\alpha = 1.5$, $\gamma = 20, GSNR = 4$ dB).

[2] and for the maximum likelihood estimator based on the Cauchy assumption (MLC).The likelihood functions are formed by using 50 space-time snapshots. In Figure 1, since the additive noise to the sensors is Gaussian ($\alpha = 2$), the MLG likelihood function is based on the correct assumption about the noise distribution. On the other hand, in Figure 2, the additive noise to the sensors is α-stable with $\alpha = 1.5$ and neither the MLG nor the MLC likelihood functions rely on the correct assumption about the noise distribution. As we can see from the figures, the MLC likelihood function, based on the Cauchy assumption, attains its maximum value very close to the true angle and Doppler values in both cases of additive stable noise. On the other hand, the MLG likelihood function, based on the Gaussian assumption, cannot localize the target accurately when the actual data distribution deviates from the Gaussian case.

The observed robustness of the MLC method is quantified in Figure 3 which shows the resulting mean-square error curve on the estimated Doppler as function of the characteristic exponent α of the additive noise. The results are based on 300 Monte Carlo runs. As we can clearly see, the Cauchy beamformer is practically insensitive to the changes of α. On the other hand, the MLG algorithm exhibits very large mean-square estimation error for non-Gaussian noise environments.

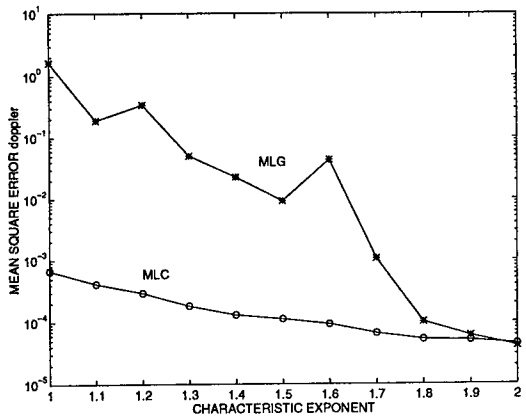


Figure 3: MSE of the estimated Doppler as a function of the characteristic exponent α.

6. CONCLUSIONS

We considered the problem of target angle and Doppler estimation with an airborne radar employing space-time adaptive processing. We derived Cramér-Rao bounds on angle and Doppler estimator accuracy for the case of additive multivariate Cauchy interference of known diagonal underlying matrix. The bounds are functions of a generalized SNR function, similarly to the Gaussian case where the bounds are functions of the SNR. As shown in (21) and (22), target angle accuracy is a function of Doppler frequency and vice-versa. In addition, we introduced a new joint spatial- and Doppler- frequency estimation technique based on the maximum likelihood Cauchy function (MLC) and we showed that the Cauchy estimator gives better results in a wide range of impulsive noise (clutter, jamming, thermal) environments.

7. REFERENCES

[1] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*. New York: John Wiley and Sons, 1995.

[2] J. Ward, "Cramér-Rao bounds for target angle and Doppler estimation with space-time adaptive processing radar," in *Twenty-Ninth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA), October 30-November 1 1995.

[3] P. Tsakalides and C. L. Nikias, "Maximum likelihood localization of sources in noise modeled as a stable process," *IEEE Trans. Signal Processing*, vol. 43, pp. 2700–2713, Nov. 1995.

[4] J. Ward, "Space-time adaptive processing for airborne radar," Tech. Rep. 1015, Lincoln Laboratory, Dec. 1994.

[5] S. U. Pillai, W. C. Lee, and J. Guerci, "Multichannel space-time adaptive processing," in *Twenty-Ninth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA), Oct. 30-Nov. 1 1995.

Higher-Order ESPRIT for Localization of Near-Field Sources: An Asymptotic Performance Analysis

Norman Yuen and Benjamin Friedlander
Department of Electrical & Computer Engineering
University of California, Davis
Davis, CA 95616
yuen@ece.ucdavis.edu
friedlan@ece.ucdavis.edu

Abstract

Most existing array processing techniques for estimating the directions of arrival or signal copy rely heavily on the plane-wave assumption of far-field sources. When the sources are located relatively close to the array, these techniques may no longer perform satisfactorily. In this paper, we present an asymptotic performance analysis of a recently proposed ESPRIT-like method for passive localization of near-field sources. The algorithm, based on fourth-order cumulants, is formulated for observations collected from a single uniformly spaced linear array. We examine the least-squares version of the algorithm and derive the expressions for the asymptotic variances of the estimated directions of arrival and estimated ranges of the sources.

1 Introduction

Most array processing methods which estimate the directions of arrival of sources make the assumption that the sources are located relatively far from the array, so that the waves emitted by the sources can be considered plane waves. However, when a source is located close to the array (*i.e.*, near-field), the plane wave assumption may no longer be valid and the wavefront must be characterized by both the azimuth and range. Methods based on the far-field assumption are not applicable to this situation. The near-field situation can occur, for example, in sonar, electronic surveillance, and seismic exploration.

In narrowband array processing, several variants of the ESPRIT algorithm using higher-order statistics have been presented. Recently, Challa and Shamsunder [1] developed a Total Least Squares ESPRIT-like algorithm, based on fourth-order cumulants, for estimating the azimuth and range of near-field sources impinging on a uniformly spaced linear array.

In this paper, we derive asymptotic expressions for the variances of estimates of the azimuth and range parameters using the higher-order ESPRIT-like algorithm of Challa and Shamsunder [1]. While Challa and Shamsunder formulated a total least squares algorithm, we give expressions based on a least squares version of the algorithm. However, it has been shown by Rao and Hari [2] that the asymptotic variances for these two versions are the same. Some of the expressions derived in this paper are based on the work presented in [4].

2 Problem Formulation

We use the narrowband model for array processing of near-field sources [1]. The output of the m^{th} sensor of the uniformly spaced linear array is given by

$$x_m(t) = \sum_{i=1}^{N} s_i(t) e^{j(\omega_i m + \phi_i m^2)} + n_m(t), \qquad (1)$$

for $m = -N_x+1, \ldots, 0, 1, \ldots, N_x$ (*i.e.*, there are $2N_x$ sensors). The array is shown in Figure 1. In matrix form, Equation (1) can be written as

$$\mathbf{x}(t) = \mathbf{B}\mathbf{s}(t) + \mathbf{n}(t) \qquad (2)$$

where the (m, n) element of $\mathbf{B}$ is given by

$$\mathbf{B}_{mn} = e^{j(-N_x+m)\omega_n + (-N_x+m)^2 \phi_n}. \qquad (3)$$

[1] This work was supported by the Office of Naval Research under contract No. N00014-95-1-0912.

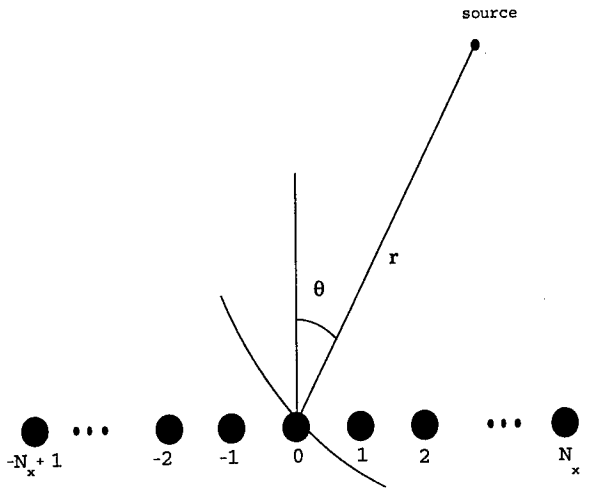


Figure 1. The uniformly spaced linear array con guration

The parameters ω_n and ϕ_n are functions of the azimuth θ_n and range r_n of the n^{th} source:

$$\omega_n = -2\pi\frac{\Delta}{L}\sin\theta_n \quad \text{and} \quad \phi_n = \pi\frac{\Delta^2}{Lr_n}\cos^2\theta_n \tag{4}$$

where L is the wavelength of the source wavefronts and Δ is the separation between adjacent sensors. The goal is to estimate the parameters $\{\theta_1, \ldots, \theta_d, r_1, \ldots, r_d\}$ given the array data $\mathbf{x}(t)$ for $0 \le t \le N_s$.

2.1 ESPRIT-like Algorithm

In this section, we summarize the higher-order ESPRIT-like algorithm proposed in [1]. Assuming that the source signals are zero-mean, non-Gaussian, statistically independent, and stationary, one can show that the matrix whose (m,n) element is

$$\begin{aligned} & C_1(m,n) \\ & \quad \triangleq \ \text{cum}\{x_m^*(t), x_{m+1}(t), x_{n+1}^*(t), x_n(t)\} \\ & \quad = \ \sum_{i=1}^{N} c_{4s_i} e^{j2\phi_i(m-n)} \end{aligned} \tag{5}$$

for $0 \le m, n \le N_x - 1$ is given by

$$\mathbf{C}_1 = \mathbf{A}\mathbf{C}_{4s}\mathbf{A}^H \tag{6}$$

which has dimensions $N_x \times N_x$. The kurtosis of the i^{th} source is c_{4s_i}. Similarly, using different sensor lags, one can compute the cumulant matrices $\mathbf{C}_2, \mathbf{C}_3$ and $\mathbf{C}_4$, as explained in [1]. Combining these matrices, one can form

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_4 & \mathbf{C}_2 \\ \mathbf{C}_4^H & \mathbf{C}_1 & \mathbf{C}_3 \\ \mathbf{C}_2^H & \mathbf{C}_3^H & \mathbf{C}_1 \end{bmatrix} = \bar{\mathbf{A}}\mathbf{C}_{4s}\bar{\mathbf{A}}^H \tag{7}$$

where

$$\bar{\mathbf{A}}^H = \begin{bmatrix} \mathbf{A}^H & \mathbf{\Phi}^H\mathbf{A}^H & \mathbf{\Omega}^H\mathbf{A}^H \end{bmatrix}. \tag{8}$$

The eigenvectors

$$\mathbf{E}_s = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_N \end{bmatrix} = \begin{bmatrix} \mathbf{E}_x \\ \mathbf{E}_y \\ \mathbf{E}_z \end{bmatrix} \tag{9}$$

corresponding to the N nonzero eigenvalues of $\mathbf{C}$ can be shown to yield

$$\mathbf{E}_y = \mathbf{E}_x\mathbf{\Psi} \quad \text{and} \quad \mathbf{E}_z = \mathbf{E}_x\mathbf{\Upsilon} \tag{10}$$

where

$$\mathbf{\Psi} = \mathbf{T}\mathbf{\Phi}\mathbf{T}^{-1} \quad \text{and} \quad \mathbf{\Upsilon} = \mathbf{T}\mathbf{\Omega}\mathbf{T}^{-1} \tag{11}$$

for some invertible matrix $\mathbf{T}$. Hence the eigenvalues of $\mathbf{\Psi}$ and $\mathbf{\Upsilon}$ allow one to compute the azimuth and range parameters. Furthermore, $\mathbf{\Psi}$ and $\mathbf{\Upsilon}$ can be computed, in a least squares sense, using

$$\mathbf{\Psi} = \mathbf{E}_x^{\#}\mathbf{E}_y \quad \text{and} \quad \mathbf{\Upsilon} = \mathbf{E}_x^{\#}\mathbf{E}_z, \tag{12}$$

respectively, where $\#$ denotes the pseudo-inverse of a matrix.

3 Asymptotic Performance Analysis

In this section, we derive the asymptotic variances of the estimated azimuth and range parameters for the higher-order ESPRIT-like algorithm of the previous section. The analysis is similar to that given in [4], where only the estimated azimuth parameter is analyzed.

Azimuth

For azimuth estimation, the quantity of interest is θ_i (assumed to be given in degrees), which is related to λ_i by

$$\lambda_i = e^{j4\pi\frac{\Delta}{L}\sin(\frac{\theta_i\pi}{180})}. \tag{13}$$

The separation between adjacent sensors is Δ and the wavelength of the impinging wavefronts is L. Using a first-order Taylor series expansion, we have

$$\delta\theta_i = \frac{\delta\lambda_i}{j4\pi\frac{\Delta}{L}(\cos(\frac{\theta_i\pi}{180}))\lambda_i}. \tag{14}$$

Skipping some algebraic steps, we get

$$\begin{aligned}&\mathrm{E}\{\delta\theta_i \cdot \delta\theta_i\}\\ &= \frac{1}{2}\left(\frac{L}{4\pi\Delta\cos(\frac{\theta_i\pi}{180})}\right)^2(\mathrm{E}\{|\delta\lambda_i|^2\} - \\ &\quad -\mathrm{real}\{\mathrm{E}\{(\delta\lambda_i)^2\}(\lambda_i^*)^2\}).\end{aligned} \tag{15}$$

Range

For range estimation, the quantity of interest is r_i, which is related to γ_i by

$$\gamma_i = e^{j2\pi\frac{\Delta^2}{Lr_i}\cos^2(\frac{\theta_i\pi}{180})}. \tag{16}$$

Using a first-order Taylor series expansion, we have

$$\delta r_i = \frac{\delta\gamma_i + \frac{\Delta}{2r_i}\sin(\frac{\theta_i\pi}{180})\gamma_i(\frac{\delta\lambda_i - \lambda_i^2\delta\lambda_i^*}{\lambda_i})}{-j2\pi\frac{\Delta^2}{Lr_i^2}\cos^2(\frac{\theta_i\pi}{180})\gamma_i}. \tag{17}$$

After some algebra, the variance of the estimated range parameter is

$$\begin{aligned}&\mathrm{E}\{\delta r_i \cdot \delta r_i\}\\ &= \frac{1}{2(2\pi\frac{\Delta^2}{Lr_i^2}\cos^2(\frac{\theta_i\pi}{180}))^2}\\ &\quad\cdot[\mathrm{E}\{|\delta\gamma_i|^2\} - \mathrm{real}\{(\gamma_i^*)^2\mathrm{E}\{(\delta\gamma_i)^2\}\\ &\quad+2(\frac{\Delta}{r_i}\sin(\frac{\theta_i\pi}{180}))^2\mathrm{E}\{|\delta\lambda_i|^2\} + \frac{\Delta}{r_i}\sin(\frac{\theta_i\pi}{180})\\ &\quad\cdot\mathrm{real}\{2\gamma_i^*\lambda_i\mathrm{E}\{\delta\lambda_i^*\delta\gamma_i\} - 2\gamma_i^*\lambda_i^*\mathrm{E}\{\delta\lambda_i\delta\gamma_i\}\}\\ &\quad-(\frac{\Delta}{2r_i}\sin(\frac{\theta_i\pi}{180}))^2\mathrm{real}\{3(\lambda_i^*)^2\mathrm{E}\{(\delta\lambda_i)^2\}\\ &\quad+\lambda_i^2\mathrm{E}\{(\delta\lambda_i^*)^2\} - 2\mathrm{E}\{|\delta\lambda_i|^2\}]\end{aligned} \tag{18}$$

where the quantities $\mathrm{E}\{|\delta\gamma_i|^2\}$, $\mathrm{E}\{\delta\lambda_i^*\delta\gamma_i\}$, $\mathrm{E}\{(\delta\gamma_i)^2\}$, and $\mathrm{E}\{\delta\lambda_i\delta\gamma_i\}$, which are the covariances of eigenvalues, are derived next.

3.1 Covariance of Eigenvalues

In LS-ESPRIT, we compute two matrices $\mathbf{\Psi}$ and $\mathbf{\Upsilon}$ using Equation (12) and then perform an eigendecomposition to get their associated eigenvalues. Let λ_i be an eigenvalue of $\mathbf{\Upsilon}$, $\mathbf{v}_i$ be the corresponding eigenvector, and $\mathbf{q}_i$ be the corresponding left eigenvector, such that

$$\begin{aligned}\mathbf{\Upsilon}\mathbf{v}_i &= \lambda_i\mathbf{v}_i\\ \mathbf{q}_i\mathbf{\Upsilon} &= \lambda_i\mathbf{q}_i.\end{aligned} \tag{19}$$

Furthermore, the left and right eigenvectors can be chosen to be orthonormal, so that

$$\lambda_i = \mathbf{q}_i\mathbf{\Upsilon}\mathbf{v}_i. \tag{20}$$

Under most circumstances, the matrix $\mathbf{\Upsilon}$ has to be estimated using finite data. An error $\delta\mathbf{\Upsilon} = \hat{\mathbf{\Upsilon}} - \mathbf{\Upsilon}$ in estimating $\mathbf{\Upsilon}$ will cause an error $\delta\lambda_i = \hat{\lambda}_i - \lambda_i$. As a first-order approximation, the error $\delta\lambda$ can be shown to be

$$\delta\lambda_i = \mathbf{q}_i\delta\mathbf{\Upsilon}\mathbf{v}_i. \tag{21}$$

It follows from Equation (10) that we have the first-order approximations

$$(\mathbf{E}_x + \delta\mathbf{E}_x)(\mathbf{\Upsilon} + \delta\mathbf{\Upsilon}) \approx \mathbf{E}_z + \delta\mathbf{E}_z, \tag{22}$$

$$\mathbf{E}_x\delta\mathbf{\Upsilon} \approx \delta\mathbf{E}_z - \delta\mathbf{E}_x\mathbf{\Upsilon}, \tag{23}$$

and

$$\delta\mathbf{\Upsilon} \approx \mathbf{E}_x^{\#}\delta\mathbf{E}_z - \mathbf{E}_x^{\#}\delta\mathbf{E}_x\mathbf{\Upsilon}. \tag{24}$$

Using (24) in (21) and noting that $|\lambda_i|^2 = 1$, we get

$$\begin{aligned}\delta\lambda_i &= \mathbf{q}_i\mathbf{E}_x^{\#}(\delta\mathbf{E}_z\mathbf{v}_i - \delta\mathbf{E}_x\mathbf{\Psi}\mathbf{v}_i)\\ &= -\lambda_i\mathbf{q}_i\mathbf{E}_x^{\#}(\mathbf{W}_1 - \lambda_i^*\mathbf{W}_3)\delta\mathbf{E}_s\mathbf{v}_i,\end{aligned} \tag{25}$$

where

$$\mathbf{W}_1 = \begin{bmatrix}\mathbf{I}_m & \mathbf{0}_m & \mathbf{0}_m\end{bmatrix}, \tag{26}$$

$$\mathbf{W}_3 = \begin{bmatrix}\mathbf{0}_m & \mathbf{0}_m & \mathbf{I}_m\end{bmatrix}. \tag{27}$$

and

$$\delta\mathbf{E}_s = \hat{\mathbf{E}}_s - \mathbf{E}_s. \tag{28}$$

The matrix $\mathbf{E}_s$ is defined in (9), $\mathbf{I}_m$ is the $m \times m$ identity matrix and $\mathbf{0}_m$ is the $m \times m$ zero matrix.

Similarly, let $\mathbf{p}_i$ and $\mathbf{b}_i$ be the left and right eigenvectors of the matrix $\mathbf{\Psi}$, respectively, and γ_i be the corresponding eigenvalue, so that

$$\gamma_i = \mathbf{p}_i\mathbf{\Psi}\mathbf{b}_i. \tag{29}$$

The error $\delta\gamma_i$ is then given by

$$\delta\gamma_i = \mathbf{p}_i\mathbf{E}_x^{\#}(\mathbf{W}_2 - \gamma_i\mathbf{W}_1)\delta\mathbf{E}_s\mathbf{b}_i, \tag{30}$$

where $\mathbf{W}_1$ is defined in (26) and $\mathbf{W}_2$ is

$$\mathbf{W}_2 = \begin{bmatrix}\mathbf{0}_m & \mathbf{I}_m & \mathbf{0}_m\end{bmatrix}. \tag{31}$$

The variance of $\hat{\lambda}_i$ is thus, to the first-order,

$$\begin{aligned}&\mathrm{E}\{|\delta\lambda_i|^2\}\\ &= \mathbf{q}_i\mathbf{E}_x^{\#}(\mathbf{W}_3 - \lambda_i\mathbf{W}_1)[\sum_{g=1}^{d}\sum_{h=1}^{d}v_{i,g}v_{i,h}^*\\ &\quad\mathrm{E}\{\delta\mathbf{s}_g\delta\mathbf{s}_h^H\}]\cdot(\mathbf{W}_3 - \lambda_i\mathbf{W}_1)^H(\mathbf{q}_i\mathbf{E}_x^{\#})^H,\end{aligned} \tag{32}$$

where $\delta\mathbf{s}_g$ is the g^{th} column of the matrix $\delta\mathbf{E}_s$ and $v_{i,g}$ is the g^{th} element of the vector $\mathbf{v}_i$. The quantities $\delta\mathbf{s}_h$ and $v_{i,h}$ are similarly defined.

Following the above steps in an analogous manner, we can also derive the following first-order expressions.

$$\mathrm{E}\{(\delta\lambda_i)^2\}$$
$$= \mathbf{q}_i\mathbf{E}_x^{\#}(\mathbf{W}_3 - \lambda_i\mathbf{W}_1)[\sum_{g=1}^{d}\sum_{h=1}^{d} v_{i,g}v_{i,h}$$
$$\mathrm{E}\{\delta\mathbf{s}_g\delta\mathbf{s}_h^T\}](\mathbf{W}_3 - \lambda_i\mathbf{W}_1)^T(\mathbf{q}_i\mathbf{E}_x^{\#})^T, \quad (33)$$

$$\mathrm{E}\{\delta\lambda_i\delta\gamma_i^*\}$$
$$= \mathbf{q}_i\mathbf{E}_x^{\#}(\mathbf{W}_3 - \lambda_i\mathbf{W}_1)[\sum_{g=1}^{d}\sum_{h=1}^{d} v_{i,g}b_{i,h}^*$$
$$\mathrm{E}\{\delta\mathbf{s}_g\delta\mathbf{s}_h^H\}](\mathbf{W}_2 - \gamma_i\mathbf{W}_1)^H(\mathbf{p}_i\mathbf{E}_x^{\#})^H, \quad (34)$$

$$\mathrm{E}\{|\delta\gamma_i|^2\}$$
$$= \mathbf{p}_i\mathbf{E}_x^{\#}(\mathbf{W}_2 - \gamma_i\mathbf{W}_1)[\sum_{g=1}^{d}\sum_{h=1}^{d} b_{i,g}b_{i,h}^*$$
$$\mathrm{E}\{\delta\mathbf{s}_g\delta\mathbf{s}_h^H\}](\mathbf{W}_2 - \gamma_i\mathbf{W}_1)^H(\mathbf{p}_i\mathbf{E}_x^{\#})^H, \quad (35)$$

$$\mathrm{E}\{(\delta\gamma_i)^2\}$$
$$= \mathbf{p}_i\mathbf{E}_x^{\#}(\mathbf{W}_2 - \gamma_i\mathbf{W}_1)[\sum_{g=1}^{d}\sum_{h=1}^{d} b_{i,g}b_{i,h}$$
$$\mathrm{E}\{\delta\mathbf{s}_g\delta\mathbf{s}_h^T\}](\mathbf{W}_2 - \gamma_i\mathbf{W}_1)^T(\mathbf{p}_i\mathbf{E}_x^{\#})^T, \quad (36)$$

$$\mathrm{E}\{\delta\lambda_i\delta\gamma_i\}$$
$$= \mathbf{q}_i\mathbf{E}_x^{\#}(\mathbf{W}_3 - \lambda_i\mathbf{W}_1)[\sum_{g=1}^{d}\sum_{h=1}^{d} v_{i,g}b_{i,h}$$
$$\mathrm{E}\{\delta\mathbf{s}_g\delta\mathbf{s}_h^T\}](\mathbf{W}_2 - \gamma_i\mathbf{W}_1)^T(\mathbf{p}_i\mathbf{E}_x^{\#})^T. \quad (37)$$

These equations depend on the covariances of the eigenvectors, $\mathrm{E}\{\delta\mathbf{s}_g\delta\mathbf{s}_h^H\}$ and $\mathrm{E}\{\delta\mathbf{s}_g\delta\mathbf{s}_h^T\}$, of the sample cumulant matrix $\hat{\mathbf{C}}$. In the following sections we derive these covariances.

3.2 Covariance of Eigenvectors

By using the first-order Taylor series expansion of eigenvectors of a matrix [3], we can show that the covariance matrix of the signal eigenvectors $\hat{\mathbf{s}}_g$ and $\hat{\mathbf{s}}_h$ is, to the first-order,

$$\mathrm{E}\{\delta\mathbf{s}_g\delta\mathbf{s}_h^H\}$$
$$= \sum_{\substack{l\neq g\\ l=1}}^{3N_x}\sum_{\substack{n\neq h\\ n=1}}^{3N_x}\sum_{a_1=1}^{3N_x}\sum_{a_2=1}^{3N_x}\sum_{b_1=1}^{3N_x}\sum_{b_2=1}^{3N_x} s_{l,a_1}^* s_{g,a_2}$$
$$s_{n,b_1}s_{h,b_2}^* \cdot \mathrm{E}\{(\hat{\mathbf{C}} - \mathbf{C})_{a_1a_2}(\hat{\mathbf{C}} - \mathbf{C})_{b_1b_2}^*\}$$
$$\frac{\mathbf{s}_l\mathbf{s}_n^H}{(\alpha_g - \alpha_l)(\alpha_h - \alpha_n)} \quad (38)$$

where α_i is an eigenvalue of the matrix $\mathbf{C}$ and s_{g,a_i} is the a_i^{th} element of $\mathbf{s}_g$, which is an eigenvector of $\mathbf{C}$. The notation $(\cdot)_{mn}$ refers to the (m,n) element of a matrix. We note that this expression is greatly simplified when the signals are Gaussian, as is assumed in [2].

Furthermore, the unconjugated covariance of the sample eigenvectors $\hat{\mathbf{s}}_g$ and $\hat{\mathbf{s}}_h$ is, to the first-order,

$$\mathrm{E}\{\delta\mathbf{s}_g\delta\mathbf{s}_h^T\}$$
$$= \sum_{\substack{l\neq g\\ l=1}}^{3N_x}\sum_{\substack{n\neq h\\ n=1}}^{3N_x}\sum_{a_1=1}^{3N_x}\sum_{a_2=1}^{3N_x}\sum_{b_1=1}^{3N_x}\sum_{b_2=1}^{3N_x} s_{l,a_1}^* s_{g,a_2}$$
$$s_{n,b_1}^* s_{h,b_2} \cdot \mathrm{E}\{(\hat{\mathbf{C}} - \mathbf{C})_{a_1a_2}(\hat{\mathbf{C}} - \mathbf{C})_{b_1b_2}\}$$
$$\frac{\mathbf{s}_l\mathbf{s}_n^T}{(\alpha_g - \alpha_l)(\alpha_h - \alpha_n)}. \quad (39)$$

The asymptotic covariance of sample fourth-order cumulants, denoted by $\mathrm{E}\{(\hat{\mathbf{C}} - \mathbf{C})_{a_1a_2}(\hat{\mathbf{C}} - \mathbf{C})_{b_1b_2}^*\}$ and by $\mathrm{E}\{(\hat{\mathbf{C}} - \mathbf{C})_{a_1a_2}(\hat{\mathbf{C}} - \mathbf{C})_{b_1b_2}\}$ are derived in [5].

4 Conclusion

In this paper we derived expressions for the asymptotic variances of azimuth and range estimates for the higher-order ESPRIT-like algorithm proposed by Challa and Shamsunder. The formulas derived can be used to evaluate the performance of the algorithm.

References

[1] R. N. Challa and S. Shamsunder, "Higher-order subspace based algorithms for passive localization of near-field sources," *Twenty-Ninth Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 1995.

[2] B. D. Rao and K. V. S. Hari, "Performance analysis of ESPRIT and TAM in determining the direction of arrival of plane waves in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1990-1995, December 1989.

[3] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.

[4] N. Yuen and B. Friedlander, "Asymptotic performance analysis of three ESPRIT algorithms," *Twenty-Ninth Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 1995.

[5] N. Yuen and B. Friedlander, "Asymptotic performance analysis of ESPRIT, higher-order ESPRIT, and virtual ESPRIT algorithms," *IEEE Transactions on Signal Processing*, submitted for publication, 1995.

Reduced-Rank Linear Regression*

P. STOICA

Systems and Control Group
Uppsala University
Box 27, S-751 03 Uppsala, Sweden
Email: ps@syscon.uu.se

M. VIBERG

Department of Applied Electronics
Chalmers University of Technology
S-412 96 Gothenburg, Sweden
Email: viberg@ae.chalmers.se

Abstract

This paper considers the problem of maximum likelihood (ML) estimation for reduced-rank linear regression equations with noise of arbitrary covariance. An explicit expression for the ML estimate of the regression matrix is derived. A generalized likelihood ratio (GLRT) test is also proposed, for estimating the rank of the regression matrix. Computer simulations and numerical examples indicate the superiority of the proposed estimator, as compared to a traditional least-squares approch that does not exploit the reduced rank property in an optimal way.

1 Introduction and Preliminaries

The focus of the present paper is on multivariate linear regression models of the following form:

$$y(t) = \phi x(t) + e(t), \quad t = 1, 2, ... \tag{1}$$

where $y(t) \in \mathbf{R}^{m\times 1}$ denotes the noise-obscured output (or explained) variables; $x(t) \in \in \mathbf{R}^{p\times 1}$ is the vector of input (or explanatory) variables; $e(t) \in \mathbf{R}^{m\times 1}$ denotes the equation noise; and $\phi \in \mathbf{R}^{m\times p}$ is the matrix of regression coefficients, or the parameter matrix for short. The following assumptions on (1) are considered to hold throughout the paper:

A1 The noise is temporally white, i.e.

$$E[e(t)e^T(s)] = 0 \ for \ t \neq s \tag{2}$$

and normally distributed with zero mean and *unknown* covariance matrix,

$$Q \triangleq E[e(t)e^T(t)]; \quad |Q| \neq 0 \tag{3}$$

(Hereafter, E stands for statistical expectation, and $|\cdot|$ denotes the determinant function).

A2 The explanatory variables $x(t)$ are deterministic signals, which are such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} x(t)x^T(t) = R_{xx}; \quad |R_{xx}| \neq 0 \tag{4}$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} x(t)e^T(t) = 0 \tag{5}$$

(the second equality above holds with probability one).

A3 The regression matrix ϕ may be rank deficient,

$$rank(\phi) = n; \quad n \leq \tilde{n} \triangleq min(p, m) \tag{6}$$

but n is *unknown* (ϕ itself is also *unknown*, of course).

The equation (1), along with the previous assumptions, define a *rank-reduced multivariate linear regression model with quasi-stationary deterministic*

*This work was supported in part by the Swedish Research Council for Engineering Sciences (TFR).

0-8186-7576-4/96 $5.00 © 1996 IEEE

inputs and random white normal noise of arbitrary covariance. The practical significance of reduced-rank regression modelling is discussed, for example, in [1, 3]. For instance, in large econometric models, several (noise-free) equations may be linearly related to one another, which renders ϕ rank deficient. An essentially equivalent situation appears when only a low-dimensional linear transformation of the explanatory variables suffices to describe the model outputs. Rank-reduced regression methods for certain signal processing problems are also discussed in e.g. [4, 7]. However, the studied problems in the latter contributions are somewhat different (the regressor ϕ is not explicitly modeled as rank-deficient) and the proposed methods are more or less ad-hoc from a statistical point of view. Another application of the reduced-rank regression occurs in state-space modelling of linear dynamic systems [2]. More exactly, it was shown in [2] that the estimation of the observability matrix (and then of the state-space equation parameters) associated with a linear system, by using subspace-based methods, is basically a reduced-rank linear regression problem as defined herein. The latter problem is also closely related to canonical correlation and factor analysis, and as such it is relevant to array signal processing applications. In fact, the estimation of the rank of a cross-covariance matrix from its sample version can be formulated as a reduced-rank linear regression problem. Note that the former estimation problem occurs in several signal processing and time series applications, including number of sources detection in sensor array signal processing (see e.g. [8]).

The distinctive feature of the above model is the reduced rank of ϕ. If $n = rank(\phi)$ were equal to $\tilde{n}$, then the equation (1) would be a standard linear regression, the parameter estimation of which is well documented in the literature (see, e.g., [1, 3, 5]). When $n < \tilde{n}$ (as stated in A3), the estimation of the parameters in (1) is a more complicated problem which has not received enough attention in the literature.

2 Main Results

Let the available observations be

$$\{y(1),\ x(1),\ldots,\ y(N),\ x(N)\}\,,\quad N > m+p\,.$$

Under assumption A1, the negative log-likelihood function of the observed data is given by (to within a constant)

$$L \sim \frac{N}{2}\left(\ln|Q| + \mathrm{tr}\left\{Q^{-1}\frac{1}{N}\sum_{t=1}^{N}[y(t)-\phi x(t)] \times [y(t)-\phi x(t)]^T\right\}\right)\,,$$

where $\mathrm{tr}(\cdot)$ is the trace operator. In view of A3, the ML estimates of ϕ and Q are obtained by solving the following problem,

$$\min_{Q;\phi} L(Q,\phi)$$

under the constraint $\mathrm{rank}(\phi) = n$. The constrained optimization problem above can be transformed into an unconstrained one by parameterizing ϕ as

$$\phi = AB^T \tag{7}$$

where both $A \in \mathbf{R}^{m\times n}$ and $B \in \mathbf{R}^{p\times n}$ are full rank matrices,

$$rank(A) = rank(B) = n$$

The factorization in (7), of course, is not unique. This fact complicates the analysis that follows to a certain degree, but the difficulties induced by the non-uniqueness of the parameterization of ϕ can be overcome.

Introduce the sample covariance matrix

$$\hat{R}_{xx} = \frac{1}{N}\sum_{t=1}^{N} x(t)x^T(t) \tag{8}$$

and similarly for the sample covariances $\hat{R}_{yx}$ and $\hat{R}_{yy}$. Assuming n to be known, the exact ML estimate of ϕ, obtained by explicitly minimizing $L(Q,\phi)$, is given by

$$\hat{\phi}_{ML} = \hat{R}_{yx}\hat{R}_{xx}^{-1/2}\hat{S}\hat{S}^T\hat{R}_{xx}^{-1/2}\,, \tag{9}$$

where the columns of the $p \times n$-matrix $\hat{S}$ are the n principal eigenvectors of the matrix $\hat{W}$, given as

$$\hat{W} = \hat{R}_{xx}^{-1/2} \hat{R}_{yx}^T \hat{R}_{yy}^{-1} \hat{R}_{yx} \hat{R}_{xx}^{-1/2} \ . \qquad (10)$$

The noise covariance estimate is obtained by inserting $\hat{\phi}$ into the expression

$$\hat{Q} = \frac{1}{N} \sum_{t=1}^{N} \left[y(t) - \hat{\phi} x(t) \right] \left[y(t) - \hat{\phi} x(t) \right]^T$$

For a proof see [6], where also the asymptotic properties of the ML estimate are derived. Note that the eigenvalues of $\hat{W}$ are the so-called *canonical correlations*, and the linearly transformed explanatory variables $\hat{S}^T x(t)$ are the *canonical variates* [1, 3].

In the more interesting case where n is not known, a generalized likelihood-ratio test (GLRT) can be performed. Let $\hat{n}$ be a candidate rank of ϕ to be evaluated. The proposed procedure is formulated as testing

$$H_0 : \hat{n} = n$$

against the opposing hypothesis that $n = \tilde{n}$, where $\tilde{n} = \min\{m, p\}$ is the maximum possible rank of ϕ. The GLRT statistic for this test is given by

$$\zeta_{\hat{n}} = -\frac{N}{2} \sum_{k=\hat{n}+1}^{\tilde{n}} \ln\left(1 - \hat{\lambda}_k\right) \ ,$$

where $\hat{\lambda}_k$ denote the eigenvalues of $\hat{W}$ in non-increasing order. Under the null hypothesis H_0, the GLRT variable $2\,\zeta_{\hat{n}}$ is shown to have an asymptotic chi-squared distribution with $(m - \hat{n})(p - \hat{n})$ degrees of freedom,

$$2\zeta_{\hat{n}} \xrightarrow{distr.} \mathcal{X}^2 \left[(m - \hat{n})(p - \hat{n}) \right] \ .$$

The proposed procedure is now to test H_0 for increasing values of $\hat{n}$ (starting at $\hat{n} = 1$ or any *a priori* known lower bound) until the hypothesis is accepted. For each $\hat{n}$, $\zeta_{\hat{n}}$ is compared to a threshold obtained from the tail area of the asymptotic distribution, and H_0 is rejected whenever the statistic exeeds the threshold.

3 Numerical Examples

The full version of this paper also presents a numerically reliable implementation of the ML-based detection/estimation scheme. Assuming $N \gg (m+p)$, as would typically be the case, the bulk of the implementation is the same QR-factorization used for solving the ordinary LS-problem. Thus, the only significant complexity increase of the exact ML method is due to the need for determining $\hat{n}$.

In the computer simulations presented below, an arbitrarily selected ϕ of dimensions $m = 10$, $p = 20$ and of rank $n = 5$ is used. The regression matrix is scaled such that $\|\phi\|_F = 1$, and then fixed throughout the simulation study. The exact ML estimate is compared to the ordinary LS estimate, as well as the same estimate truncated to rank n, using the singular value decomposition. The signal $x(t)$ and the noise $e(t)$ are both generated as zero-mean white Gaussian random processes. The covariance matrix of the signal is fixed at $R_{xx} = I$, whereas the noise covariance matrix is given by

$$\{Q\}_{kl} = \sigma^2 \left(-0.9\right)^{|k-l|} \ ,$$

which is reminiscent of a first order spatial AR-process with a pole at -0.9. In Figure 1, the total MSE is displayed versus the SNR. The estimates are calculated using a batch of $N = 100$ samples. Note that the MSE of the MLE follows the theoretical curve at high SNR values, and also that the LS-based methods perform notably worse in this scenario. The probability of correctly determining the rank of the regressor is displayed in Figure 2. A confidence level of 0.05 (according to the asymptotic distribution of the GLRT variable) is selected in the detection procedure. The SNR is here fixed at 10 dB and the number of samples is varied. About $N = 300$ samples are required for determining the correct rank of ϕ with high probability in this case. This might seem a large figure, but recall that the number of estimated parameters in the unconstrained ϕ is 200. Notice also that the probability of detection appears to settle at 95% for large N, as predicted by the theory.

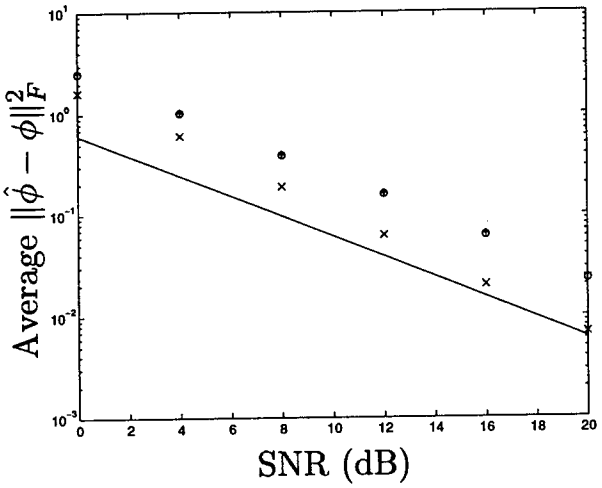


Figure 1. *Theoretical (solid curve) and empirical total mean square error versus signal-to-noise ratio. MLE ('x'), LS ('+') and Modified LS ('o').*

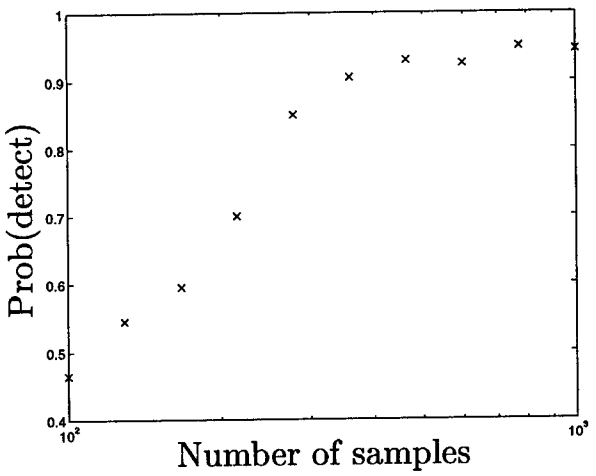


Figure 2. *Empirical probability of correct rank determination versus the number of samples.*

4 Conclusions

The exact ML estimator for a linear regression problem, where the regression matrix is known to be rank-deficient, is derived. An explicit expression for the estimator is found, employing a truncated canonical correlation decomposition. The computational complexity is similar to that of the ordinary least-squares (LS) estimator. However, the proposed technique takes into account the reduced rank in an optimal way, which can yield a significant performance improvement in difficult situations. A GLR test is proposed for determining the rank of the regressor. The asymptotic distribution of the parameter estimates are presented in the full version of this paper. The computer simulations indicate that the derived asymptotic results are useful in predicting the behavior in samples of practical lengths.

References

[1] T. Anderson. *An Introduction to Multivariate Statistical Analysis, 2nd edition.* John Wiley & Sons, New York, 1984.

[2] M. Jansson and B. Wahlberg. "4SID Linear Regression". In *Proc. 33rd CDC*, pages 2858–2863, Orlando, FL, 1994.

[3] C. Rao. *Linear Statistical Inference and its Applications.* Wiley, New York, 1973.

[4] L. Scharf. "The SVD and reduced rank signal processing". *Signal Processing (EURASIP)*, 25(2):113–133, 1991.

[5] T. Söderström and P. Stoica. *System Identification.* Prentice-Hall, London, U.K., 1989.

[6] P. Stoica and M. Viberg. "Maximum-Likelihood Parameter and Rank Estimation in Reduced-Rank Multivariate Linear Regressions". Technical Report CTH-TE-31, Chalmers University of Technology, Gothenburg, Sweden, July 1995. Submitted to IEEE Trans SP.

[7] D. Tufts and R. Kumaresan. "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood". *Proc. IEEE*, 70:975–989, Sept. 1982.

[8] Q. Wu and K. Wong. "UN-MUSIC and UN-CLE: An Application of Generalized Canonical Correlation Analysis to the Estimation of the Directions of Arrival of signals in unknown correlated noise ". *IEEE Trans. SP*, 42:2331–2341, Sept. 1994.

Robust Adaptive Beamforming Under Uncertainty in Source Direction-of-Arrival

Kristine L. Bell, Yariv Ephraim, and Harry L. Van Trees
Center of Excellence in C^3I
George Mason University, Fairfax, VA 22030-4444
kbell@gmu.edu, yephraim@gmu.edu, hlv@gmu.edu

Abstract

Adaptive beamforming can be used as a method for estimating an unknown waveform from a source impinging on an array of sensors. When the direction-of-arrival (DOA) of the incoming signal is known, the minimum variance distortionless response (MVDR) beamformer provides a distortionless version of the signal while suppressing noise and interference. However, if there is a mismatch between the look direction of the beamformer and the actual DOA of the signal, there can be significant degradation in performance. In this paper, we use a Bayesian approach with the MVDR criterion to derive an adaptive beamformer which has nearly optimal performance under good conditions, and is robust to uncertainty in DOA under poor conditions.

1 Introduction

Adaptive beamforming can be used as a method for estimating an unknown waveform from a source impinging on an array of sensors. When the direction-of-arrival (DOA) of the incoming signal is known, the minimum variance distortionless response (MVDR) beamformer [1] provides a distortionless version of the signal while suppressing noise and interference. However, if the source DOA is not known exactly, or if the source or array is moving, the mismatch between the actual DOA of the signal and the look direction of the beamformer can cause a significant degradation in performance [2].

Numerous methods have been proposed to overcome this sensitivity to pointing errors. These can generally be separated into two categories, "robust" adaptive beamformers, and "responsive" adaptive beamformers. Robust beamformers reduce sensitivity by widening and flattening the main beam around the presumed DOA. Some commonly used techniques are to impose point, derivative, or quadratic constraints on the beamformer output. Robust techniques generally work well under a wide range of scenarios, but sacrifice some performance with respect to the optimal beamformer informed of the true DOA.

Responsive beamformers attempt to "respond" to the current environment by learning or estimating the signal DOA from the observations, then using this information as if it were known exactly. Techniques of this type include estimating the DOA directly using maximum likelihood (ML) or some other estimation procedure, and learning the DOA indirectly by estimating the signal subspace. Under conditions where good DOA estimates can be obtained, i.e. for high signal-to-noise ratio (SNR) and a slowly fluctuating DOA, the responsive techniques have nearly the same performance as the beamformer informed of the true DOA. However, responsive techniques can have very poor performance under less favorable conditions.

In this paper, we use a Bayesian approach with the MVDR criterion to derive an adaptive beamformer which tends to be "responsive" under good conditions and "robust" under poor conditions.

2 Problem Formulation

We consider the problem of recovering, or estimating, the waveform of a narrowband planewave signal incident on an array of M sensors from DOA θ. The $M \times 1$ vector of received signals consists of a desired signal component and an additive noise component and has the form

$$\mathbf{x}(t) = \mathbf{a}(\theta)s(t) + \mathbf{n}(t), \tag{1}$$

where $s(t)$ is the desired signal, $\mathbf{n}(t)$ is the $M \times 1$ vector of additive noise, and $\mathbf{a}(\theta)$ is the $M \times 1$ "array response" or "steering vector" in the direction θ.

The beamformer applies a set of complex weights $\mathbf{w}$ to the received signals and sums to form the beamformer output

$$y(t) = \mathbf{w}^H \mathbf{x}(t). \tag{2}$$

0-8186-7576-4/96 $5.00

When the DOA θ is known, the weights for the MVDR beamformer are chosen to minimize the output power of the beamformer, $E\left\{|y(t)|^2\right\}$, while maintaining a distortionless response in the direction of the desired signal. The weights are found from the solution to

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_\mathbf{X} \mathbf{w} \quad \text{subject to} \quad \mathbf{a}(\theta)^H \mathbf{w} = 1, \tag{3}$$

where $\mathbf{R}_\mathbf{X}$ is the data correlation matrix

$$\mathbf{R}_\mathbf{X} = E\left\{\mathbf{x}(t)\mathbf{x}(t)^H\right\}. \tag{4}$$

The MVDR weight vector has the form

$$\mathbf{w} = \frac{\mathbf{R}_\mathbf{X}^{-1}\mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \mathbf{R}_\mathbf{X}^{-1}\mathbf{a}(\theta)}. \tag{5}$$

When θ is not known exactly, sensitivity to pointing errors can be reduced by imposing constraints on the shape of the main beam to widen and flatten it. One possibility is to impose constraints on the beamformer output at K values of θ near the presumed DOA. The weights are found from

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_\mathbf{X} \mathbf{w} \quad \text{subject to} \quad \mathbf{C}^H \mathbf{w} = \mathbf{c}, \tag{6}$$

where $\mathbf{C}$ is the $M \times K$ matrix of steering vectors for the constrained DOAs

$$\mathbf{C} = [\mathbf{a}(\theta_1) \ \cdots \ \mathbf{a}(\theta_K)], \tag{7}$$

and $\mathbf{c}$ is the $K \times 1$ vector of constraints. For a distortionless response to all the constrained DOAs, $\mathbf{c}$ is a vector of ones. The constrained weight vector has the form

$$\mathbf{w} = \mathbf{R}_\mathbf{X}^{-1}\mathbf{C}\left(\mathbf{C}^H \mathbf{R}_\mathbf{X}^{-1}\mathbf{C}\right)^{-1}\mathbf{c}. \tag{8}$$

Additional constraints can improve robustness to pointing errors, but hamper noise cancellation because they reduce adaptive degrees of freedom.

Alternatively, the unknown DOA θ can be estimated from L snapshots of the received data vector taken at times $t_1, \ldots, t_L$,

$$\mathbf{x}_L = [\mathbf{x}(t_1)^T \ \cdots \ \mathbf{x}(t_L)^T]^T. \tag{9}$$

The weight vector then has the same form as (5) with θ replaced by $\hat{\theta}(\mathbf{x}_L)$,

$$\mathbf{w} = \frac{\mathbf{R}_\mathbf{X}^{-1}\mathbf{a}(\hat{\theta}(\mathbf{x}_L))}{\mathbf{a}(\hat{\theta}(\mathbf{x}_L))^H \mathbf{R}_\mathbf{X}^{-1}\mathbf{a}(\hat{\theta}(\mathbf{x}_L))}. \tag{10}$$

This technique works well when the observed data is sufficient to yield good estimates of the DOA, but can result in significant mismatch when the estimates are poor.

In practice, the data correlation matrix $\mathbf{R}_\mathbf{X}$ is rarely known, and the beamformers weights in (5), (8), and (10) are implemented by substituting an estimate of $\mathbf{R}_\mathbf{X}$ such as the sample correlation matrix obtained from N snapshots of the data

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}(t_i)\mathbf{x}(t_i)^H. \tag{11}$$

The number of snapshots, N, used in estimating $\hat{\mathbf{R}}$ and the number of snapshots, L, used in estimating θ need not be the same. Both are chosen to tune the performance of the processor for the situation at hand. As a rule, both are set large enough so that good estimates of the desired quantity can be obtained, but small enough so that the estimates can follow temporal fluctuations.

3 Beamformer

We will use a Bayesian approach with the MVDR criterion to derive an adaptive beamformer which tends to be "responsive" under good conditions and "robust" under poor conditions. It is assumed that θ is a random parameter with a priori probability density function (pdf) $q(\theta)$, which reflects the level of uncertainty in the source DOA. The Bayesian approach has been used for detecting signals under directional uncertainty in [3], with averaging over the a priori pdf $q(\theta)$. The resulting detector was was robust, but required numerical integration over the a priori pdf. In order to obtain a simpler and more responsive beamformer, we will use an technique similar to that in [4]. Where averaging is needed, we will use the a posteriori pdf $p(\theta|\mathbf{x}_L)$ given L snapshots of the data vector. Furthermore, we will assume that $q(\theta)$ is defined only on a discrete set of P points, $\Theta = \{\theta_1 \ \cdots \ \theta_P\}$, in the a priori parameter space.

The objective is still to minimize the output power, but now the constraint is for a distortionless response on the average, i.e.,

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_\mathbf{X} \mathbf{w} \quad \text{subject to} \quad \bar{\mathbf{a}}^H \mathbf{w} = 1, \tag{12}$$

where $\bar{\mathbf{a}}$ is an average, or composite, steering vector averaged over $p(\theta|\mathbf{x}_L)$

$$\bar{\mathbf{a}} = \sum_{i=1}^{P} \mathbf{a}(\theta_i)p(\theta_i|\mathbf{x}_L) = \mathbf{A}\mathbf{p}, \tag{13}$$

where $\mathbf{A}$ is the $M \times P$ matrix of steering vectors

$$\mathbf{A} = [\mathbf{a}(\theta_1) \ \cdots \ \mathbf{a}(\theta_P)], \tag{14}$$

and the i^{th} element of $\mathbf{p}$ is $p(\theta_i|\mathbf{x}_L)$.

This results in beamformer weights of the form

$$\mathbf{w} = \frac{\hat{\mathbf{R}}^{-1}\mathbf{A}\mathbf{p}}{\mathbf{p}^T \mathbf{A}^H \hat{\mathbf{R}}^{-1}\mathbf{A}\mathbf{p}} \tag{15}$$

where $\hat{\mathbf{R}}$ has been substituted for $\mathbf{R}_{\mathbf{X}}$. A similar beamformer was derived in in [5] under different considerations. In [5], the vector $\mathbf{p}$ is not related to the a posteriori pdf, but is determined from a complicated optimization rule.

If we assume that the source and noise waveforms are sample functions of uncorrelated, zero-mean, stationary Gaussian random processes with variance σ_s^2, and covariance $\sigma_n^2\mathbf{I}$, respectively, then $p(\mathbf{x}_L|\theta_i)$ is a complex, zero-mean, Gaussian density with covariance

$$\mathbf{R}_{\mathbf{X}}(\theta_i) = \sigma_s^2\mathbf{a}(\theta_i)\mathbf{a}(\theta_i)^H + \sigma_n^2\mathbf{I}. \tag{16}$$

Applying Bayes rule, $p(\theta_i|\mathbf{x}_L)$ has the form:

$$p(\theta_i|\mathbf{X}) = \frac{q(\theta_i)\exp\left\{\beta L\mathbf{a}(\theta_i)^H\hat{\mathbf{R}}_L\mathbf{a}(\theta_i)\right\}}{\sum_{k=1}^{P} q(\theta_k)\exp\left\{\beta L\mathbf{a}(\theta_k)^H\hat{\mathbf{R}}_L\mathbf{a}(\theta_k)\right\}}. \tag{17}$$

where $\hat{\mathbf{R}}_L$ is the sample correlation matrix of $\mathbf{x}_L$ and β is a monotonically increasing function of SNR ($\gamma \equiv \frac{\sigma_s^2}{\sigma_n^2}$):

$$\beta = \frac{\gamma}{\sigma_n^2(1+M\gamma)}. \tag{18}$$

The SNR is not usually known, but β can be viewed as a variable which may be adjusted to tune the responsiveness of beamformer to the source SNR, just as N and L can be chosen to tune temporal responsiveness. The beamformer is updated in two steps. First the a posteriori pdf is found from (17), then the weights are calculated from (15).

The beamformer uses the same amount of observed data as was used in estimating θ in (10), and similar a priori information in determining the a priori pdf $q(\theta)$ as was needed in defining the point constraints in (8). In this beamformer, increasing number of DOAs in Θ does not reduce adaptive degrees of freedom, because they are averaged to form a composite steering vector. Adding points increases the computational complexity, and the number of points is chosen to cover the a priori parameter space sufficiently densely while keeping the computational requirements low.

4 Performance Example

We now consider a simple example to illustrate the performance of the proposed "a posteriori" beamformer as compared to the the MVDR beamformer informed of the source DOA, a "responsive" beamformer which uses the maximum likelihood estimate (MLE) of the DOA, and a "robust" beamformer which uses a set of point constraints over the a priori interval. The array is a uniform linear array (ULA) with half-wavelength spacing and $M = 8$ elements. The a priori uncertainty in the DOA is over the region $u = \sin(\theta) \in [-0.3, 0.3]$. For an 8-element array, this interval is slightly larger than the width of the mainlobe in the ideal beampattern. The set Θ is composed of $P = 13$ evenly spaced points on the interval $[-0.3, 0.3]$. For the constrained beamformer, we must use less than 8 constraints. Five distortionless constraints were used at the points $\{-0.3, -0.15, 0, 0.15, 0.3\}$. The source DOA was chosen to be $u_s = 0.223$, which does not coincide exactly with any of the constraint points or any of the points in Θ.

In Figures 1-4, typical performance is illustrated for a high SNR (0 dB) and low SNR (-20 dB) case, respectively. Figures 1 and 3 show the a posteriori pdf and typical beampatterns for a single trial in the two cases, and Figures 2 and 4 show a histogram of array gain for the different beamformers obtained from 500 trials. In the high SNR case, the a posteriori pdf is sharply peaked near the true DOA and the Bayesian beamformer, as well as the informed and MLE beamformers have nearly the same beampatterns, providing high gain to the source, and relatively low gain elsewhere. The array gain is relatively stable over all trials and close to the optimal value of $M = 8$ (9 dB). The constrained beamformer, in attempting to provide good gain over the entire a priori interval, does not suppress noise as well as the other beamformers, and has a lower array gain, close to 0 dB.

At low SNR, the a posteriori pdf is nearly equal to the a priori pdf, with some small peaks. The MLE of the DOA attempts to find the most likely estimate, but is not always accurate, resulting in a beamformer which does not always point at the desired signal. The histogram of array gain values for the MLE beamformer shows that the MLE is accurate enough to provide optimal performance only about half of the time, and can be so inaccurate as to reduce array gain as low as -25 dB. The constrained beamformer still provides good gain over the entire a priori region and a stable array gain of about 0 dB. Our Bayesian beamformer is now more robust, providing reasonable gain over the entire a priori interval, with increased gain at local maxima in the a posteriori pdf. The array gain is stable near a value which is less than optimal, but still better than the constrained processor.

References

[1] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE*, 57(8):1408–1418, August 1969.

[2] H. Cox. Resolving power and sensitivity to mismatch of optimum array processors. *J. Acoust. Soc. Amer.*, 54(3):771–785, September 1973.

[3] W. S. Hodgkiss and L. W. Nolte. Optimal array processor performance trade-offs under directional uncertainty. *IEEE Trans. Aero. Elec. Syst.*, 12(5):605–615, September 1976.

[4] D. T. Magill. Optimal adaptive estimation of sampled stochastic processes. *IEEE Trans. Automat. Contr.*, 10(4):434–439, October 1965.

[5] J. C. Preisig. Robust maximum energy adaptive matched field processing. *IEEE Trans. Sig. Proc.*, 42(7):1585–1595, July 1994.

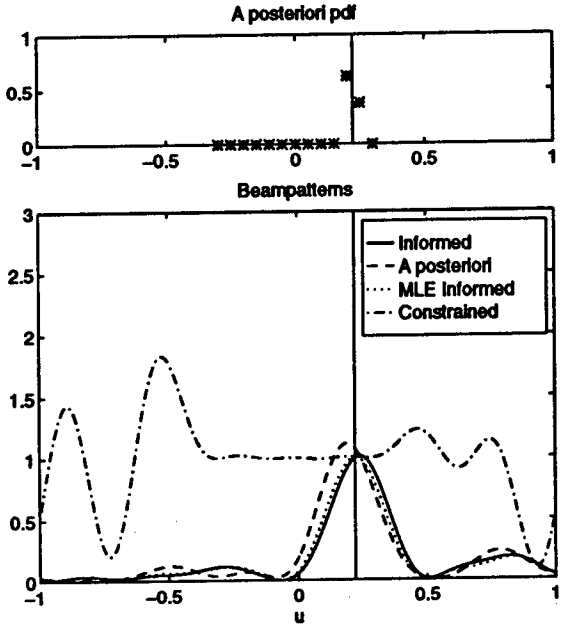


Figure 1. A posteriori pdf and beampatterns of adaptive beamformers for SNR = 0 dB. A priori interval = [-0.3,0.3], source DOA = 0.223.

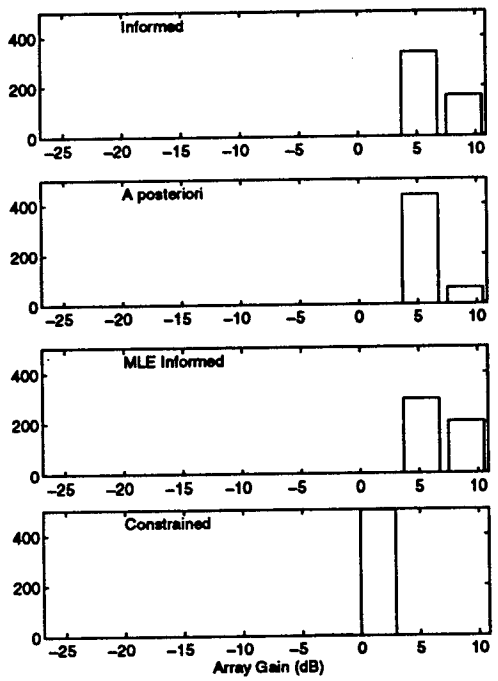


Figure 2. Histogram of array gain for beamformers from 500 trials for SNR = 0 dB.

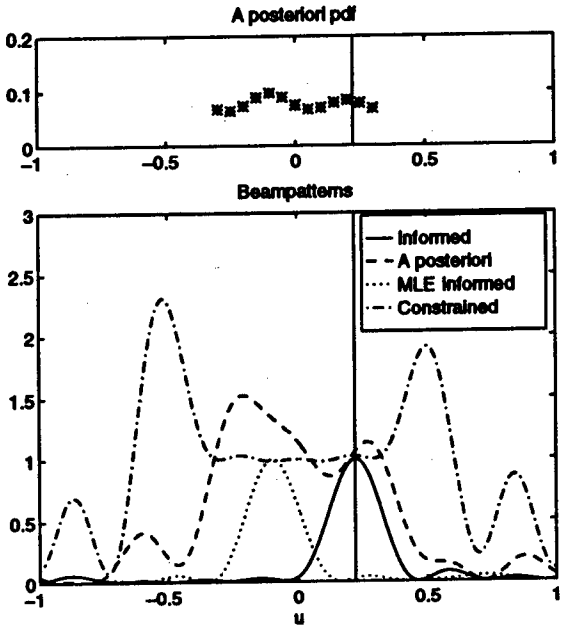


Figure 3. A posteriori pdf and beampatterns of adaptive beamformers for SNR = -20 dB. A priori interval = [-0.3,0.3], source DOA = 0.223.

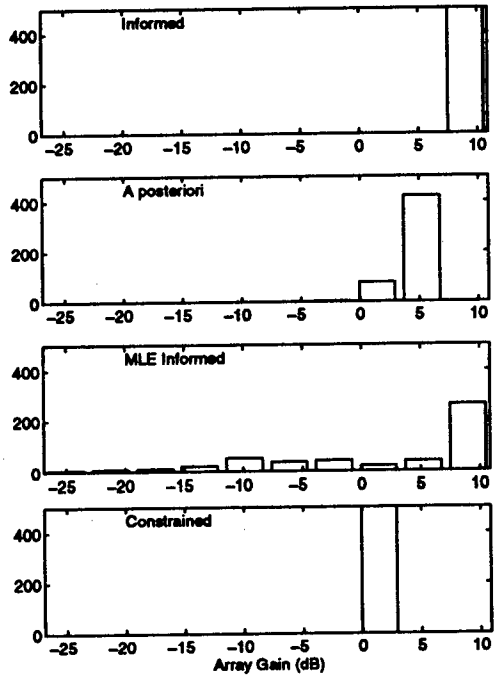


Figure 4. Histogram of array gain for beamformers from 500 trials for SNR = -20 dB.

Positive-Definite Toeplitz Completion in DOA Estimation for Partially-Augmentable Nonuniform Linear Antenna Arrays

Y. I. Abramovich* & N. K. Spencer
CSSIP, Technology Park, The Levels,
South Australia, 5095, Australia
yuri@cssip.edu.au & nspencer@cssip.edu.au

A. Y. Gorokhov
Département Signal, Télécom Paris,
46 rue Barrault, 75634, Paris, Cedex 13, France
gorokhov@sig.enst.fr

Abstract

Maximum-entropy positive-definite completion for partially-specified Toeplitz covariance matrices is developed for DOA estimation in partially-augmentable antenna arrays (those that have an incomplete set of covariance lags).

1. Introduction

This paper considers the problem of DOA (direction-of-arrival) estimation for multiple uncorrelated plane waves incident upon partially-augmentable antenna arrays. This type of array has an incomplete set of covariance lags [6]. Specifically, consider a nonuniform linear array (NLA) geometry specified by the sensor positions d_i $(i = 1, \ldots, M)$ and set $d_1 = 0$ for convenience. Let the unit spacing d be the greatest common divisor of the difference set

$$\mathcal{D} = \{d_i - d_j \mid i, j = 1, \ldots, M;\ i > j\}. \qquad (1)$$

Denote the maximum inter-element distance (array aperture) by $d(M_\alpha - 1)$. *Fully-augmentable* arrays have the property that all intermediate integral distances are realised; *ie.* given the sequence of natural numbers $\kappa = 1, \ldots, M_\alpha - 1$, we have $\kappa d \in \mathcal{D}$. On the other hand, *partially-augmentable* arrays have some nonzero number (G) of missed lags ("gaps"). It is clear that a partially-augmentable array gives rise to an incomplete augmented covariance Toeplitz matrix T, since some lags are missing. Thus both the spatial covariance matrix estimation problem and the spatial spectrum estimation problem must be formulated as p.d. (positive definite) Toeplitz completion problems.

The latter problem is investigated in this paper for two cases; firstly in the case where the available covariance lags are supposed to be precisely known (deterministic completion). Here we define the unique maximum-entropy p.d. Toeplitz completion and discuss its DOA-estimation performance under the condition that the number of uncorrelated plane waves (m) exceeds the number of antenna elements (M).

Secondly, we shall investigate the case where we assume we have sufficient statistics for the DOA estimates in the form of the direct data covariance (DDC) matrix $\hat{R}$, obtained by sample averaging on a set of N independent vectors ("snapshots") originating from a complex Gaussian distribution $\mathcal{CN}(M, 0, R)$.

Our benchmark will be the limiting accuracy provided by the Cramer-Rao bound.

2. Deterministic Matrix Completion

Consider the covariance matrix R of an M-element sparse array with assumed Gaussian processes observed as a combination of m uncorrelated plane waves with DOA's $\theta = [\theta_1, \ldots, \theta_m]^T$, powers $P = \mathrm{diag}[p_1, \ldots, p_m]$ and white noise of power σ:

$$R = BPB^H + \sigma I_M \qquad (2)$$

where the signal manifold matrix $B = [B(\theta_1), \ldots, B(\theta_m)]$,

$$B(\theta_i) = \left[1, \exp\left(i2\pi\frac{d_2}{\lambda}\sin\theta_i\right), \ldots, \exp\left(i2\pi\frac{d_M}{\lambda}\sin\theta_i\right)\right]^T \qquad (3)$$

is the so-called steering vector, and λ is the wavelength of incident radiation.

Let the set of presented covariance lags t_κ be $\mathcal{S}$, where $t_\kappa = R_{ij}$, $\kappa = (d_i - d_j)/d$, $i, j = 1, \ldots, M$ and let the set of missing lags be $\overline{\mathcal{S}}$ (in the language of matrix completion theory, these are the *specified* and *unspecified* values respectively). Fortunately the deterministic augmented covariance matrix $T = \mathrm{Toep}[t_\kappa]$ has at least one p.d. Toeplitz completion (T_{ULA} which is the covariance matrix of the corresponding uniform linear array) and hence the feasibility of the covariance matrix estimation problem is guaranteed.

*This study was partly supported by the INTAS SASPARC grant.

0-8186-7576-4/96 $5.00 © 1996 IEEE

Let the set of all possible p.d. Toeplitz completions $\mathcal{T}$ be introduced as follows:

$$\mathcal{T} = \left\{ z : T(z) = T + \sum_{\kappa \in \overline{\mathcal{S}}} \left(z_\kappa E_+^\kappa + z_\kappa^* E_-^\kappa \right) > 0 \right\} \quad (4)$$

where

$$E_+ = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{bmatrix}, \; E_- = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ & 1 & 0 & & \\ & & \ddots & \ddots & \\ & & & 1 & 0 \end{bmatrix}. \quad (5)$$

It can be shown [2] that the function

$$\phi(z) = \begin{cases} \log\det T^{-1}(z) & z \in \mathcal{T} \\ \infty & \text{otherwise} \end{cases} \quad (6)$$

is strictly convex on the feasible set, and so has a unique minimiser which we denote by z_{ME}^*:

$$z_{ME}^* := \arg\min_{\mathcal{T}} \phi(z) = \arg\max_{\mathcal{T}} \det T(z). \quad (7)$$

We refer to z_{ME}^* as the analytic centre of the linear matrix inequality $T(z) > 0$. For Gaussian distributions, we may evidently treat the analytic centre as a maximum-entropy completion.

A recently-developed convex programming approach [2] can be directly applied to find z_{ME}^*. The Ellipsoid Algorithm is first applied to find an arbitrary feasible point z_0^* such that $T(z_0^*) > 0$. If such a feasible point is found, any convergent minimiser of $\phi(z)$ should find the analytic centre. Nesterov and Nemirovskii's [5] version of Newton's method has been implemented here to find the optimum solution.

Thus the unique maximum-entropy p.d. Toeplitz completion may now be found for the partially specified covariance matrix T induced by the partially-augmentable array. Since the ME completion $T(z_{ME}^*)$ does not coincide with the true M_α-variate covariance matrix R, a further step is proposed to truncate the signal subspace dimension.

The ME-optimum completion may be treated as an initial estimate to the solution T_{mME} of the following optimisation problem:

$$\text{Find } \inf \left\| T(z_{ME}^*) - T_m \right\|_F \text{ subject to } T_m \in \mathcal{C} := \mathcal{C}_1 \cap \mathcal{C}_2 \quad (8)$$

where

$$\begin{aligned} \mathcal{C}_1 &= \left\{ T_m \in \mathcal{H}^{M_\alpha} : T_m \text{ is Toeplitz and p. d.} \right\} \\ \mathcal{C}_2 &= \left\{ T_m \in \mathcal{H}^{M_\alpha} : (T_m - \lambda_{min} I) \text{ is of rank } m \right\}, \end{aligned} \quad (9)$$

and where $\mathcal{H}^p$ is the space of $p \times p$ Hermitian matrices, and λ_{min} is the minimum eigenvalue of T_m, of multiplicity $(M_\alpha - m)$. Convergent alternating projection methods described in [3] are used to find the (at least local) extremum for this problem.

3. Stochastic Matrix Completion

When the direct augmentation approach [1] is used to obtain the specified lags $\hat{t}_\kappa$ via the stochastic DDC matrix $\hat{R}$, the feasibility condition is no longer guaranteed to hold. In our terms, the feasibility condition deals with conditions under which a p.d. Toeplitz matrix completion exists for the given $\hat{t}_\kappa$, $\kappa \in \mathcal{S}$.

Unfortunately, the necessary and sufficient feasibility conditions for the general p.d. Toeplitz completion problem have not yet been found [4]. One of the obvious necessary conditions is that every specified principal sub-matrix should be positive definite. Denoting the greatest specified sub-matrix of $\hat{T}$ by $\hat{T}_{N_{\max}}$, this necessary but obviously not sufficient condition is

$$\hat{T}_{N_{\max}} > 0. \quad (10)$$

Thus the success of the Ellipsoid Algorithm can be treated as our only feasibility condition. When the condition at Eqn. (10) is not satisfied, or the Ellipsoid Algorithm fails to find a feasible point, we need to modify the initial set of specified and unspecified sample covariance lags in order to achieve feasibility with the minimum possible deflection from the initial set of (maximum likelihood) estimates $\hat{t}_\kappa \in \mathcal{S}$. Let

$$\hat{T}(z) = \hat{T}_o + \sum_{\kappa \in \mathcal{S}} \left(z_\kappa E_+^\kappa + z_\kappa^* E_-^\kappa \right) + \sum_{\kappa \in \overline{\mathcal{S}}} \left(z_\kappa E_+^\kappa + z_\kappa^* E_-^\kappa \right) \quad (11)$$

then the minimum deflective feasible point search is:

$$\text{Find } \min \sum_{\kappa \in \mathcal{S}} |z_\kappa|^2 \quad \text{subject to} \quad \hat{T}(z) > 0. \quad (12)$$

Procedures elaborated in [2] provide a straight-forward approach to finding the unique optimum solution for this problem.

One such procedure uses the Ellipsoid Algorithm to find the optimal feasible point $\hat{T}(z_0^{opt})$. Another approach adopts the penalty function method, which admits simultaneous completion and deflection:

$$\text{Find} \quad \min \Phi(z) = \log\det_{\kappa \in \overline{\mathcal{S}}} \hat{T}^{-1}(z_\kappa) + \mu \sum_{\kappa \in \mathcal{S}} |z_\kappa|^2 \quad (13)$$

$$\text{subject to} \quad \hat{T}(z) > 0, \quad (14)$$

$$\text{with} \quad \hat{T}(z) = \hat{T}_0 + \sum_{\kappa=1}^{M_\alpha - 1} \left(z_\kappa E_+^\kappa + z_\kappa^* E_-^\kappa \right). \quad (15)$$

Evidently, as $\mu \to \infty$ we expect the optimum solution of this problem to coincide with the previous approach. Since both the entropy and the deflection norms relate to the DOA estimation accuracy, an appropriate trade-off between the two which optimises the DOA estimation accuracy is desired.

4. Simulation Results

To demonstrate the efficiency of the proposed ME completion approach, we consider a sparse array with $M = 5$ elements. The sensors are at positions $\boldsymbol{d} = \{0, 1, 4, 9, 11\}$, measured by half-wavelengths. The single missed covariance lag is t_6. We start with the deterministic matrix completion problem, where the exact values of the specified covariance lags were calculated. We have $m = 8$ sources having a common SNR ratio of 20 dB with respect to white noise, *ie.* $p_i = 100; \sigma = 1$ in Eqn. (2). The sources are uniformly separated in spatial frequency ($w = \pi sin\theta$) and the inter-source separation is $\Delta w = 0.08$ for the first scenario (Fig. 1) and $\Delta w = 0.12$ for the second scenario (Fig. 2).

The dotted line in Fig. 1 demonstrates the behaviour of the ME spectrum obtained when the 12-variate spatial covariance matrix T_{ME} is restored via the ME-completion algorithm. For comparison, the dashed line shows the ME spectrum obtained for the corresponding 12-element uniform linear array exact covariance matrix T_{ULA}. Also illustrated are the DOA's estimated by root-MUSIC applied to the ME-completed covariance matrix T_{ME}. Fig. 2 compares the behaviour of the ME spectra and root-MUSIC DOA estimates for the ME-completion algorithm with the associated ULA spectra for the case $\Delta w = 0.12$.

These results demonstrate that the ME spectra obtained via the ME-completion algorithm for sparse arrays practically coincide with the corresponding ULA ME spectrum. In this sense the "maximum entropy" properties of the ULA are fully restored by the ME-completion approach applied to the sparse array.

Meanwhile, for both scenario, the ME spectral maxima do not correspond to the true DOA's and even the number of main peaks is erroneous. Nevertheless for the greater spatial separation, the root-MUSIC DOA estimates calculated for T_{ME} locate the true DOA's with negligible errors. For the smaller separation, the root-MUSIC DOA estimates are essentially erroneous. This once again demonstrates that the maximum entropy criterion is inconsistent with the harmonic analysis criterion, especially for severe "super-resolution" conditions.

To verify this, Fig. 3 illustrates the behaviour of the Cramer-Rao bound for DOA estimation accuracy as a function of spatial frequency separation Δw for the same 8-source scenario with $N = 1000$ snapshots. The maximum RMSE from the eight sources is depicted by the dotted line. Note that a separation of ($\Delta w = 0.08$) is far beyond the realistic resolution capabilities of the antenna examined. However in the area where the accuracy limit is reasonable ($\Delta w > 0.15$), ME-completion applied to the deterministic matrix provides practically unbiased DOA estimates via the MUSIC / root-MUSIC approach. Thus our approach for situations in this region provides asymptotically-unbiased estimations and we may now analyse the mean and RMSE of the stochastic errors for the finite sample size N.

The solid and dashed lines in Fig. 3 illustrate the maximum sample DOA RMSE and bias respectively for the eight sources as a function of the inter-source spatial separation. In each of the 1000 trials, the MUSIC algorithm was applied to the p.d. finite signal subspace Toeplitz matrix T_{mME}, which is obtained via the ME completion algorithm with an initially modified data set (Eqns. (11) and (12)).

These results clearly define the pre-asymptotic domain in this case as $\Delta w \leq 0.16$, where the false peaks of the MUSIC sample pseudo spectra often give rise to completely erroneous DOA estimates. Following [1], we define "abnormal" DOA estimates as those estimates $\hat{\omega}_i$ lying outside the range $[\omega_i - \frac{\Delta\omega}{2},\ \omega_i + \frac{\Delta\omega}{2}]$. For this simulation, a total of 627 abnormal trials were rejected (in the process of ensuring that 1000 normal trials were finally obtained) for $\Delta w = 0.15$; while there were 29 abnormal trials for $\Delta w = 0.16$. In the asymptotic domain $\Delta w > 0.16$ (where the number of abnormal estimates vanishes), DOA estimation accuracy is reasonably close to the Cramer-Rao bound, similarly to fully-augmentable antenna arrays [1], although the bias here remains nonzero.

It is interesting to note that in this simulation, the necessary condition $\hat{T}_{N_{max}} > 0$ was able to detect between 0% and 60% of all the initial unfeasible sets of covariance lag estimates, depending on source separation.

In some simulations, minimum-deflection completion (Eqns. (11) – (12)) is significantly improved by the penalty method (Eqns. (13) – (15)). For example, for $m = 6$ sources separated by $\Delta w = 0.19$ and a penalty value $\mu = 10^{-5}$, the maximum RMSE and maximum bias are reduced from 0.011 to 0.005 and 0.034 to 0.021 respectively.

5. Summary

The above convex programming technique provides a unique solution to the problem of maximum-entropy p.d. Toeplitz completion (spectral estimation) for the incomplete Toeplitz augmented covariance matrix that meets the feasibility condition. When the Ellipsoid Algorithm fails to find an arbitrary feasible point for stochastic covariance lag estimates, the modification approach is proposed.

It has been shown that the deterministic ME spectra obtained by this technique are practically identical to the ULA ME spectra for all situations examined. However, this similarity does not necessarily imply that the corresponding root-MUSIC DOA estimates are true.

For situations when the number of abnormal estimates vanishes, the actual DOA estimation accuracy obtained by this new approach is demonstrated to be reasonably close to the Cramer-Rao limit.

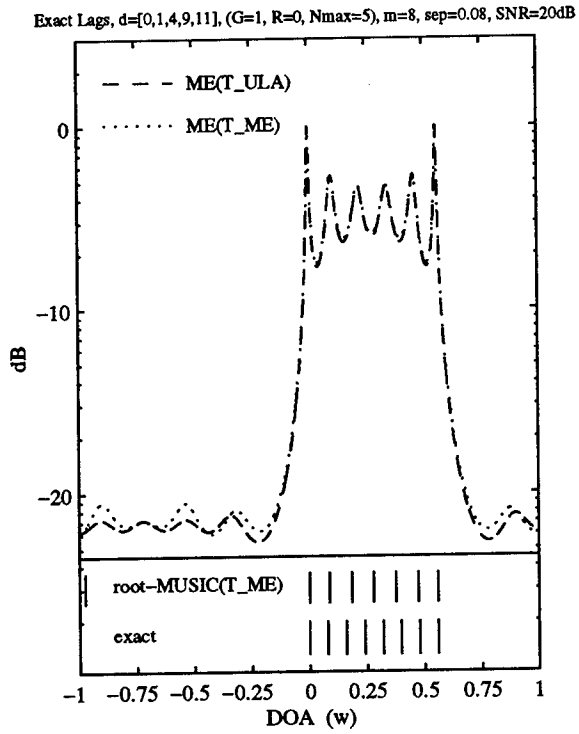


Figure 1. Deterministic completion, $\Delta w = 0.08$.

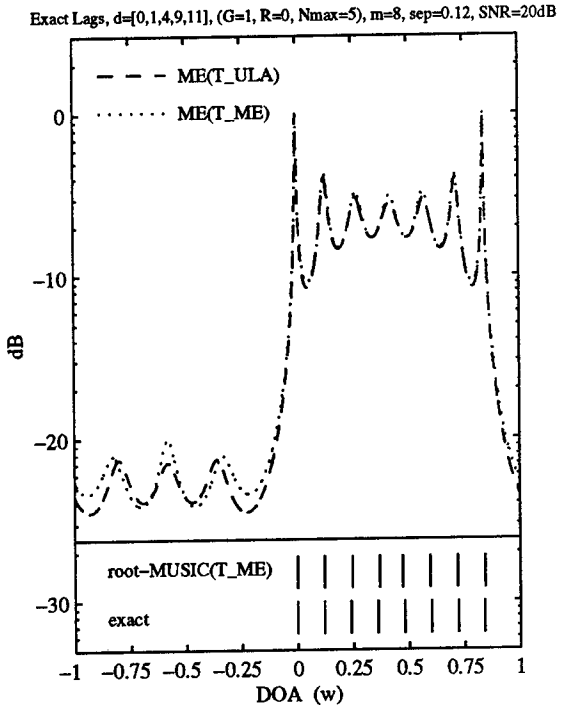


Figure 2. Deterministic completion, $\Delta w = 0.12$.

References

[1] Y. Abramovich, D. Gray, A. Gorokhov, and N. Spencer. Positive-definite Toeplitz completion in DOA estimation for fully-augmentable nonuniform linear antenna arrays. In *Proc. ICASSP 96*, Atlanta, 1996.

[2] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. Society for Industrial and Applied Mathematics, Philadelphia, 1994. SIAM Studies in Applied Mathematics Vol. 15.

[3] K. Grigoriadis, A. Frazho, and R. Skelton. Application of alternating convex projection methods for computation of positive definite Toeplitz matrices. *IEEE Trans. on Signal Processing*, 42 (7):1873–1975, 1994.

[4] C. Johnson. Matrix completion problems: a survey. *Proc. Symp. Applied Math.*, 40:171–198, 1990.

[5] Y. Nesterov and A. Nemirovskii. *Interior-point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, Philadelphia, 1994. SIAM Studies in Applied Mathematics Vol. 13.

[6] M. Sverdlik. *Optimal Discrete Signals*. Sovietsckoe Radio, Moscow, 1975. In Russian.

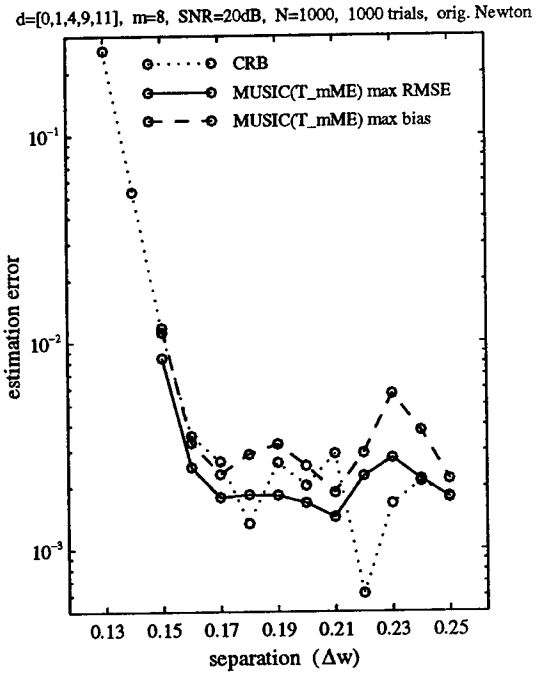


Figure 3. DOA estimation performance statistics compared with the Cramer-Rao bound.

Wide-Band Maximum Likelihood Direction Finding by Using Tree-Structured EM Algorithm

Nail Çadallı and Orhan Arıkan
Department of Electrical and Electronics Engineering, Bilkent University,
TR-06533 Ankara, TURKEY. Phone: 90-312-2664307, Fax: 90-312-2664126,
e-mail: cadalli@ee.bilkent.edu.tr, oarikan@ee.bilkent.edu.tr

Abstract

A tree structured Expectation Maximization (EM) algorithm is proposed and applied to the wide-band angle of arrival estimation. It may be seen as a generalization on EM using the ideas of Cascade EM algorithm and Space Alternating Generalized EM algorithm. Also, for passive data acquisition, robust and efficient alternatives for the estimation of the source signals are investigated.

1. Introduction

In many data acquisition systems, reception data acquired by an array of sensors are processed to obtain information about the source locations. When the sources are located relatively far away from the sensors, only the direction of arrivals of the acquired source signals can be reliably obtained. Although the Maximum likelihood (ML) estimation provides more accurate estimates for the direction of arrivals, due to the higher computational cost of obtaining the ML estimates, it has not found much use in practice. However, by exploiting the superposition property of the data acquisition system, the complexity of the ML estimation can be greatly reduced by using the Expectation Maximization (EM) algorithm [1, 2, 5]. In EM formalism, the observation, *incomplete data* is obtained via a many-to-one mapping from the *complete data* space that includes signals which we would obtain as the sensor outputs if we were able to observe the effect of each source separately. The EM algorithm iterates between estimating the log-likelihood of the complete data using the incomplete data and the current parameter estimates (E-step) and maximizing the estimated log-likelihood function to obtain the updated parameter estimates (M-step). Under mild regularity conditions, the iterations of the EM algorithm converges to a stationary point of the observed log-likelihood function, where at each iteration the likelihood of the estimated parameters is increased [11]. In this study, a tree structured hierarchy is used for the description of relation between the complete data space and the observations. Within this hierarchy it is possible to combine in one algorithm the ideas of the Cascade EM and Space Alternating Generalized EM algorithms [3, 7]. For the estimation of unknown signals arriving from different directions to a passive array, alternative regularized estimation schemes to the common least squares solution are investigated. For this purpose two different methods are used. The first one is an adaptive Tikhonov type regularized least-squares (RLS) estimation method, which is computationally intensive and the second one is an averaged least-squares estimation (LSSET) method over a set of angles in a neighborhood of the nominal angles. It has been demonstrated that when RLS or LSSET methods are used in the estimation of the received signals, the EM algorithm has better convergence behavior.

2. Signal Model

For the case of M sources with direction of arrivals θ_l, $1 \leq l \leq M$, the measured signal at the i'th sensor of an array with P sensors is

$$y_i(t) = \sum_{l=1}^{M} a_i(t,\theta_l) * s_l(t - \tau_i(\theta_l)) + u_i(t)$$
$$1 \leq i \leq P\ , \quad \text{t=0,T,2T,...,(N-1)T} \qquad (1)$$

where $s_l(t)$ is the wide-band signal of the l'th source, $u_i(t)$ is the 0 mean spatially and temporally white Gaussian noise at the i'th sensor, $\tau_i(\theta)$ is the time delay of the source signal from the direction θ as it propagates to the i'th sensor relative to the phase center of

0-8186-7576-4/96 $5.00

the array, $a_i(t,\theta)$ is the time domain function for the gain of the i'th sensor which is dependent on frequency and the direction of arrival, θ. The frequency domain representation of (1) is,

$$Y_i(k) = \sum_{l=1}^{M} A_i(k,\theta_l) e^{-j\frac{2\pi k}{F}\frac{\tau_i(\theta_l)}{T}} S_l(k) + U_i(k)$$
$$1 \le i \le P, \quad 0 \le k < F \qquad (2)$$

where F is the DFT size which is chosen sufficiently large and $Y_i(k)$, $A_i(k,\theta)$, $S_l(k)$ ve $U_i(k)$ are the transforms of $y_i(t)$, $a_i(t,\theta)$, $s_l(t)$ and $u_i(t)$ respectively. Let the following definitions be made ($'$ is the transpose operator)

$$\begin{aligned}
\mathbf{b}(k,\theta) &= [A_1(k,\theta)e^{-j\frac{2\pi k}{F}\frac{\tau_1(\theta)}{T}} \cdots \\
&\qquad A_P(k,\theta)e^{-j\frac{2\pi k}{F}\frac{\tau_P(\theta)}{T}}]' \\
&= [b_1(k,\theta)\cdots b_P(k,\theta)]' \\
\mathbf{B}(k,\Theta) &= [\mathbf{b}(k,\theta_1)\cdots\mathbf{b}(k,\theta_M)] \\
\mathbf{S}(k) &= [S_1(k)\cdots S_M(k)]' \\
\mathbf{Y}(k) &= [Y_1(k)\cdots Y_P(k)]'
\end{aligned}$$

Using these definitions (2) becomes

$$\mathbf{Y}(k) = \mathbf{B}(k,\Theta)\mathbf{S}(k) + \mathbf{U}(k) \quad 0 \le k < F \qquad (3)$$

This final compact form of the measurement relation, which is the same as the signal model of the Cramer-Rao Lower Bound formula in [8], will be used in our derivations.

3. Wide-Band EM Algorithm

Since, the measurement noise is modeled as normally distributed additive noise, the probability density of the observations are Gaussian. Hence, the log-likelihood function of the observations has the following familiar form († is the conjugate transpose operator),

$$\mathcal{L}(\Theta,\mathbf{S};\mathbf{Y}) = -\sum_{k=0}^{F-1} [\mathbf{Y}(k) - \mathbf{B}(k,\Theta)\mathbf{S}(k)]^\dagger$$
$$[\mathbf{Y}(k) - \mathbf{B}(k,\Theta)\mathbf{S}(k)] \qquad (4)$$

In order to find the ML estimate, likelihood function of the observations should be maximized with respect to Θ and $\mathbf{S}(k)$. However, the direct maximization of this function is not only computationally demanding but also due to the local maxima structure of the likelihood function it is not guaranteed to converge to the global maxima. The Expectation Maximization (EM) method of obtaining the ML estimate overcomes this difficulty by an iterative search in much lower dimensional parameter spaces [1]. The EM method requires the identification of so called complete data space. In our application the commonly used complete data is $\mathbf{X}_l(k) = [X_{1l}(k)\cdots X_{Pl}(k)]'$ which is the signal that would be observed at the sensors if we were able to see the effect of l'th source only. Then the many-to-one mapping for all sources from the complete data space to the incomplete data space can be written as

$$\mathbf{Y}(k) = \sum_{l=1}^{M} \mathbf{X}_l(k) \qquad 0 \le k < F \qquad (5)$$

The mean of the complete data $\mathbf{X}_l$(k) is $\mathbf{b}(k,\theta_l)S_l(k)$ and it is normally distributed. The log-likelihood function of the complete data is

$$\mathcal{L}_c(\Theta,\mathbf{S};\mathbf{X}) = -\sum_{k=0}^{F-1}\sum_{l=1}^{M} \|\mathbf{X}_l(k) - \mathbf{b}(k,\theta)S_l(k)\|^2 \qquad (6)$$

Here, the observed signal is decomposed to M constituents. Therefore to estimate θ_l and $S_l(k)$, only $\mathbf{X}_l(k)$ is used besides the observation. At the n'th iteration of the EM algorithm expectation step conditionally estimates the likelihood of the complete data $\mathcal{L}_c(\Theta,\mathbf{S} \mid \Theta^n,\mathbf{S}^n)$. Maximization step then finds the maximizer of the estimated likelihood and assigns to θ_l^{n+1}. To find $\mathbf{b}(k,\theta_l)S_l(k)$ it is sufficient to know $\mathbf{X}_l(k)$, therefore in expectation step $\mathbf{X}_l(k)$ is estimated. It can be shown that, ([6], p. 164),

$$\begin{aligned}
\mathbf{X}_l^n(k) &= \mathcal{E}\{\mathbf{X}_l(k)|\theta_l^n, S_l^n(k), \mathbf{Y}(k)\} \\
&= \mathbf{b}(k,\theta_l^n)S_l^n(k) + \frac{1}{M}[\mathbf{Y}(k) - \mathbf{B}(k,\Theta^n)\mathbf{S}^n(k))] \\
&\qquad 0 \le k < F \qquad (7)
\end{aligned}$$

In maximization step complete data likelihood which is formed by using $\mathbf{X}_l^n$(k) is maximized with respect to θ_l and $S_l(k)$. The θ_l update is found as

$$\theta_l^{n+1} = arg\max_{\theta} \left\{ \max_{S_l} \left\{ -\sum_{k=0}^{F-1}\sum_{l=1}^{M} \|\mathbf{X}_l^n(k) - \mathbf{b}(k,\theta)S_l(k)\|^2 \right\} \right\} \qquad (8)$$

where there is two maximization problems inside one another. If $S_l(k)$ is unknown they must be simultaneously solved. For a given θ value, the solution of the inner maximization is

$$\begin{aligned}
S_l(k) &= [\mathbf{b}(k,\theta)\mathbf{b}^\dagger(k,\theta)]^{-1}\mathbf{b}^\dagger(k,\theta)\mathbf{X}_l(k) \\
&= \frac{\mathbf{b}^\dagger(k,\theta)\mathbf{X}_l(k)}{\|\mathbf{b}(k,\theta)\|^2} \qquad (9)
\end{aligned}$$

Inserting this expression into (8) and solving for the outer maximization θ_l^{n+1} is found. For that maximization linear search may be used. Finally, at the n'th iteration of the EM algorithm the update formulas are as follows,

$$\theta_l^{n+1} = arg\max_{\theta} \sum_{k=0}^{F-1} \frac{\mathbf{b}^\dagger(k,\theta)\mathbf{X}_l{}^n(k)\mathbf{X}_l{}^{\dagger n}(k)\mathbf{b}(k,\theta)}{\|\mathbf{b}(k,\theta)\|^2} \quad (10)$$

$$S_l^{n+1}(k) = \frac{\mathbf{b}^\dagger(k,\theta_l^{n+1})\mathbf{X}_l^n(k)}{\|\mathbf{b}(k,\theta_l^{n+1})\|^2} \quad (11)$$

If $S_l(k)$ is known, as in active array applications, (8) is simply reduced to one maximization problem and there remains no need for (11). If (10) and (11) are run together, i.e., in the case of unknown source signals, θ_l^{n+1} should be close to true direction values for S_l^{n+1} to converge to true signal waveforms.

After (10), $\mathbf{\Theta}^{n+1}$ is available. If it is inserted into (3), $\mathbf{S}^{n+1}(k)$ can be solved for by using a number of alternatives. For instance the least squares (LS) solution is as follows,

$$\begin{aligned}\mathbf{S}(k) &= arg\min_{\mathbf{S}(k)} \|\mathbf{Y}(k) - \mathbf{B}(k,\mathbf{\Theta})\mathbf{S}(k)\|^2 \\ &= [\mathbf{B}^\dagger(k,\mathbf{\Theta})\mathbf{B}(k,\mathbf{\Theta})]^{-1}\mathbf{B}^\dagger(k,\mathbf{\Theta})\mathbf{Y}(k) \quad (12)\end{aligned}$$

Regularization may be applied on the LS solution which is called regularized least squares, RLS,

$$\mathbf{S}(k) = [\mathbf{B}^\dagger(k,\mathbf{\Theta})\mathbf{B}(k,\mathbf{\Theta}) + \mu\mathbf{I}]^{-1}\mathbf{B}^\dagger(k,\mathbf{\Theta})\mathbf{Y}(k) \quad (13)$$

It is important to chose μ in the regularization and it can be chosen optimally [4, 9]. Another alternative in source signal estimation may be the following which will be referred to as LSSET solution, where $\mathcal{K}$ is a set of angles in a neighborhood of $\mathbf{\Theta}$,

$$\mathbf{S}(k) = arg\min_{\mathbf{S}(k)} \int_{\mathcal{K}} \|\mathbf{Y}(k) - \mathbf{B}(k,\mathbf{\Theta})\mathbf{S}(k)\|^2 d\mathbf{\Theta} \quad (14)$$

EM algorithm starts with $n = 0$ at which time $\mathbf{\Theta}^0$ is available obtained by using a rough estimation. To find $\mathbf{X}_l^0(k)$ in (7), S_l^0 is needed and it is estimated by one of the methods mentioned above. EM shows monotonic increase of the likelihood and its convergence issues have been investigated [1, 11].

4. Tree-Structured EM

In this section we will use a different mapping from the complete data to the incomplete data which is structured as a binary tree as shown in Figure 1. $\mathbf{Y}_{i,j}(k)$ is the intermediate incomplete data between the observation $\mathbf{Y}(k)$ and the complete data $\mathbf{X}_l(k)$'s. In this setting EM algorithm is run for two sources at a time using the intermediate data at the joint node of two leaves. This provides an update for the corresponding DOA and source signals. The value of the intermediate data is found by using, in (3), the original observation $\mathbf{Y}(k)$ and the current source signal estimates other than the ones which are to be updated by the current run. For instance, to run EM algorithm for $\mathbf{X}_1(k)$ and $\mathbf{X}_2(k)$ we form the required incomplete data as

$$\mathbf{Y}_{2,1}(k) = \mathbf{Y}_{1,1}(k) - \mathbf{B}(k, \begin{bmatrix}\theta_3 \\ \theta_4\end{bmatrix}) \begin{bmatrix}\mathbf{S}_3(k) \\ \mathbf{S}_4(k)\end{bmatrix} \quad (15)$$

where $\mathbf{Y}_{1,1}(k)$ is found by using $\mathbf{Y}(k)$ and the current estimates for the last there source signals in (3). $\mathbf{Y}_{2,2}(k)$ may be found similarly and EM algorithm is run for that branch too. This may be repeated a number of times and then by using the updates obtained for the first 4 source signals and DOA's, branch of $\mathbf{Y}_{1,2}(k)$ may be processed. The idea of putting inter-

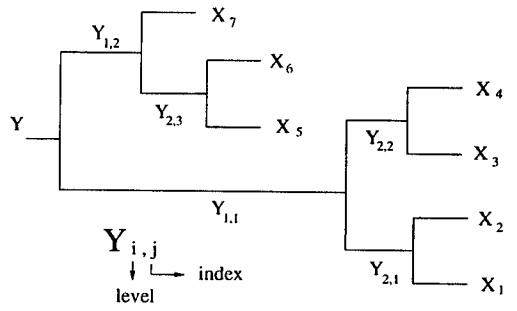


Figure 1. An example for the tree structure.

mediate data mappings between $\mathbf{Y}(k)$ and $\mathbf{X}_l(k)$'s can be associated with that of the Cascade EM, CEM, algorithm but here there is more than one intermediate data space. Due to the limited space, the generalization of CEM to multiple levels is not presented here. The tree structure may also be associated with Space Alternating Generalized EM algorithm in the sense that not all of the parameters are updated at a time. Also EM is run on a more noisy data reducing the information content of intermediate observations and this is reported to speed the convergence [3].

5. Simulations and Conclusions

Observation signals are obtained by simulation of a linear array of sensors. The number of signals are assumed to be known since there are studies in detection

[10]. The source signals are taken as coherent pulse modulated chirp signals with bandwidth comparable to the center frequency. Noise is assumed to be independent identical Gaussian distributed. First the

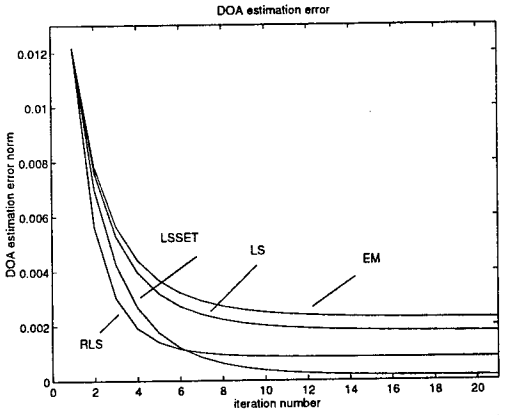


Figure 2. Averaged traces of DOA error norms.

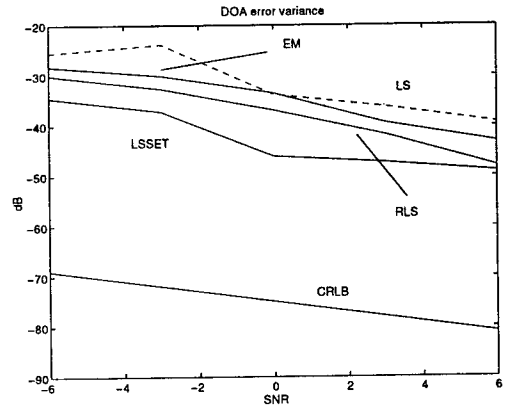


Figure 3. DOA error variances.

signal estimation alternatives are inserted in the EM algorithm and their relative performances are compared. EM algorithm is run for two sources impinging from 35° and −50° at an SNR level of 0dB. The averaged traces of error norm of DOA estimation, which describes the convergence behaviours, can be seen in Figure 2 where EM, LS, RLS, LSSET refers to (11), (12), (13), (14) respectively. The DOA error variance together with the CRLB for each alternative is plotted in Figure 3. For LSSET, $\mathcal{K}$ consists of 5 angles in a 1° neighborhood of the current DOA. This figures out to be computationally less complex than RLS.

To compare the tree-structured EM algorithm with the original EM algorithm four sources from directions $\Theta = [35° \ -50° \ -20° \ 50°]'$ are used at SNR=10dB. Initial DOA's are given as $\Theta_0 = [33° \ -48° \ -18° \ 48°]'$. The DOA error norms for iterations of original EM and tree-structured EM algorithms are shown in the next table. The original EM algorithm, could not converge to true DOA values. Furthermore, it diverges from the initial angle values. But within the same number of total iterations the tree-structured EM converges with much lower DOA error to $\Theta = [35.3 \ -50.0 \ -20.0 \ 50.7]'$.

iteration no. →	10	20	50	100
EM (10^{-3})	5.3	5.5	5.5	5.5
Tree-EM (10^{-4})	6.3	6.1	4.6	2.2

By this study, an improvement on EM algorithm is realized not only by using robust signal estimation schemes but also by changing the data mapping of the original algorithm.

References

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.*, B-39:1–37, 1977.

[2] M. Feder and E. Weinstein. Parameter estimation of superimposed signals using the EM algorithm. *IEEE Trans. Acoust., Speech, Signal Processing*, 36(4):477–489, April 1988.

[3] J. A. Fessler and A. O. Hero. Space alternating generalized expectation-maximization algorithm. *IEEE Trans. Signal Processing*, 42(10):2664–2677, Oct. 1994.

[4] A. Hoerl and W. Kennard. Rigde regression: biased estimation for non-orthogonal problems. *Technometrics*, 12:55–68, 1970.

[5] M. I. Miller and D. R. Fuhrmann. Maximum-likelihood narrow-band direction finding and the EM algorithm. *IEEE Trans. Acoust., Speech, Signal Processing*, 38(9):1560–1577, Sep. 1990.

[6] A. Papoulis. *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, 3rd edition, 1991.

[7] M. Segal and E. Weinstein. The cascade em algorithm. *Proc. IEEE*, 76(10):1388–1390, Oct. 1988.

[8] P. Stoica and A. Nehorai. MUSIC, Maximum Likelihood and Cramer-Rao Bound. *IEEE Trans. Acoust., Speech, Signal Processing*, 37(5):720–741, May 1989.

[9] H. D. Vinod. Survey of ridge regression and related techniques for improvements over ordinary least squares. *Review of Economics and Statistics*, 60:121–131, 1978.

[10] H. Wang and M. Kaveh. Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-33(4):823–831, Aug. 1985.

[11] C. F. J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 1983.

High Performance Doa Trackers Derived From Parallel Low Resolution Detectors

Ana Perez-Neira, Miguel A. Lagunas*.
Department of Signal Theory and Communications. Universitat Politècnica de Catalunya
Campus Nord UPC. Edificio D-5. c/ Gran Capità, s/n.
08034 BARCELONA, SPAIN
Phone:34-3-4016436. Fax: 34-3-4016447.
e-mail: anuska@gps.tsc.upc.es

Abstract

Traditionally, high resolution spectral Direction Of Arrival (DOA) estimation has been associated with algorithms rather than with a processing scheme or architecture. Motivated by a previous work on feasible implementations of the Estimate and Maximize algorithm [1], the authors show that classical bank filter approach [see 2 and its references] can get similar, even better, performance than the most sophisticated algorithms, in terms of performance versus complexity. In fact, the practicality and robustness required for DOA trackers, both in radar and in the mobile communication scenarios to alleviate data fusion and hand-over respectively, makes evident the use of filter-bank or scanning beams for DOA tracking at the expense of resolution. The herein reported tracker enhances complexity and robustness of these schemes, achieving high resolution from the EM architecture. The result is a low complexity tracker with robustness against coherent sources and a resolution close to Singular Value Decomposition (SVD) based methods.

1. Introduction

Motivated by a previous work on feasible implementations of the Estimate and Maximize algorithm [1], the authors show that classical bank filter approach [2] can get similar, even better, performance than the most sophisticated trackers in terms of performance versus complexity. The present summary is organized as follows: Section 2 goes over the scanning beam procedures for DOA estimation and brings in the modifications of interest in this work. Next, Section 3 brings out the EM-based architecture in order for Section 4 to propose a multiple source tracker that uses this architecture together with the beamforming scanning approach briefly described in the previous section. The result is a DOA tracker architecture and algorithm whose robustness and performance is associated to the intrinsic clarity and simplicity of the processing scheme and the DOA algorithms used inside.

* This work has been supported by PRONTIC/CICYT: TIC-95-1022-C05-01 and CIRIT/GENERALITAT de Cat. GRQ 93-3021

II. Scanning beam procedures for Doa estimation

In face of DOA detectors, usually based on SVD of the data matrix or its covariance, the oldest approach, referred to as the bank filter approach, uses a dedicated beam to explore all the scenario looking for the steering directions where a local maximum of received power is produced. As it can be viewed in Figure 1, the DOA estimator is implemented by a steerable beam $\mathbf{a}$ ($\mathbf{s}_d(\theta)$ denotes the steering vector, focused on angle θ, to which the beam is steered) which, followed by a power device (envelope detector plus integration) produces the power density Φ (power/solid angle) for every search direction $\mathbf{s}_d$ (the spatial bandwidth B_N is the noise bandwidth [2]). Finally, the DOA estimate will be the maximum of the spatial power density.

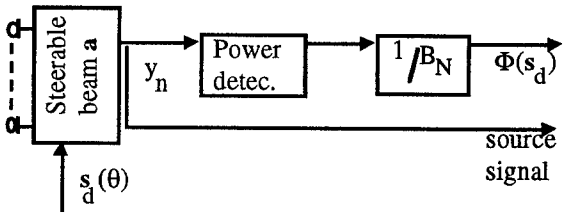


Figure 1. Scanning beam scheme for DOA estimation

From the simplicity of the scheme depicted in figure 1, it can be concluded that complexity and robustness of these procedures are their main features. It is in terms of resolution when the main criticism appears. Any DOA estimation procedure using a beamvector to measure power density has to face the uncertainty principle being the product of the aperture size in wavelenghts by the beamvector bandwidth bounded. An example is the classic phased-array scanning procedure.

The phased-array scanning (1) can be formulated as a beamformer $\mathbf{a}$ with 0 dB gain in the steered direction $\mathbf{s}_d$ and minimizing the response to the non-directional noise (with identity covariance matrix)

$$a^H s_d = 1 \quad (1.a) \qquad a^H s_d = 1 \quad (2.a)$$

$$a^H a|_{min} \quad (1.b) \qquad a^H R\, a|_{min} \quad (2.b)$$

The phased-array response is distorted whenever non-uniform spatial noise or source distributions are to deal with. To alleviate in part the resulting leakage problem the so-called Capon's beamformer is designed in a data dependent fashion. The Capon's beamformer adapts to the spectral content of the input process at each DOA of interest. For each scan direction, it reduces the interference contributions to noise level. Its basic formulation is shown in (2), where **R** is the data correlation matrix measured from the snapshot vector **x**.

Nevertheless, the case of interest of the hereafter tracker is that both approaches, i.e. the data independent or phased array and the data-dependent or Capon's beamvector, are suitable for introducing additional constrains. Specifically, the modification we are interested in is when a given direction s_0 has to be nulled out in order to reduce leakage in any s_d due to the potential presence of an interference source at s_0. The resulting beamvector comes from the following formulation in (3) and (4)

$a^H (s_d\ s_0) = (1\ 0)$ (3a) $a^H (s_d\ s_0) = (1\ 0)$ (4a)

$a^H a|_{min}$ (3b) $a^H R\, a|_{min}$ (4b)

where (3) and (4) depart from the phased-array and the Capon's beamformer philosophy respectively. The corresponding beamvectors are easily derived by means of the Lagrange multipliers. For its simplicity, we pay special attention to the beamvector that is derived from (3) and formulated in (5).

$$a = A\,[A^H A]^{-1}\,(1;0) = A^{\#}\,(1;0) \qquad (5)$$

where $A = [s_d\ s_0]$ and # stands for the pseudo-inverse. This minimum norm beamvector leads to minimum loss of desired signal response if the coefficients a_i are achieved by attenuation and to smallest sensitivity to errors in construction. Additionally, its design is completely data free. It is also interesting to note that if the steered direction s_d is the same as the desired source direction, the beamvector formulated in (5) offers the Deterministic Maximum Likelihood estimate of the signal waveform $\hat{e}$ coming from that source.

$$\hat{e} = A^{\#}\,(1;0)\,x \qquad (6)$$

We recall the importance of the noise bandwidth (B_N) normalization in order to get a reliable DOA estimate from the spatial power density Φ in (3)instead of directly using the spatial power.

$$\Phi = a^H R\, a / B_N \qquad (7)$$

To be more specific, spatial bandwidth may introduce substantial power leakage from sources or directional noise impinging on the aperture from other directions than the desired one.

It is important to remark the robustness and low complexity of both procedures associated with the principle of the beamvector scanning. The only problem they face is spatial frequency leakage or resolution loss for the multiple source case.

Next, we will propose to use either (3) or (4) in a EM based architecture to provide a high resolution tracker yet preserving the low complexity and robustness previously mentioned.

3. The EM-based architecture

After a detailed exam of the EM algorithm [1] both in the deterministic and the stochastic approach, the Estimate step can be viewed as a blocking step where the multiparameter estimation problem is reduced to a single parameter estimation. Being more specific, the steps Estimate and Maximize, when implemented in a signal processing architecture for DOA estimation, can be renamed as blocking and single source estimation respectively. In other words, given the original data snapshot $\mathbf{x}_n$ containing NS sources, the blocking step produces NS snapshots $\mathbf{x}_{n,k}$ (k=1,NS) such that a single source is relevant in every snapshot or, at least, the other sources are highly attenuated with respect to this source.

In consequence, the blocking step could be implemented as NS matrices $\mathbf{B}_k$ (k=1,NS) that produces from $\mathbf{x}_n$ the single source snapshot

$$x_{n,k} = B_k\, x_n \qquad (k=1,NS) \qquad (8)$$

As the blocking step requires the source DOA's, it is necessary to feed the DOA's obtained in the second stage back to the first or blocking stage. This is the other main feature of the EM and EM-based algorithms. As the reader can observe in figure 2, the maximum at the output of each branch governs the nulls of the other branches (i.e. cross-feedback). This fact prevents two or more branches from collapsing into the same angle estimation.

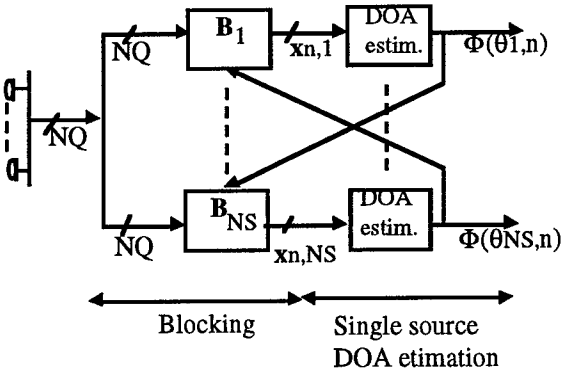


Figure 2. The EM-based architecture with the blocking stage followed by single source DOA detectors

The resulting architecture, depicted in the figure 2, can be found in detail in [3], where the links with the deterministic and stochastic EM algorithm are presented in full. The purpose of this work is to use this architecture together with the beamforming scanning approach briefly described in the previous section.

From now on, the presentation will be reduced to the two source case for a linear array. As the reader may

conclude there is no formal difficulties to extend the application to the case of planar arrays or to the case of multiple sources. Nevertheless, in a radio communication scenario, the probability of more than two users demanding Space Diversity Multiple Access (SDMA) is very low being, in consequence, the case of two active sources the closest to real scenarios.

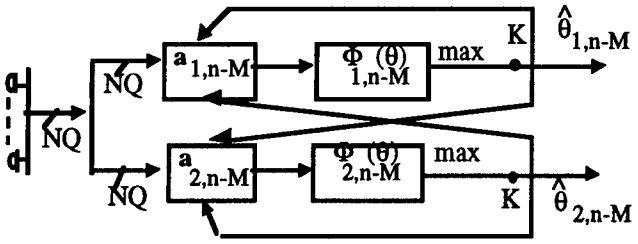


Figure 3. A DOA tracker with beamformer scanning procedures in an EM-based architecture.

Finally, and going back to the architecture of figure 2, it should be mentioned that both steps can be implemented as a single one when the procedures shown in (3) and (4) are used. These procedures allow the packing of both steps in a single one; since, being single source estimates, they include the blocking of DOAs a priori selected. This proposed architecture is the one depicted in figure 3.

Next section will explain how a two source DOA detector/tracker based in this architecture works.

IV. The proposed multiple source tracker.

IV.1. A DOA tracker architecture and algorithm

In the EM-based architecture that is depicted in figure 3, two source DOA's are produced: $\hat{\theta}_{1,n\text{-}M}$ and $\hat{\theta}_{2,n\text{-}M}$. Initially, the data correlation matrix $\mathbf{R}_n$, which is required to compute the spatial power density Φ_i, is initalized with a number of snapshots equal to ten times the number of sensors and afterwards this matrix is updated during M snaphots following the rule:

$$R_{n-M+k} = \beta R_{n-M} + (1-\beta)\, x_{n-M+k}\, x^{H}_{n-M+k} \qquad (9)$$

being M equal to 1/1-β. This interval M is the number of samples between successive updates of the DOAs provided by the system. Its choice is a trade-off between radial source velocity and scanning time.

To make the system robust to bad initializations in whatever kind of scenario (i.e. very different power sources and even presenting strong correlation), initially, just one branch sets out to work. Once, this branch has detected one source DOA, this DOA can then drive the null of the second branch. In this way, both branches cannot collapse into the same source DOA. If both DOA's are far enough, both branches in figure 3 can then begin to scan parallelly. Next, the procedure to update each angle estimate is described.

During the mentioned M snapshots and in the case of the data independent design (see 3), the beamformers are obtained as it is shown in (10) for the beamforemer labelled 1 in figure 3 and in the same manner for the beamformer 2

$$a^{H}_{1n}\,[s_1(\theta)\;\; s_2(\hat{\theta}_{2,n})\,] = (1\;\;0) \qquad (10)$$

$$a^{H}_{1n}\; a_{1n}|_{min}$$

Note that for the beamformer labelled 1, the second branch drives its null at $\hat{\theta}_{2,n}$ and in the same manner for the beamformer 2

Once the beamformers have been designed, they scan on s as:

$$\Phi_{1n}(\theta) = \frac{a^{H}_{1n}\, R_n\, a_{1n}}{a^{H}_{1n}\, a_{1n}} \qquad (11)$$

where the spatial bandwidth B_N has been approximated by the norm of the beamvector $\mathbf{a}_{1n}$ [3]; the new estimate will be the DOA that maximizes the estimated spatial power density

$$\hat{\theta}_{1,n} = max\; \Phi_{1n}(\theta) \qquad (12)$$

We remark that, in order to save in time and computational burden the DOA's that are scanned on s can be close to the previous $\hat{\theta}_{1,n\text{-}M}$. However, the system is no more a detector but just a tracker.

Note that simultaneously to the acquisition period M, the architecture may iterate over (10)-(12) in the same fashion as it was in the original EM algorithm. It should be pointed out that the number of iterations, to be useful, requires a high precision scanning through s and at the end may face the upper bound in resolving two close sources from a M interval data correlation estimate. At least two iterations will be necessary in any case.

IV.2. Tracking subsystem

The concept of a global tracker includes not only the DOA detection scheme, but also the parameter filtering which enables to cope with eventual fadings of bounded time duration, as it may occur in crossing radial trajectories of two targets. This additional processing uses to contain two additional stages of time-trackers of each parameter and data fusion or, in some cases, image processing of the two image produced by parameter values versus time.

Most of the cases, the DOA detection scheme (we just described an alternative in the previous section) does not take profit from the powerful processing that follows in forming the global tracker system. We comment on it to state that the blocking plus estimation scheme described can take a great advantage of the tracker subsystem. We will refer hereafter as parameter tracking, since most of the success of the DOA estimation is based on adequate

nulling or inhibition of the non-steered sources.

Once the detected angle has stabilised at each of the branches, an elevation tracker is used in the scheme of figure 3 (insertion point K). In this way, the performance of the system may improve since tracking and prediction of "next" location is of capital importance in inhibition or blocking.

This work has employed the most basic parameter tracker that can be used: a Kalman filter tracking radial position $\hat{\theta}_{ie,n-M}$ and $\hat{v}_{ie,n-M}$ velocity. A complete description can be found in [4]. Next, we just comment some specific aspects on the state equation and the measurement equation. The state model is

$$\begin{bmatrix} \hat{\theta}_{ie\ n+M} \\ \hat{v}_{ie\ n+M} \end{bmatrix} = \begin{bmatrix} 1 & M \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\theta}_{ie\ n} \\ \hat{v}_{ie\ n} \end{bmatrix} + w_n \qquad (13)$$

where $\mathbf{w}_n$ is the uncertainty (associated to the maneuverability of the sources) with covariance matrix $\mathbf{Q}$. The measurement model is (14)

$$\hat{\theta}_{i\ n} = \theta_{ie\ n} + v_n \qquad (14)$$

where $\hat{\theta}_{i\ n}$ is the estimate produced after the detection and v_n is the noise in the observation of the elevation angle. This noise, of covariance c_{vn}, is due to air-interface, down conversion mismatching, noise and DOA detection errors.

Both covariance $\mathbf{Q}$ and c_v have to be matched to the specific application. In our work we have set Q to diagonal (10^{-4} 10^{-7}) for a mobile communication scenario. We have commented before that, simultaneously to the acquisition period M, the architecture of figure 2 may iterate over (10)-(12). During these L iterations, each of the detectors adjust their DOA estimates until each one stabilizes. The, the Kalman sub-system filters the noise out of these estimates and produces $\hat{\theta}_{ie\ n}$. Therefore, in order to set the measurement covariance c_{vn}, it can be estimated as $\overset{L}{c}_{vn}$ in (15), that is, the error power between the angle predicted by Kalman $\hat{\theta}_{ie\ n}$ and the angle detected $\hat{\theta}_{i\ n}$ over L realizations

$$\overset{k+1}{c}_{vn} = (1-1/L)\ \overset{k}{c}_{vn} + 1/L \left| \overset{k}{\hat{\theta}}_{i\ n} - \hat{\theta}_{ie\ n} \right|^2; \quad k=1...L \qquad (15)$$

V. Simulations and Conclusions

In order to validate the proposed DOA detection/tracking technique two simulations have been conducted. Figure 4 and 5 show the case of two moving sources tracked by the system of figure 3, where the beamformers $\mathbf{a}_i$ are simple phased-arrays that follow the design rule of (10) and where the Kalman sub-system is incorporated. First. figure 4 shows the performance of the system in a scenario of two sources received with very different powers: 15 and 5 dB respectively. The uniform linear array consists of 8 sensors and each scan is carried out after 30 snaphsots. Radial velocity is 0.01°/snapshot and an initial angle estimate of 20° has been considered.

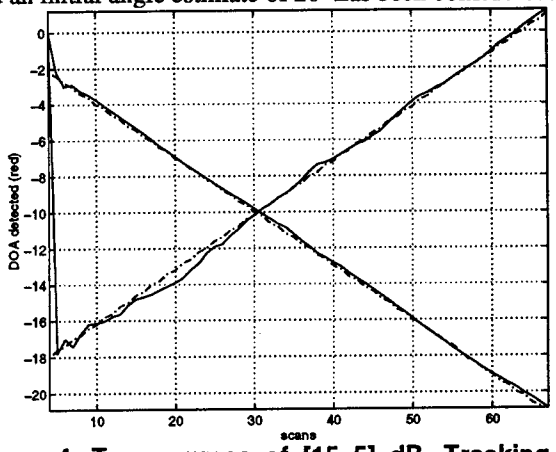


Figure 4. Two sources of [15 5] dB. Tracking by an 8 sensor array. Each scan consists of 30 snaphots.

Next, figure 5 is carried out in the same scenario but with fully coherent sources.

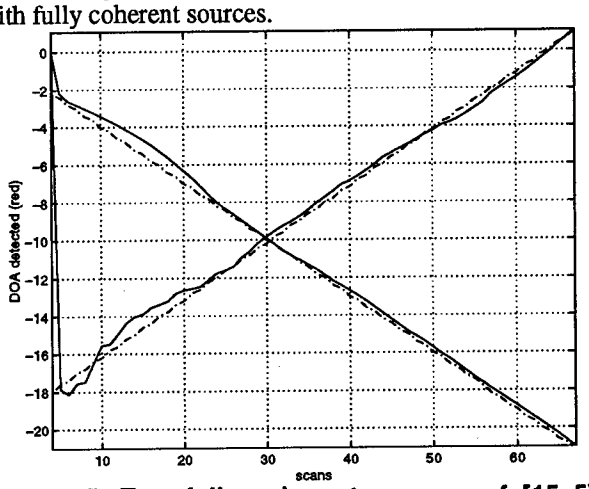


Figure 5. Two fully coherent sources of [15 5] dB. Tracking by an 8 sensor array. Each scan consists of 30 snaphots.

The proposed technique offers a good trade-off between performance against complexity and cost. Its robustness is associated to the intrinsic clarity and simplicity of the processing scheme and the DOA algorithm used inside.

Future work will consider the impact of the deviation in element locations, mutual coupling an quantization effects in using digitally controlled attenuators and shifters.

References

[1] M.Feder, E.Weinstein, "Parameter Estimation of Superimposed Signals Using the EM Algorithm," IEEE Trans. on ASSP, vol. 36, no. 4, April 1988, pp. 477-489

[2] M.A.Lagunas et al., "Maximum Likelihood Filters in Spectral Estimation Problems," Signal Processing 10, pp. 19-34, 1986.

[3] A.Perez-Neira, M.A. Lagunas, L.Kirkin,"Cross-Coupled DOA trackers," submitted to the Special Issue on Advanced Communications of the IEEE S.P. Trans.

[4] B.O.Anderson,J.B.Moore,Optimal filtering, Prentice-Hall, Electrical Engineering Series, pp 54- 59.

Adaptive Array Processing in Non-Gaussian Environments

Christ D. Richmond

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139.

Abstract

In several adaptive array application areas the Gaussian distribution has not proven to be an accurate model of the measured data. Nevertheless, Gaussian based processors have demonstrated robust performance in spite of this statistical mismatch. A need therefore exists for the consideration of (i) problem reformulation and (ii) performance analysis in non-Gaussian environments. The theory of complex multivariate elliptically contoured (MEC) distributions provides an attractive theoretic framework for these considerations especially in the adaptive array setting. We replace the Gaussian data assumption with one of MEC distributed and reexamine the optimality and performance of widely used adaptive detection and beamforming structures.

I. Introduction

IN several radar/sonar and other array application areas the Gaussian distribution is an inadequate model for measured data (see [5] and its bibliography). Thus, extending classical array processing to non-Gaussian distributions has been a longstanding desire of the array community. In addition the observed robustness of Gaussian based processors in some non-Gaussian environments motivates the need for statistical performance analyses of these processors which address in a relatively general sense a plurality of possible non-Gaussian distributions. Such analyses are of special interest in adaptive array scenarios which often involve estimation of the data covariance via the sample covariance matrix (SCM). The theory of complex multivariate elliptically contoured (MEC) distributions provides one such vehicle to perform such analyses. MEC distributions represent a fairly attractive set of data models for the adaptive array scenario for several reasons: (1) similar models have had success historically and contemporarily speaking under the guise of spherically invariant random vectors (SIRVs or processes SIRPs) and Gaussian mixtures [5], (2) MECs provide a theoretic framework which (i) allows one to address a plurality of non-Gaussian distributions simultaneously, and (ii) provides for optimal estimation of the (typically unknown) data covariance, (3) MECs often allow for tractable performance analyses in the presence of the SCM, which has historically been a limiting factor when deviating from the Gaussian assumption.

In this paper we replace the classic assumption of data normality with one of MEC distributed data and reexamine important Gaussian based results of adaptive array detection and signal estimation. In particular, Kelly's generalized likelihood ratio test (GLRT) [1] and Robey's adaptive matched filter (AMF) [2] are shown to be detection structures which arise not necessarily from Gaussianity per se, but rather as by products of the elliptical symmetry which the Gaussian happens to possess. Indeed, it is shown that a large class of MEC distributions lead to the same detection structures. The probability of false alarm (PFA) and constant false alarm rate (CFAR) loss relative to the complex Gaussian of known covariance are shown invariant over the MEC class. Concerning adaptive beamforming, exact statistical analyses of the sample covariance based (SCB) linearly constrained minimum variance (LCMV) beamformer and its SCB generalized sidelobe cancellor (GSC) implementation, which include pdfs for their weightings, beam responses, and beamformer outputs are given. All results suggest significant robustness implications to adaptive array processing in non-Gaussian environments.

II. Array Processing

In array processing the multisensor array data is often modeled by the following vector observation (all vectors are complex)

$$\mathbf{x}_{(N\times 1)} = \mathbf{G}_{(N\times E)}\mathbf{s}_{(E\times 1)} + \mathbf{n}_{(N\times 1)} = \sum_{i=1}^{E} \mathbf{g}_i S_i + \mathbf{n}. \quad (1)$$

The dimensions of the corresponding matrices are indicated in subscript. $\mathbf{x}$ is the received array data with covariance $\mathbf{R}$, containing the desired signal vector

0-8186-7576-4/96 $5.00

$\mathbf{s} = [S_1, S_2, \ldots, S_E]^T$ (T denotes matrix transpose and H the conjugate transpose). $\mathbf{n}$ is additive noise. The columns of the matrix $\mathbf{G} = [\mathbf{g}_1|\mathbf{g}_2|\cdots|\mathbf{g}_E]$ model the system transfer functions, also known as the *steering vector(s)*. It is assumed that $\mathbf{G}$ is full column rank and known exactly.

In this paper we consider the problems of adaptive array detection/estimation and adaptive beamforming in the specific class of MEC non-Gaussian environments.

III. MEC Distributed Data

Consider the $N \times (L+1)$ data matrix $\mathbf{X}_0 = [\mathbf{x}_1|\mathbf{x}_2|\cdots|\mathbf{x}_L|\mathbf{x}] = [\mathbf{X}|\mathbf{x}]$ which has as its first L columns training data and its last column the primary array snapshot under interrogation. Traditionally the columns of $\mathbf{X}_0$ are assumed independent Gaussian; however, the Gaussian distribution is one member of a broad class of distributions known as elliptically contoured (EC) distributions, which likewise often allow for tractable analysis. Multivariate EC distributions extend classical Gaussian based sampling theory of multivariate statistical analysis to the case of observations which are dependent and/or drawn from nonnormal populations [10], [11]. Note the following definitions:

Definition 1: If a complex random vector $\mathbf{h}_{N\times 1}$ has a characteristic function (c.f.) of the form

$$E\{e^{j\mathrm{Re}(\mathbf{t}_0^H\mathbf{h})}\} = \exp[j\mathrm{Re}(\mathbf{t}_0^H\mathbf{m})]\cdot\phi(\mathbf{t}_0^H\mathbf{R}\mathbf{t}_0) \quad (2)$$

where $E\{\cdot\}$ is the expectation, $\mathbf{m}$ is $N\times 1$, $\mathbf{R}$ is $N\times N$ positive semi-definite (≥ 0), $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote real and imaginary parts, we say that $\mathbf{h}$ is complex EC distributed with parameters $\mathbf{m}, \mathbf{R}$, ϕ, and we denote this by

$$\mathbf{h} \sim \mathcal{CEC}_N(\mathbf{m}, \mathbf{R}, \phi). \quad (3)$$

If the density function of $\mathbf{h}$ exists (is nondegenerate) then it necessarily has the form

$$|\mathbf{R}|^{-1}\, g[(\mathbf{h}-\mathbf{m})^H\mathbf{R}^{-1}(\mathbf{h}-\mathbf{m})] \quad (4)$$

where $|\cdot|$ denotes the matrix determinant.

Definition 2: A multivariate EC (MEC) distribution generalizes the vector EC to the case of a matrix, and is similarly defined. An $N\times L$ matrix $\tilde{\mathbf{H}} = [\mathbf{h}_1|\mathbf{h}_2|\cdots|\mathbf{h}_L]$ whose c.f. has the form

$$E\left(\exp\{j\mathrm{Re}[\mathrm{tr}(\mathbf{T}_0^H\tilde{\mathbf{H}})]\}\right) = \exp\left(j\mathrm{Re}[\mathrm{tr}(\mathbf{T}_0^H\mathbf{M})]\right) \quad (5)$$

$$\times\, \phi(\mathbf{t}_1^H\mathbf{R}_1\mathbf{t}_1 + \mathbf{t}_2^H\mathbf{R}_2\mathbf{t}_2 + \cdots + \mathbf{t}_L^H\mathbf{R}_L\mathbf{t}_L) \quad (6)$$

where $\mathrm{tr}(\cdot)$ denotes the matrix trace, $\mathbf{T}_0 = [\mathbf{t}_1|\mathbf{t}_2|\cdots|\mathbf{t}_L]$ is $N\times L$, and $\mathbf{R}_i \geq 0$ for $i = 1, 2, \ldots, L$, is said to be complex MEC and we write

$$\tilde{\mathbf{H}} \sim \mathcal{CMEC}_{N\times L}(\mathbf{M}; \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_L; \phi). \quad (7)$$

If the density of $\tilde{\mathbf{H}}$ exists, then it has the form

$$\left[\prod_{i=1}^{L}|\mathbf{R}_i|^{-1}\right] \times g\left[\sum_{i=1}^{L}\mathrm{tr}\mathbf{R}_i^{-1}(\mathbf{h}_i-\mathbf{m}_i)(\mathbf{h}_i-\mathbf{m}_i)^H\right].$$

Definition 3: When $\mathbf{M} = [\mathbf{m}|\mathbf{m}|\cdots|\mathbf{m}]$, and $\mathbf{R}_1 = \mathbf{R}_2 = \cdots = \mathbf{R}_L = \mathbf{R}$ in definition 2, then, say $\mathbf{H}$, is said to be complex Louiville EC (LEC) distributed and we write

$$\mathbf{H} \sim \mathcal{CLEC}_{N\times L}(\mathbf{m}, \mathbf{R}, \phi). \quad (8)$$

The functional form of $\phi(\cdot)$ uniquely determines $g(\cdot)$ and distinguishes one type of EC/MEC distribution from another. For example, if $\phi(u) = e^{-u}$ then $g(u) \propto e^{-u}$ and $\mathbf{H}$, for example, is a complex Gaussian data matrix with columns independent identically distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{R})$.

IV. Adaptive Array Detection

In adaptive array detection, signal presence is sought in the single vector snapshot $\mathbf{x}$ called the *primary* data vector. One of the two following hypotheses is true:

$$H_0 : \mathbf{x} = \mathbf{n}, \qquad \text{or} \qquad H_1 : \mathbf{x} = \mathbf{G}\mathbf{s} + \mathbf{n}; \quad (9)$$

either the primary data vector is simply noise only (under the null hypothesis H_0), or it contains a target signal plus noise (under hypothesis H_1). It is assumed that $\mathbf{s}$ and $\mathbf{R}$ are unknown and that a *secondary* data set (or training set) $\mathbf{X} = [\mathbf{x}_1|\cdots|\mathbf{x}_L]$ is available to help compensate for ignorance of these nuisance parameters. The target free snapshots $\mathbf{x}_i$ are zero mean and share the same covariance as the primary data vector, *i.e.* $\mathrm{cov}(\mathbf{x}_i) = \mathbf{R}$ for $i = 1, \ldots, L$. The decision about signal presence is based on the totality of the data summarized by the $N\times(L+1)$ data matrix $\mathbf{X}_0 = [\mathbf{X}|\mathbf{x}]$.

A. The GLRT and AMF Detectors

Under both hypotheses and throughout the remainder of this paper we assume that $L \geq N$ and that the data matrix $\mathbf{X}_0$ is MEC distributed with a density of the form

$$g_{H_i} = |\mathbf{R}|^{-(L+1)}g[\,\mathrm{tr}\mathbf{R}^{-1}(\mathbf{X}_0-\mathbf{M}_i)(\mathbf{X}_0-\mathbf{M}_i)^H\,] \quad (10)$$

$i = 0, 1$, where under the corresponding hypotheses we have

$$H_0 : \mathbf{M}_0 = \mathbf{0}_{N\times(L+1)}, \quad H_1 : \mathbf{M}_1 = [\mathbf{0}_{N\times L}|\,\mathbf{G}\mathbf{s}\,]. \quad (11)$$

(Note that this data model includes the complex Gaussian considered by Kelly [1] when we choose $g(r) = e^{-r}\pi^{-N(L+1)}$.) Assuming this data distribution, we follow two theoretic/heuristic approaches to the adaptive array detection problem; namely, (i) that outlined by the GLRT procedure, and (ii) the heuristic AMF approach.

B. Key Results on Detection

- (1) Taking the same GLRT approach as Kelly [1] we find that the resulting GLRT decision statistic is mathematically unchanged under this class of complex MEC distributions.
- (2) Following the heuristic AMF approach [2] likewise leads to the same detection structure one would obtain under the data Gaussianity assumption.
- (3) For both the GLRT and AMF the PFA in invariant over the complex MEC class, *i.e.* it does not depend on $g(\cdot)$. Hence, the CFAR loss relative to the Gaussian is the same.
- (4) For both detectors the PD is dependent on the functional form of the density given by $g(\cdot)$.
- (5) The SCB LCMV beamformer remains the maximum-likelihood (ML) estimate of the signal parameters $\mathbf{s}$. (See [6], [8]).

V. Adaptive Beamforming

A. Clairvoyant Beamformer Weightings

If $\mathbf{R}$ is known exactly, then under the Gaussian assumption the signal estimates and beamformer outputs are respectively of the form

$$\widetilde{\mathbf{s}} = \mathbf{W}^H\mathbf{x} \qquad \text{and} \qquad \widetilde{y} = \mathbf{w}^H\mathbf{x} \qquad (12)$$

[9]. The specific weightings performing the linear transformations on the data $\mathbf{x}$ for the clairvoyant ($\mathbf{R}$ known) array processors considered in this paper are summarized by

Clairvoyant Weightings

$$\begin{aligned}
\text{ML:}\quad & \mathbf{W}_{ML} = \mathbf{R}^{-1}\mathbf{G}(\mathbf{G}^H\mathbf{R}^{-1}\mathbf{G})^{-1} \\
\text{LCMV:}\quad & \mathbf{w}_{LCMV} = \mathbf{W}_{ML}\mathbf{f} \text{ (also MVDR)} \\
\text{GSC:}\quad & \mathbf{w}_{gsc} = \mathbf{W}_{GSC}\mathbf{f} \\
& \mathbf{W}_{GSC} = [\mathbf{I}_N - \mathbf{G}_\perp\mathbf{\Omega}_{GSC}]\,\mathbf{G}(\mathbf{G}^H\mathbf{G})^{-1} \\
& \mathbf{\Omega}_{GSC} = (\mathbf{G}_\perp^H\,\mathbf{R}\,\mathbf{G}_\perp)^{-1}\mathbf{G}_\perp^H\,\mathbf{R}.
\end{aligned}$$

B. SCB Beamformer Weightings

Typically $\mathbf{R}$ is an unknown parameter which must be estimated from a secondary data set. The common heuristic procedure is simply to replace $\mathbf{R}$ with the SCM, which we denote by $\hat{\mathbf{R}}$. This method leads to the following SCB processors:

SCB Weightings

$$\begin{aligned}
\text{ML:}\quad & \widehat{\mathbf{W}}_{ML} = \hat{\mathbf{R}}^{-1}\mathbf{G}(\mathbf{G}^H\hat{\mathbf{R}}^{-1}\mathbf{G})^{-1} \\
\text{LCMV:}\quad & \widehat{\mathbf{w}}_{LCMV} = \widehat{\mathbf{W}}_{ML}\mathbf{f} \text{ (also MVDR)} \\
\text{GSC:}\quad & \widehat{\mathbf{w}}_{gsc} = \widehat{\mathbf{W}}_{GSC}\mathbf{f} \\
& \widehat{\mathbf{W}}_{GSC} = \left[\mathbf{I}_N - \mathbf{G}_\perp\hat{\mathbf{\Omega}}_{GSC}\right]\mathbf{G}(\mathbf{G}^H\mathbf{G})^{-1} \\
& \hat{\mathbf{\Omega}}_{GSC} = (\mathbf{G}_\perp^H\,\hat{\mathbf{R}}\,\mathbf{G}_\perp)^{-1}\mathbf{G}_\perp^H\,\hat{\mathbf{R}}.
\end{aligned}$$

The hat " $\widehat{\ }$ " accent is used to denote the SCM as well as the dependence of the weightings and other quantities on the SCM via the above heuristic procedure. Although originally a heuristic procedure, we show that these SCB beamformers are optimal in the ML sense under the MEC distributed assumption for $\mathbf{X}_0$ given by eq(10) when maximizing over both parameters $\mathbf{s}$ and $\mathbf{R}$ [6].

C. Key Results on Beamforming

C.1 SCB Weightings

The key result is a unified stochastic representation of the SCB weightings. All of these weightings can be written equal in distribution to their clairvoyant counterparts plus a stochastic term:

$$\begin{aligned}
\widehat{\mathbf{W}} &\overset{d}{=} \mathbf{W} - \mathbf{A}\mathcal{T}\mathbf{B} \text{ or} \\
\widehat{\mathbf{w}} &\overset{d}{=} \mathbf{w} - \mathbf{C}\mathcal{T}\mathbf{d}
\end{aligned} \qquad (13)$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{d}$ are deterministic matrices/vector and $\mathcal{T}$ is an $(N-E)\times E$ random matrix with pdf

$$P_{\mathcal{T}}(\mathcal{T}_0) = k\cdot\left|\mathbf{I}_{N-E} + \mathcal{T}_0\mathcal{T}_0^H\right|^{-(L+E)}, \qquad (14)$$

a *standardized* complex multivariate t-distribution. k is the normalizing constant of the pdf. This stochastic representation is completely independent of the functional form of g (or ϕ), and is used to derive exact means, covariances, and pdfs for the beam responses, beamformer outputs, and signal estimates which result from these SCB weightings [7], [8].

C.2 SCB Beam Responses

All SCB beam responses $\hat{\mathbf{b}}(\theta,\omega) = \widehat{\mathbf{W}}^H\mathbf{d}_\theta(\omega)$ or

$\widehat{b}(\theta,\omega) = \widehat{\mathbf{w}}^H \mathbf{d}_\theta(\omega)$, can also be written equal in distribution to their clairvoyant counterparts plus a noise term:

$$\boxed{\begin{array}{rcl} \widehat{\mathbf{b}}(\theta,\omega) & \stackrel{d}{=} & \mathbf{W}^H \mathbf{d}_\theta(\omega) - \mathbf{A}_0\,\mathbf{t} \\ \widehat{b}(\theta,\omega) & \stackrel{d}{=} & \mathbf{w}^H \mathbf{d}_\theta(\omega) - B_0 \times t \end{array}} \qquad (15)$$

where $\mathbf{A}_0$ and B_0 are deterministic and $\mathbf{t}$ $(E \times 1)$ and t are random with pdfs that are special cases of eq(14). This representation is likewise invariant over the class of complex MECs considered.

C.3 Beamformer Outputs

The SCB signal estimates $\widehat{\mathbf{s}} = \widehat{\mathbf{W}}^H \mathbf{x}$ and SCB beamformer outputs $\widehat{y} = \widehat{\mathbf{w}}^H \mathbf{x}$ respectively can be written equal in distribution to:

$$\boxed{\begin{array}{rcl} \widehat{\mathbf{s}} & \stackrel{d}{=} & \mathbf{s} + \mathbf{K}\,\mathbf{z}_g/\sqrt{\beta} \\ \widehat{y} & \stackrel{d}{=} & \mathbf{f}^H \mathbf{s} + K \times Z_g/\sqrt{\beta} \end{array}} \qquad (16)$$

where $\mathbf{K}$ and K are deterministic quantities, $\mathbf{z}_g$ $(E \times 1)$ and Z_g are complex spherically symmetric noise terms whose pdfs depend on the functional form of $g(\cdot)$, and β is complex beta distributed with parameters $L - N + E + 1$ and $N - E$ which is independent of both $\mathbf{Z}_g$ and Z_g [8].

References

[1] E. J. Kelly, "An Adaptive Detection Algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, Vol. AES-22, 115 - 127 (1986).

[2] F. C. Robey, D. R. Furhmann, E. J. Kelly, R. A. Nitzberg, "A CFAR Adaptive Matched Filter Detector," *IEEE Trans. Aerosp. Electron. Syst.* Vol. 28, No. 1, 208–216 (1992).

[3] C. G. Khatri, C. Radhakrishna Rao, "Test for a Specified Signal when the Noise Covariance Matrix is Unknown," *Technical Report 85-47*, Center for Multivariate Analysis, University of Pittsburgh, PA (Nov. 1985).

[4] I. S. Reed, J. D. Mallett, L. E. Brennan, "Rapid Convergence Rate in Adaptive Arrays,' *IEEE Trans. Aerospace and Electronic Systems*, Vol. AES-10, No. 6, 853–863 (1974).

[5] K. J. Sangston, K. R. Gerlach, "Coherent Detection of Radar Targets in a Non-Gaussian Background," *IEEE Trans. Aeros. Electr. Sys.*, Vol. 30, No. 2, 330–340 (1994).

[6] C. D. Richmond, "A Note on Non-Gaussian Adaptive Array Detection and Signal Parameter Estimation," submitted to *IEEE Signal Processing Letters* September 1995.

[7] C. D. Richmond, "PDFs, Confidence Regions, and Relevant Statistics for a Class of Sample Covariance Based Array Processors," to appear *IEEE Trans. Signal Processing* July 1996.

[8] C. D. Richmond, A. O. Steinhardt, A. B. Baggeroer, "Adaptive Array Processing in Non-Gaussian Environments," to be submitted to *IEEE Trans. on Signal Processing.*

[9] S. Haykin, A. O. Steinhardt (Editors), *Adaptive Radar Detection and Estimation*, John Wiley & Sons, Inc. Chapters 3 and 4 (1992).

[10] Fang Kai-Tai and T. W. Anderson (Editors), *Statistical Inference in Elliptically Contoured and Related Distributions*, Allerton Press Inc., New York (1990).

[11] Fang Kai-Tai, Zhang Yao-Ting, *Generalized Multivariate Analysis*, Springer–Verlag, New York (1990).

Chaotic Real-Time Encryption Using Systems of Difference Equations with Large Parameter Spaces

S. Papadimitriou[1] A. Bezerianos[2] T. Bountis[3]

1 Dept. of Comp. Engineering & Informatics, University of Patras, School of Engineering,
26110 Patras, Greece, *tel:*+30-61-997804, *fax:* +30-61-991909, *e-mail:* *pdimitri@grpatvx1.bitnet*
2 Dept. of Medicine, University of Patras, School of Medical Physics,
26500 Patras, Greece, *tel:* +30-61-997620 , *email:* *bezer@upatras.gr*
3 Dept. of Mathematics, Section of Applied Analysis, University of Patras
26110 Patras, Greece, *tel:* +30-61-997381, *fax:* +30-61-997163, *email:* *bountis@math.upatras.gr*

Abstract

We present a new class of discrete chaotic systems (i.e. chaotic maps) that can effectively encrypt information. The nonlinearity of these systems is achieved by designing proper piecewise linear functions and by using modulo operations. The chaotic maps are used as pseudo-noise generators and as the synchronization mechanism of a secure spread-spectrum communication system design. The potential for automatic synchronization, the lack of periodicity and the extremely large parameter spaces that our chaotic maps exhibit offer great advantages over the traditional Linear Feedback Shift Registers pseudo-noise generators for spread spectrum system design.

1 Introduction

The paper presents encryption algorithms that operate by restricting the system's parameters in ranges that guarantee *geometric convergence* of all the variables of the receiving system to those of the transmitting one under the influence of a common transmitted variable (called the *drive* signal). This process has only few features in common with the *chaotic synchronization* process [1,2,5,6,9,10] studied extensively for flows (i.e. systems of differential equations). The similarity is the utilization of one (or more) variable(s), constituting the drive signal, which is used in order to entrain the receiver's system to the transmitter. However, the design of the presented *discrete* chaotic systems is completely different.

The improvements we introduce over previous related work [1,3,6,11] are threefold:

1. We present a methodology based on *general piecewise linear functions that exhibit chaotic dynamics*. These functions are easily implementable with the available electronic technology [4]. They exploit the *modulo operator* in order to achieve both bounded chaotic evolution and extreme sensitivity on the initial conditions. Due to the modulo operation both the range of parameter space over which the evolution is chaotic and the sensitivity of the system to parameter variations are dramatically increased. The consequence is an immense parameter space even for small encryption systems.
2. We present a *general form* for our chaotic enciphering systems and we establish *systematically* a set of *convergence conditions* on the variables of these systems.
3. Although the presented chaotic systems offer very strong encryption security and the possibility to encrypt bulk data (e.g. video data), at fast real-time rates, they are very sensitive to transmission noise. We combine the scheme presented in [11] with our chaotic enciphering systems and obtain the design of a *secure* spread spectrum communication system that can operate *reliably* even in the presence of a strongly noisy background. The designed robust communication systems offer high security, automatic and robust synchronization between the transmitting and receiving spreading sequences and tolerance to intense noise levels. In contrast to the traditional spread spectrum techniques, the security of the system arises mainly from the *inability to synchronize* without the possesion of the encryption key and secondly from the spreading of the spectrum. The chaotic enciphering systems can be implemented efficiently by exploring the parallelism of the computational operations with dedicated array hardware [13].

The paper proceeds by presenting the chaotic encryption/decryption method in Section 2. Section 3 deals with the transmission of digital information over channels with strong noise background. Section 4 discusses the complexity of the encryption method, while in section 5 the conclusions are presented along with directions for future work.

2 The Chaotic Enciphering System

The transmitter encodes the information by implementing the following system of difference equations:

$$
\begin{aligned}
x_1(n+1) &= f_1(x_1(n)) + \varepsilon \cdot s(n) \\
x_2(n+1) &= f_2(x_3(n)) + \sum_{k=1}^{K} a_{2k} x_k(n) + c_2 x_2(n) \sin(d_2 x_3(n)) \\
x_i(n+1) &= f_i(x_{i+1}(n)) + a_{i2} x_2(n) + \sum_{k=i,k\neq 2}^{K} a_{ik} x_k(n) \\
&\quad + c_i x_{i+1}(n) \sin(d_i x_{i+1}(n)), \quad i = 3,\ldots,K-1 \\
x_K(n+1) &= \alpha_{KK} x_K(n) + \alpha_{K2} x_2(n) + c_K x_2(n) \sin(d_K x_2(n))
\end{aligned}
\qquad (1)
$$

where α_{ij}, c_i, d_i , $i = 2,\ldots,K$, $j = 1,\ldots,K$ are constants, $s(n)$ is the information signal and $x_2(n)$ the *drive* signal, i.e. the signal which is transmitted in order to force the synchronization.

The values of these parameters together with the values of the parameters of the functions f_i, $i = 1,\ldots,K-1$ form the *encryption key*. The equation for $x_1(n)$ adds the signal $s(n)$ to the chaotically evolving variable $x_1(n)$.

It is important to stress beforehand that in contrast to the traditional encryption methods the information signal is not transmitted in an encrypted form; rather it is reconstructed by the variables of the *proper* chaotic system at the receiver. The above statement formally means that we cannot express the encoded information **C** as a function of

0-8186-7576-4/96 $5.00

the encryption key $\mathbf{K}$ and the information $\mathbf{I}$, i.e. $\mathbf{C}=\mathbf{F}(\mathbf{K},\mathbf{I})$. Given an information vector $\mathbf{I}$ and an encryption key $\mathbf{K}$, the ciphertext $\mathbf{C}$ can take an *infinity* of possible values (due to the non-periodicity of the chaotic motion).

The parameters of the *piecewise-linear* functions f_i are evaluated modulo R_i i.e. $x_i(n)=x_i(n)\,(\mathrm{mod}\ R_i)+L_i$, where $R_i=U_i-L_i$ and $[L_i,U_i]$ is the domain of definition of the function f_i. The effect of this rule is to limit the evolution of each function within its domain. The detailed expressions of the piecewise linear canonical form and the analysis of the chaotic dynamics of the system (1) is presented in [12,13].

The receiver extracts the information (*information reconstruction*) by implementing the following system which is very similar to (1):

$$\begin{aligned}
&x'_K(n+1)=\alpha_{KK}x'_K(n)+\alpha_{K2}x'_2(n)+c_K x'_2(n)\sin(d_K x'_2(n))\\
&x'_i(n+1)=f_i(x'_{i+1}(n))+a_{i2}x'_2(n)+\sum_{k=i,k\neq 2}^{K}a_{ik}x'_k(n)\\
&\qquad+c_i x'_{i+1}(n)\sin(d_i x'_{i+1}(n)),\quad i=K-1,\ldots,3\\
&x'_1(n)=\frac{(x'_2(n+1)-f_2(x'_3(n))-\sum_{k=2}^{K}a_{2k}x'_k(n)-c_2x'_2(n)\sin(d_2x'_3(n)))}{a_{21}}\\
&r(n)=(x'_1(n+1)-f_1(x'_1(n)))/\varepsilon
\end{aligned}\qquad(2)$$

where $x'_2(n)=x_2(n)$ (since $x_2(n)$ is the transmitted signal that also functions as the *drive* [6] signal) and $r(n)$ is the recovered information signal.

Convergence Conditions We prove that the information signal can be perfectly recovered, when a set of convergence conditions on the parameters is satisfied. Let $x'_2(n)=x_2(n)$ be the common drive signal. We subtract the equations for the Kth variable (x_K and x'_K) of the systems (1) and (2) to get: $\Delta x_K(n+1)=(\alpha_{KK}+c_K\sin(d_K x_2(n)))\cdot\Delta x_K(n)$. Clearly, a sufficient condition for $x'_K\to x_K$ is: $(|\alpha_{KK}|+|c_K|)<1$.

Similarly, the conditions $(|\alpha_{mm}|+|c_m|)<1$ for $m=K-1,K-2,\ldots,3$ can be imposed as sufficient conditions for the convergence of variables $x_{K-1},x_{K-2},\ldots,x_3$ respectively. Finally, by the equations for the drive variable $x_2(n)$ of both systems (1) and (2) we find: $x'_1(n)\to x_1(n)$, $n\to\infty$. Now, it is straightforward to conclude that $r(n)\to s(n)$. Thus, the information is reconstructed at the receiver.

Multiple time-lags The recurrent dependence can be easily extended to $M>1$ time lags. By taking (again) the differences between (1) and (2) we derive the recurrence

$$\Delta x_i(n+1)=\sum_{m=0}^{M-1}a_{iim}\Delta x_i(n-m),\quad i=3,\ldots,K\qquad(3)$$

This recurrence is stable if and only if all the roots $\rho_j, j=0,\ldots M-1$ of its characteristic equation $\rho^{n+1}-\sum_{m=0}^{M-1}a_{iim}\rho^{n-m}=0$, have modulus less than unity. Practically, we design stable (i.e. able to synchronize) systems by selecting M values ρ_i, $i=0,\ldots,M-1$, such that $|\rho_i|<1$. Then we determine the coefficients of the characteristic polynomial with roots ρ_i, $i=0,\ldots,M-1$. These coefficients yield the appropriate values for the parameters a_{ikm}, $m=0,\ldots,M-1$.

3 Noise Tolerance

The presented chaotic encryption systems offer great security levels (encryption complexity is discussed in Section 4). They are, however, very sensitive to distortion by transmission noise. Some schemes for noise robust chaotic modulation have been proposed; notably by a method presented in [11]. These schemes allow reliable transmission of information over channels that exhibit large noise levels (even with negative Signal to Noise Ratio). However, they require *accurate synchronization* between the spreading sequences and, in case synchronization is lost (even temporarily), the communication fails. We apply the presented chaotic enciphering maps - of the type of equation (1) - in order to *automatically* keep intact the synchronization between the chaotic spreading sequences of the receiver and the transmitter. Moreover, the presented design requires the possession of the encryption key for the message retrieval. The extremely large parameter spaces that discrete chaotic enciphering systems exhibit also guarantees high level of security.

The proposed method operates as follows. The transmitter and the receiver implement the chaotic encryption systems (1) and (2) respectively. The variable $x_2(n)$ is used as the *drive* signal which synchronizes the receiving system. The spreading of the information signal for noise tolerance is achieved by using any of the chaotically evolving variables, e.g. $x_1(n)$. The two systems (labeled by S_1 and S_2 in Figure 1) can synchronize and thus generate the same chaotic sequence $x_1(n)$ (i.e $x'_1(n)\to x_1(n)$ where $x'_1(n)$ is the reconstructed signal $x_1(n)$ at the receiver).

We should note here that the synchronization information (i.e. the drive signal) is transmitted *reliably* as the information signal itself. In order to achieve this objective we use multiple (instead of one) chaotic time-series generators for information spreading. These generators are controlled by the main chaotic systems that are capable of achieving synchronization (i.e. S_1 and S_2). Figure 1 illustrates the cluster of K systems used for chaotic spreading (spreading systems) and the main chaotic systems (synchronizing or *entraining* systems). The design operates as follows:

Each sample $x_1(n)$ is fed into each of the K spreading chaotic time series generators. These generators in turn produce $K\cdot N_w$ samples (the parameter N_w controls the window size of the evolution of the spreading systems after the resetting with the value $x_1(n)$ from the entraining system). The generator n, $n=0,\ldots,K-1$, provides to the vector $\mathbf{w}$ the values:

$$w(nN_w+j),\quad j=0,\ldots,N_w-1,$$

where $w(nN_w)=x_1(n)$, for each n. These values (i.e. $w(nN_w+j), j=1,\ldots,N_w-1$) are generated by the evolution of the nth chaotic system initialized to $x_1(n)$. We refer to these generators as the *spreading* systems.

We use many spreading systems instead of one since the exponential divergence of nearby trajectories that the chaotic systems exhibit prevents the use of large parameter N_w (since the very small differences between the entraining systems are enlarged in a few iterations). Moreover, in order to keep intact the pseudo-noise

properties of the spreading sequence we require for the evolution of the spreading systems (in addition to the entraining) to be *chaotic*.

Following the approach of [11] each bit $s(k) \in \{-1,+1\}$ is encoded by using N_s elements of the spreading chaotic time series $w(n)$, $n=(k-1)N_s+1,\ldots,kN_s$. The parameter N_s, called the *spreading parameter*, controls the spreading of the information over the transmitted signal. When the transmitted bit is $s(k)=+1$ we transmit unaltered the N_s values, i.e. $w'(n)=w(n)$, $n=(k-1)N_s+1,\ldots,kN_s$, while if $s(k)=-1$ we transmit $w'(n)=-w(n)$, $n=(k-1)N_s+1,\ldots,kN_s$, i.e. the negatives of the computed values. The drive signal $\mathbf{x}_2$ is converted to a bit stream $\mathbf{x}_{2b}$ and it is transmitted with the same method. The receiver accumulates the reconstructed synchronization bit stream $\mathbf{x}'_{2b}$ in order to build the reconstructed drive signal $\mathbf{x}'_2$. These conversions can be accomplished easily with the use of UARTs (Universal Asynchronous Receiver Transmitter) chips.

In order to proceed with an analysis of the tolerance to noise of this transmission scheme we denote the N-dimensional vectors $\mathbf{w}$ and $\mathbf{N}_v$ as:

$$\mathbf{w}(k)=[w((k-1)N_s+1),\ldots,w(kN_s)],$$
$$\mathbf{N}_v(k)=[v((k-1)N_s+1),\ldots,v(kN_s)]$$

where $v(i)$ denotes the noise component added to the value $w(i)$.

Since noise affects the quality of synchronization, the reconstructed signal $\mathbf{w}_r$ can be expressed as: $\mathbf{w}_r=\mathbf{w}+\Delta\mathbf{w}$, where $\Delta\mathbf{w}$ is the deviation from the synchronization state. The correlation sum $S_N(k)$ can be expanded by introducing the inner products:

$$S_N(k)=\sum_{n=(k-1)N_s+1}^{kN_s} w'(n)\cdot w_r(n)=$$
$$=\sum_{n=(k-1)N_s+1}^{kN_s}(s(k)\cdot w(n)+N_v(n))\cdot(w(n)+\Delta w(n)) \quad (4)$$
$$=s(k)\langle\mathbf{w}(k),\mathbf{w}(k)\rangle+\langle\mathbf{N}_v(k),\mathbf{w}(k)\rangle$$
$$+s(k)\langle\mathbf{w}(k),\Delta\mathbf{w}(k)\rangle+\langle\mathbf{N}_v(k),\Delta\mathbf{w}(k)\rangle$$

As the spreading parameter increases, the probability that the noise vector $\mathbf{N}_v(k)$ has strong components in the direction of $\mathbf{w}(k)$ reduces rapidly. Thus $|\langle\mathbf{N}_v(k),\mathbf{w}(k)\rangle|<<\langle\mathbf{w}(k),\mathbf{w}(k)\rangle$ for large spreading parameter. On the other hand, the inner product $\langle\mathbf{N}_v(k),\mathbf{w}(k)\rangle$, for large Signal to Noise Ratio (SNR), takes much smaller values than $<\mathbf{w}(k),\mathbf{w}(k)>$, but for small SNR it is necessary to use large spreading parameters in order to obtain reliable transmission. The terms $\langle\mathbf{N}_v(k),\Delta\mathbf{w}(k)\rangle$, $\langle\mathbf{w}(k),\Delta\mathbf{w}(k)\rangle$ obey similar rules, since both $\mathbf{N}_v(k)$ and $\Delta\mathbf{w}(k)$ are random vectors. These theroretical conclusions are supported by the simulation results of Figure 2.

Therefore the correlation sum $S_N(k)$ in (4), for sufficiently large spreading parameter, can be approximated by:

$$S_N(k)\approx s(k)\langle\mathbf{w}(k),\mathbf{w}(k)\rangle$$

Since $\langle\mathbf{w}(k),\mathbf{w}(k)\rangle>0$, *the sign of* $S_N(k)$ *determines the transmitted information bit.*

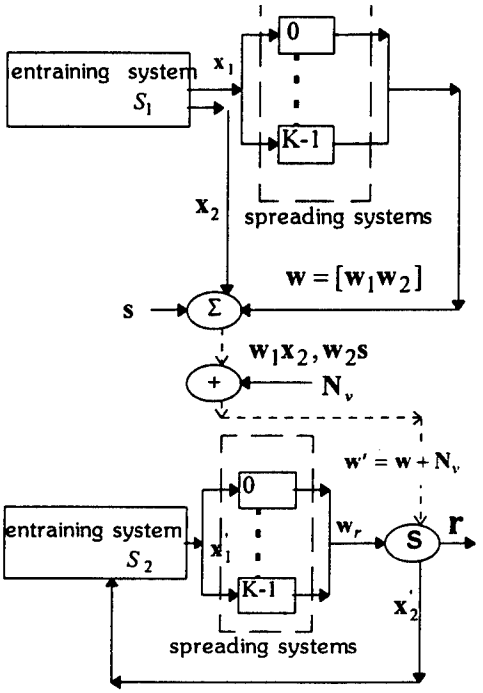


Figure 1 *The architecture of the extended secure communication system*

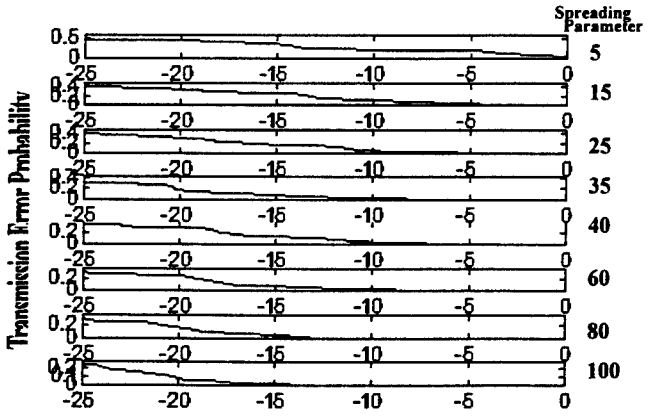


Figure 2 *The probability of transmission error (vertical axis) as a function of the Signal to Noise Ratio (horizontal axis, dB units) for some values of the spreading parameter.*

4 Complexity of the Enciphering Systems

We have already presented the detailed analysis of the fifth-order chaotic system in [13]. Therein we prove that the size of the parameter space is of the order $K_s \cdot 10^{327}$ where K_s is a large constant. By increasing the complexity of the piecewise linear functions, we can increase even further this huge number at the cost of more complex implementation. Generally, for a *[K,M]* chaotic enciphering system (i.e. K variables and M lags), the parameter space size grows approximately with an expression, derived analytically in [13]:

$$O\left(\left(\frac{R_{av}}{S_{av}}\right)^{K^2M}\prod_{i=1}^{N_f}\left(\frac{{D_{f_i}}^2}{\Delta E_i \Delta f_i}\right)^{N_{bi}}\right)$$

where R_{av}, S_{av} are the average range and sensitivity of parameters, N_f is the number of piecewise linear functions, N_{bi} the number of breakpoints of the piecewise linear function f_i, D_{f_i} the domain of definition of f_i, and $\Delta E_i, \Delta f_i$ are the sensitivities over the breakpoint position and breakpoint value respectively. Some typical values for these parameters are:

$$\Delta E_i \approx \Delta f_i \approx 10^{-4}, S_{av} \approx 3.3 \cdot 10^{-8}, R_{av} \approx 2$$
$$N_{b_i} = 20, N_f = K - 1 = 4, M = 3, D_{f_i} = 100$$

With the above values the parameter space size becomes of the order $O(1.1 \cdot 10^{250})$!

It is straightforward to conclude that the parameter space size grows *doubly exponentially* both with the order *[K,M]* of the chaotic system and *exponentially* with the complexity of the piecewise linear functions (note that the exponentation with N_{bi} is multiplied N_f times).

This double exponential complexity should be contrasted with the exponential ones for the more common encryption algorithms in use. Specifically, 2^{56} operations required to break the DES (Data Encryption Standard) algorithm and 2^{128} operations for the newer IDEA block cipher system [8].

The Chaotic Enciphering Systems can be implemented easily with simple hardware (e.g. current mode techniques [4]). The computational complexity of the presented coding algorithms is limited to a few additions and multiplications. Specifically, in order to evaluate a piecewise linear function we have to evaluate only $n+1$ multiplications and $2 \cdot n+2$ additions (where n is the number of breakpoints of these functions). Furthermore, with the exploitation of parallel hardware, $\log_2 n+2$ levels of gate delays are sufficient. Thus besides the flexibility, scalability (with the arbitrary choice of breakpoints, domains and ranges) and simplicity that the piecewise-linear design offers, it also allows faster enciphering/deciphering rates compared with the selection of more complex alternatives (e.g. DES [7], IDEA [8]).

5 Conclusions

In this paper we have presented and analysed chaotic systems of difference equations which display *chaotic* evolution over a wide range of parameter configurations and are amenable for parallel systolic array implementation. These chaotic systems exhibit an immense sensitivity to the parameter configuration. This makes them ideal for application to secure communication systems over *reliable* computer digital networks. In addition we have presented secure transmission schemes (also based on chaotic difference systems) that are capable of transmitting reliably digital information over channels with very low (even negative) Signal to Noise Ratio.

Although the presented cryptosystems provide an effective method for data encryption, they are inefficient for solving the problem of key management. In contrast, public key cryptosystems, such as RSA [7], support an effective scheme for key management. This suggests the use of a hybrid approach exploiting the best of both cryptosystems as the basis for the practical design of a secure communication system. For example, the RSA algorithm may be used for authentication, and the chaotic difference systems for the bulk encryption at very fast rates.

References

[1] K. Sean Halle. Chai Wah Wu, Makoto Itoh, Leon O. Chua, "Spread Spectrum communication through modulation of Chaos", *International Journal on Bifurcation and Chaos,* Vol. 3, No. 2, pp. 469-477, 1993

[2] U. Parlitz, L. O. Chua, Lj. Kocarev. K. S. Halle, A. Shang, "Transmission of digital signals by chaotic synchronization", *International Journal of Bifurcation and Chaos,* Vol. 2, No. 4 (1992) 973-977

[3] K. Murali, M. Lakshmanan, "Synchronizing Chaos in driven Chua's circuit", *International Journal of Bifurcation and Chaos,* Vol. 3, No. 4, pp. 1057-1066, 1993

[4] A. Rodriguiez-Vazquez, M. Delgado-Restituto, "Generation of Chaotic signals using Current-Mode techniques", *Journal of Intelligent and Fuzzy Systems,* Vol. 2, pp. 15-31, 1994

[5] T. L. Carrol, L.M. Pecora, "Cascading synchronized chaotic systems", *Physica D,* 67, pp. 126-140, 1993

[6] Louis M. Pecora and Thomas L. Carrol, "Synchronization of Chaotic Systems", *Physical Review Letters,* Vol. 64, No. 8, pp. 821-823, 1990

[7] Man Young Rhee, *Cryptography and Secure Communications*, McGraw-Hill, 1994

[8] Bruce Schneier, "The IDEA Encryption Algorithm", Dr. Dobb's Journal, December 1993, pp. 50-56

[9] K. M. Cuomo, A. V. Oppenheim, "Circuit implementation of synchronized chaos with applications to communication", *Phys. Rev. Lett. 71,* 1993, pp. 65-68

[10] K. M. Cuomo, A. V. Oppenheim and S. H. Strogatz, "Synchronization of Lorenz-based chaotic circuits with applications to communications", *IEEE Trans. CAS 40,* 1993, pp. 626-633

[11] U. Parlitz and S. Ergezinger, "Robust communication based on chaotic spreading sequences", *Phys. Lett. A 188,* pp. 146-150, 1994

[12] S. Papadimitriou, G. Pavlides, A. Bezerianos, T. Bountis, "Chaotic Systems of Difference Equations for Real-Time Encryption", *1995 IEEE Workshop On Nonlinear Signal and Image Processing*, June 20-22, 1995, pp. 145-149

[13] S. Papadimitriou, T. Bountis, G. Pavlides, "Secure Communication with Chaotic Systems of Difference Equations", accepted for publication to the *IEEE Transactions On Computers*

Statistical Analysis of the Propagator Method for DOA Estimation Without Eigendecomposition

J. Sanchez-Araujo and S. Marcos
Laboratoire des Signaux et Systèmes, CNRS-ESE,
Plateau de Moulon, 91192 Gif-sur-Yvette cedex, France

Abstract

The Propagator method (PM), as well as SWEDE (subspace method without eigendecomposition) [5] and BEWE (bearing estimation without eigendecomposition) [4], [8], belong to a class of subspace-based methods for direction-of-arrival (DOA) estimation which do not require the eigendecomposition of the sample covariance matrix of the received signals and which only use linear operations on the covariance matrix of the received data. These methods can therefore be implemented with a reduced complexity compared to MUSIC. In [1], [2], a method was proposed for estimating the power of sensor noise and the DOA using the PM. The goal of the present paper is to statistically analyse these noise power and DOA estimates.

1 Introduction

Most of the subspace-based methods for DOA estimation require the eigendecomposition of the sample covariance matrix or the singular value decomposition of the data matrix to estimate the signal and/or noise subspaces. Unfortunately, in applications like high resolution passive sonar systems where the number of sensors is large, the use of such methods is unattractive owing to their intensive computational implementation. The PM as well as SWEDE and BEWE, belong to a class of subspace-based methods for DOA estimation which do not need any eigendecomposition and which only use linear operations on the covariance matrix of the sensor outputs. These methods then have a clear potential for real-time applications. The PM uses a linear operator referred to as the "Propagator" which only depends on the steering vectors and which can be easily extracted from the data by a least square process. A non-asymptotical (i.e. finite amount of snapshots) performance analysis of the PM has been reported in [3]. It was found that the PM performs like MUSIC at high and moderate SNR. In [1], [2], a joint estimation of the noise power and the Propagator from the data was proposed. The goal of the present paper is to analyse the asymptotical (large number of snapshots) performance of the PM for estimating the noise power and the DOAs. In section 2 the PM proposed in [1], [2] for the joint estimation of the noise power and the DOAs is briefly recalled. Section 3 is devoted to the asymptotical statistical analysis of the PM and expressions for the variance of the noise power and DOA estimates are established. Section 4 provides numerical examples exhibiting a comparison of the performance of the PM with BEWE, SWEDE and MUSIC. Section 5 concludes the paper.

2 The Propagator Method

Consider an array of M sensors on which K incident narrowband point sources impinge ($M > K$). The observation vector $\boldsymbol{x}$ of the sensor outputs, can be written:

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{x} \in C^{M \text{ x } 1}$ is the noisy data vector, $\boldsymbol{s} \in C^{K \text{ x } 1}$ is the vector of the signal amplitudes, $\boldsymbol{n} \in C^{M \text{ x } 1}$ is an additive noise, and $\boldsymbol{A} = [\boldsymbol{a}(\theta_1), ..., \boldsymbol{a}(\theta_K)] \in C^{M \text{ x } K}$ is the matrix of the steering vectors $\boldsymbol{a}(\theta_i) \in C^{M \text{ x } 1}$ and θ_i, $i = 1, ..., K$ is the direction of arrival of source i, mesured relative to the normal of the array. Under the assumption that the noise is spatially and temporally white, the covariance matrix of $\boldsymbol{x}$ is given by

$$\boldsymbol{R} = E[\boldsymbol{x}\boldsymbol{x}^H] = \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H + \sigma^2\boldsymbol{I}_M \tag{2}$$

where $\boldsymbol{S} = E[\boldsymbol{s}\boldsymbol{s}^H]$ is the signal covariance matrix of dimension K x K assumed to be nonsingular.

The definition of the Propagator relies on the partition of the steering matrix according to

$$\boldsymbol{A}^H = \begin{matrix} \overset{K}{} & \overset{M-K}{} \end{matrix} \begin{bmatrix} \boldsymbol{A}_1^H & \boldsymbol{A}_2^H \end{bmatrix} \tag{3}$$

Under the assumption that $\boldsymbol{A}_1$ is nonsingular, the Propagator is the unique linear operator $\boldsymbol{P} \in C^{K \text{ x } (M-K)}$, equivalently defined by:

$$\boldsymbol{P}^H\boldsymbol{A}_1 = \boldsymbol{A}_2, \quad \text{or} \quad \boldsymbol{A}^H \begin{bmatrix} \boldsymbol{P} \\ -\boldsymbol{I} \end{bmatrix} = \boldsymbol{A}^H\boldsymbol{Q} = 0 \tag{4}$$

0-8186-7576-4/96 $5.00

It follows that matrix $\boldsymbol{Q}$ spans the nullspace of $\boldsymbol{A}^H$. This matrix can be estimated from the sample covariance matrix:

$$\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{x}(t)\boldsymbol{x}^H(t) \tag{5}$$

where N is the number of snapshots. In [1], [2], a method was proposed for estimating both the noise power σ^2 and the Propagator $\boldsymbol{P}$ from the sample covariance matrix (5). This relies on the following. Consider the modified covariance matrix

$$\boldsymbol{R} - \delta \boldsymbol{I}_M = [\boldsymbol{G}(\delta), \boldsymbol{H}(\delta)] \tag{6}$$

where δ is positive and $\boldsymbol{G}(\delta)$ and $\boldsymbol{H}(\delta)$ are matrices of dimension M x K and M x $(M-K)$, respectively. The following proposition has been proved in [1].

Proposition: Assuming that the $(M-K)$ x K matrix $\boldsymbol{A}_2$ in (3) is of rank K, $(\widehat{\boldsymbol{P}} = \boldsymbol{P}, \delta = \sigma^2)$ is the unique solution of

$$\boldsymbol{G}(\delta) = \boldsymbol{H}(\delta)\widehat{\boldsymbol{P}} \tag{7}$$

if, and only if, $M - K > K$.

Consider now the following partition

$$\boldsymbol{R} = \begin{array}{c} \begin{matrix} \scriptstyle K & \scriptstyle M-K \end{matrix} \\ \begin{bmatrix} \boldsymbol{G}_1 & \boldsymbol{H}_1 \\ \boldsymbol{G}_2 & \boldsymbol{H}_2 \end{bmatrix} \end{array} \begin{matrix} \\ \scriptstyle K \\ \scriptstyle M-K \end{matrix} \tag{8}$$

According to (2), matrices $\boldsymbol{G}_2$ and $\boldsymbol{H}_2$ satisfy:

$$\begin{aligned} \boldsymbol{G}_2 &= \boldsymbol{A}_2 \boldsymbol{S} \boldsymbol{A}_1^H \\ \boldsymbol{H}_2 &= \boldsymbol{A}_2 \boldsymbol{S} \boldsymbol{A}_1^H + \sigma^2 \boldsymbol{I}_{M-K} \end{aligned} \tag{9}$$

It can easily be seen that

$$\sigma^2 = \frac{\mathrm{tr}\{\boldsymbol{H}_2 \boldsymbol{\Pi}\}}{\mathrm{tr}\{\boldsymbol{\Pi}\}} \tag{10}$$

where $\boldsymbol{\Pi} = \boldsymbol{I}_{M-K} - \boldsymbol{G}_2\boldsymbol{G}_2^\dagger = \boldsymbol{I}_{M-K} - \boldsymbol{A}_2\boldsymbol{A}_2^\dagger$ and where $\mathrm{tr}\{.\}$ denotes the trace operator and $(.)^\dagger$ is the Moore-Penrose pseudo-inverse. Then, a possible estimate of σ^2 can be obtained by (see also [2])

$$\hat{\sigma}^2 = \frac{\mathrm{tr}\{\widehat{\boldsymbol{H}}_2 \widehat{\boldsymbol{\Pi}}\}}{\mathrm{tr}\{\widehat{\boldsymbol{\Pi}}\}} \tag{11}$$

where $\widehat{\boldsymbol{H}}_2$ and $\widehat{\boldsymbol{\Pi}}$ are estimates of $\boldsymbol{H}_2$ and $\boldsymbol{\Pi}$, respectively, and where

$$\widehat{\boldsymbol{H}}_2 = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{x}_2(t)\boldsymbol{x}_2^H(t), \quad \widehat{\boldsymbol{G}}_2 = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{x}_2(t)\boldsymbol{x}_1^H(t) \tag{12}$$

where $\boldsymbol{x}^H = \begin{matrix} \scriptstyle K & \scriptstyle M-K \\ [\boldsymbol{x}_1^H & \boldsymbol{x}_2^H] \end{matrix}$

3 Statistical Analysis

According to the central limit theorem, it can be checked that $\widehat{\boldsymbol{R}} - \boldsymbol{R} = O\left(\frac{1}{\sqrt{N}}\right)$ which means that $\sqrt{N}(\widehat{\boldsymbol{R}} - \boldsymbol{R})$ is bounded in probability when $N \to \infty$. It then follows from (5) and (12) that $\widehat{\boldsymbol{H}}_2 - \boldsymbol{H}_2 = O\left(\frac{1}{\sqrt{N}}\right)$ and that $\widehat{\boldsymbol{G}}_2 - \boldsymbol{G}_2 = O\left(\frac{1}{\sqrt{N}}\right)$. With a first order expansion of $\widehat{\boldsymbol{H}}_2$ and $\widehat{\boldsymbol{G}}_2$ we easily derive [9] that the estimator (11) provides a consistent estimate $\hat{\sigma}^2$ of σ^2. Similarly, a consistent estimate $\widehat{\boldsymbol{P}}$ of $\boldsymbol{P}$ can be obtained from $\widehat{\boldsymbol{R}} - \hat{\sigma}^2\boldsymbol{I}_M$ according to (7) as

$$\widehat{\boldsymbol{P}} = \widehat{\boldsymbol{G}}^\dagger \widehat{\boldsymbol{H}} \tag{13}$$

We now derive large sample variance expressions for the noise power estimate (11) and the DOA estimates which minimize

$$f(\theta) = \boldsymbol{a}^H(\theta)\boldsymbol{\Pi}_{\hat{\boldsymbol{Q}}}\boldsymbol{a}(\theta) \tag{14}$$

where $\boldsymbol{\Pi}_{\hat{\boldsymbol{Q}}} = \hat{\boldsymbol{Q}}\hat{\boldsymbol{Q}}^\dagger$ is the orthogonal projector onto the noise subspace.

Theorem 1: The large sample (for $N >> 1$) variance of $\hat{\sigma}^2$ (11), is given by

$$E\left[(\hat{\sigma}^2 - \sigma^2)^2\right] = \frac{\sigma^4}{N(M-2K)} \tag{15}$$

Proof: See the Appendix.

Now from a first-order approximation of the first derivative $f'(\hat{\theta}_i)$ of $f(\theta)$ around the DOA estimates $\hat{\theta}_i$, and after a first order expansion of $\boldsymbol{\Pi}_{\hat{\boldsymbol{Q}}}$ it can be shown that

$$\hat{\theta}_i - \theta_i \simeq -\frac{\mathrm{Re}\{\boldsymbol{d}_i^H \boldsymbol{\Pi}_{\tilde{\boldsymbol{Q}}} \boldsymbol{a}_i\}}{\boldsymbol{d}_i^H \boldsymbol{\Pi}_{\boldsymbol{Q}} \boldsymbol{d}_i} \tag{16}$$

where $\boldsymbol{\Pi}_{\tilde{\boldsymbol{Q}}} \simeq \boldsymbol{Q}(\boldsymbol{Q}^H\boldsymbol{Q})^{-1}\tilde{\boldsymbol{Q}}^H$, $\boldsymbol{d}_i = \dfrac{\partial \boldsymbol{a}(\theta_i)}{\partial \theta_i}$, $\boldsymbol{a}_i = \boldsymbol{a}(\theta_i)$ and where

$$\tilde{\boldsymbol{Q}} = \begin{bmatrix} \tilde{\boldsymbol{P}} \\ \boldsymbol{0} \end{bmatrix}, \qquad \tilde{\boldsymbol{P}} = (\boldsymbol{G}^H\boldsymbol{G})^{-1}\boldsymbol{G}^H[\widehat{\boldsymbol{H}} - \widehat{\boldsymbol{G}}\boldsymbol{P}] \tag{17}$$

In expressions (17), $\tilde{\boldsymbol{Q}}$ and $\tilde{\boldsymbol{P}}$ are the estimation error matrices of $\boldsymbol{Q}$ and $\boldsymbol{P}$, respectively. We now state and prove the following result.

Theorem 2: Let $\{\hat{\theta}_i\}$ be the DOAs estimated by the PM. The asymptotical variance (for $N \to \infty$) of $\hat{\theta}_i$ is given by

$$E[(\hat{\theta}_i - \theta_i)^2] = \frac{\sigma^2}{2N\gamma_i}\left[(\boldsymbol{S}^{-1})_{ii} + \sigma^2(\boldsymbol{S}^{-1}(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{S}^{-1})_{ii}\right] \tag{18}$$

where $(.)_{ii}$ denotes the i-element in the diagonal of the bracketed matrix and $\gamma_i = \boldsymbol{d}_i^H \boldsymbol{\Pi}_Q \boldsymbol{d}_i$. Note that this expression is identical to that obtained for MUSIC [6].

Proof: See the Appendix.

4 Simulations

In this section we present some numerical examples to lend support to the theoretical results here obtained. In all the examples, we consider two uncorrelated sources of equal power impinging on a uniform linear array of equispaced sensors.

Example 1: In this example we illustrate the performance of the proposed method for estimating the noise power. Consider $\theta_1 = 0^o$ and $\theta_2 = 20^o$. First, the number of sensors M varies from 5 to 40 with SNR=0 dB. Next, the SNR varies from -2 to 12 dB with $M = 20$. In all the cases $N = 400$ snapshots were used for each of the 100 independent trials. Figure 1 displays the empirical root mean square errors (RMSE) and the theoretical standard deviations calculated using expressions (11) and (15), respectively. In the same figure, the performance of the proposed noise power estimator is compared with another 'linear method' (LM) and a so-called 'eigenvalue method' (EM) proposed and statistically analysed in [7]. It can be seen that the RMSE of the proposed noise power estimator and that of the EM are comparable for large values of M. Note that, unlike the proposed method, the variance of the LM does not decrease as M increases.

Example 2: We compare in this example the performance of the PM with SWEDE (G), BEWE and MUSIC for estimating the DOAs: $\theta_1 = 5^o$ and $\theta_2 = 15^o$ from 50 snapshots. First, the number of sensors M varies from 10 to 45 with SNR=0 dB and then, the SNR varies from -4 to 10 dB with $M = 20$. In both cases the empirical RMSE is based on 200 independent trials. The empirical RMSE and the theoretical st. dev. of the estimates of the source $\theta_1 = 5^o$ are exhibited in Figure 2. This figure verifies our theoretical result, i.e., for a spatially and temporally white noise model, the PM variance is equal to that of MUSIC. Consequently, this version of the PM performs better than SWEDE (G) and BEWE.

5 Concluding remarks

The purpose of this work was to statistically analyse the performance of the PM for DOA estimation when the noise power is estimated and removed from the sample covariance matrix of the sensor outputs according to the method proposed in [1], [2]. It has been shown that this version of the PM asymptotically performs like MUSIC in spatially white noise scenarios. While the eigendecomposition of a M x M matrix requires $O(M^3)$ operations, the implementation of the PM is of order $O(M^2)$ which is more interesting from the computational point of view.

Appendix

Proof of theorem 1. First-order approximations for $\widehat{\boldsymbol{H}}_2\widehat{\boldsymbol{\Pi}}$ and $\widehat{\boldsymbol{\Pi}}$ which appear in (11) and the use of $\boldsymbol{\Pi}\boldsymbol{G}_2 = \boldsymbol{0}$, allow us to write

$$\hat{\sigma}^2 = \mathrm{Re}\left\{\frac{\alpha + \sigma^2\beta}{\eta}\right\} \tag{A.1}$$

where

$$\alpha = \mathrm{tr}\{\widehat{\boldsymbol{H}}_2\boldsymbol{\Pi} - \boldsymbol{H}_2\boldsymbol{G}_2^{\dagger H}\widehat{\boldsymbol{G}}_2^H\boldsymbol{\Pi} - \boldsymbol{H}_2\boldsymbol{\Pi}\widehat{\boldsymbol{G}}_2\boldsymbol{G}_2^{\dagger}\},$$

$$\beta = \mathrm{tr}\{\boldsymbol{G}_2^{\dagger H}\widehat{\boldsymbol{G}}_2^H\boldsymbol{\Pi} + \boldsymbol{\Pi}\widehat{\boldsymbol{G}}_2\boldsymbol{G}_2^{\dagger}\}, \quad \text{and} \quad \eta = \mathrm{tr}\{\boldsymbol{\Pi}\} \tag{A.2}$$

By noting that $\boldsymbol{H}_2\boldsymbol{\Pi} = \sigma^2\boldsymbol{\Pi}$ and $\boldsymbol{G}_2^{\dagger}\boldsymbol{\Pi} = \boldsymbol{\Pi}\boldsymbol{G}_2^{\dagger H} = \boldsymbol{0}$ and by substituting (12) into (A.2) it can easily be verified that $\beta = 0$ and

$$\alpha = \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{x}^H(t)\boldsymbol{M}\boldsymbol{x}(t), \qquad \boldsymbol{M} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Pi} \end{bmatrix} \tag{A.3}$$

After a straightforward derivation we obtain

$$E[(\hat{\sigma}^2 - \sigma^2)^2] = E\left[\frac{\alpha^2}{\eta^2}\right] - \sigma^4 \tag{A.4}$$

Using the formula for the expectation of four random matrices proposed in [10], we obtain:

$$E[\alpha^2] = \mathrm{tr}\{\boldsymbol{M}\boldsymbol{R}\}^2 + \frac{1}{N}\mathrm{tr}\{(\boldsymbol{M}\boldsymbol{R})^2\} \tag{A.5}$$

It follows from expressions (A.3), (8), and (9) that

$$E[(\hat{\sigma}^2 - \sigma^2)^2] = \frac{\sigma^4}{N\mathrm{tr}\{\boldsymbol{\Pi}\}} \tag{A.6}$$

Finally, note that $\mathrm{tr}\{\boldsymbol{\Pi}\} = (M - K) - \mathrm{tr}\{\boldsymbol{A}_2\boldsymbol{A}_2^{\dagger}\}$. Under the assumptions that $\boldsymbol{A}_2\boldsymbol{A}_2^{\dagger}$ has full rank, equal to K, and that $M - K > K$, it can easily be shown that $\boldsymbol{A}_2\boldsymbol{A}_2^{\dagger}$ has $M - 2K$ eigenvalues which are equal to 0 and that the K remaining eigenvalues are equal to 1. Hence $\mathrm{tr}\{\boldsymbol{\Pi}\} = M - 2K$.

Proof of theorem 2. Note from (16) that

$$E[(\hat{\theta}_i - \theta_i)^2] = \frac{E\left[(\mathrm{Re}\{\rho\})^2\right]}{\gamma_i^2} \tag{A.7}$$

where $\rho = -\boldsymbol{\alpha}_i^H\widetilde{\boldsymbol{P}}^H\boldsymbol{a}_{1,i}$ with $\boldsymbol{a}_{1,i}$ being the i-column of $\boldsymbol{A}_1$ and $\boldsymbol{\alpha}_i = \boldsymbol{Q}^{\dagger}\boldsymbol{d}_i$. Then, it follows from (5) and (6)

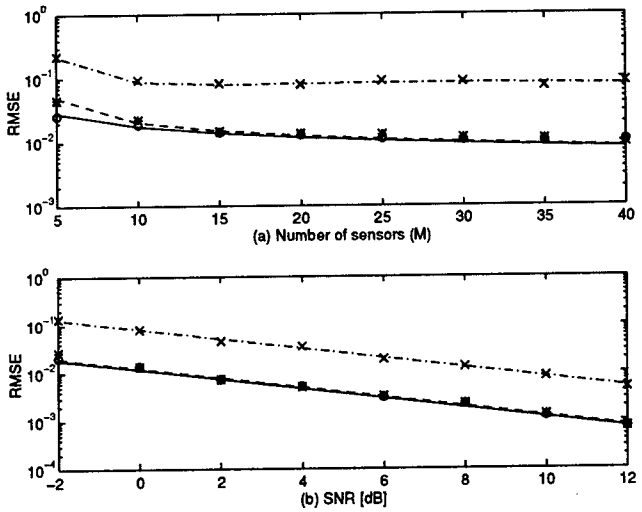


Figure 1. Empirical RMSE and theoretical st. dev. of the noise power estimates versus (a) Number of sensors, (b) SNR. The proposed method (*)-(- -), the LM (x)-(-.), and the EM (o)-(–), (empirical RMSE)-(theoretical st. dev.).

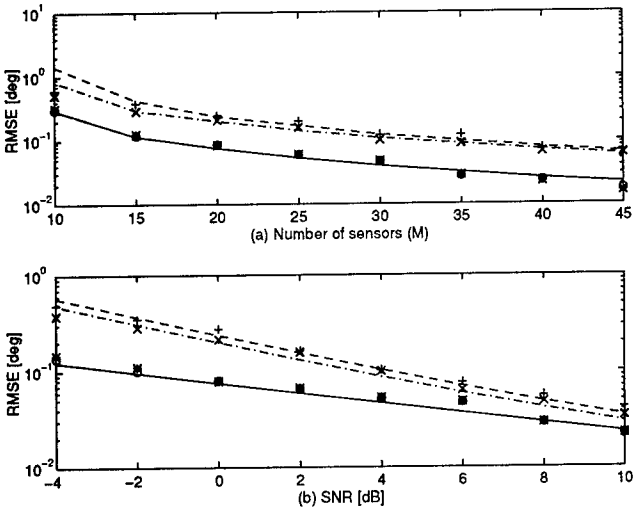


Figure 2. Empirical RMSE and theoretical st. dev. of the estimate of $\theta_1 = 5^o$ versus (a) Number of sensors, (b) SNR. The PM (*)-(–), SWEDE (G) (x)-(-.), BEWE (+)-(- -), and MUSIC (o)-(–), (empirical RMSE)-(theoretical st. dev.).

that $\widehat{G} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{x}(t)\boldsymbol{x}_1^H(t)$, $\widehat{H} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{x}(t)\boldsymbol{x}_2^H(t)$ and ρ becomes

$$\rho = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{x}^H(t)\bar{\boldsymbol{M}}\boldsymbol{Q}^H\boldsymbol{x}(t) \qquad \text{(A.8)}$$

where matrix $\bar{\boldsymbol{M}} = \boldsymbol{G}(\boldsymbol{G}^H\boldsymbol{G})^{-1}\boldsymbol{a}_{1,i}\boldsymbol{\alpha}_i^H$ is of dimension M x $(M-K)$. By inserting (A.8) in (A.7), using once again the result of [10] and checking that $\boldsymbol{Q}\bar{\boldsymbol{M}}^H\boldsymbol{R} = \boldsymbol{0}$, $\bar{\boldsymbol{M}}\boldsymbol{Q}^H\boldsymbol{R} = \boldsymbol{0}$, we find

$$E[(\hat{\theta}_i - \theta_i)^2] = \frac{1}{2N\gamma_i^2}\left[\mathrm{tr}\{\bar{\boldsymbol{M}}\boldsymbol{Q}^H\boldsymbol{R}\boldsymbol{Q}\bar{\boldsymbol{M}}^H\boldsymbol{R}\}\right] \qquad \text{(A.9)}$$

It can easily be verified that

$$\mathrm{tr}\{\bar{\boldsymbol{M}}\boldsymbol{Q}^H\boldsymbol{R}\boldsymbol{Q}\bar{\boldsymbol{M}}^H\boldsymbol{R}\} = \sigma^2\gamma_i\left[\boldsymbol{S}^{-1} + \sigma^2\boldsymbol{S}^{-1}(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{S}^{-1}\right]_{ii} \qquad \text{(A.10)}$$

Hence (18).

References

[1] S. Marcos and M. Benidir, "On a high resolution array processing method non-based on the eigenanalysis approach," *Proc. ICASSP*, pp. 2955-2958, Albuquerque, NM, Apr. 1990.

[2] S. Marcos and M. Benidir, "Source bearing estimation and sensor positioning with the propagator method," in: F.T. Luk, ed., *Adv. Signal Processing Alg., Arch., and Implementations.*, SPIE, vol. 1348, pp. 312-323, San Diego, Jul. 1990.

[3] S. Marcos, A. Marsal, and M. Benidir, "The Propagator method for source bearing estimation," *Signal Processing*, vol. 42, no. 2, pp. 121-138, Mar. 1995.

[4] C. C. Yeh, "Simple computation of projection matrix for bearing estimations," *Proc. Inst. Elec. Eng.*, Part F, vol. 134, no. 2, pp. 146-150, Apr. 1987.

[5] A. Eriksson, P. Stoica, and T. Söderström, "On-line subspace algorithms for tracking moving sources," *IEEE Trans. Signal Processing*, vol. 42, no. 9, pp. 2319-2330, Sep. 1994.

[6] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 38, no. 12, pp. 2140-2150, Dec. 1990.

[7] P. Stoica, and T. Söderström, and V. Symonyte, "On estimating the noise power in array processing," *Signal Processing*, vol. 26, no. 2, pp. 205-220, Feb. 1992.

[8] P. Stoica and T. Söderström, "Statistical analysis of a subspace method for bearing estimation without eigendecomposition," *Proc. Inst. Elec. Eng.*, Part F, vol. 139, no. 4, pp. 301-305, Aug. 1992.

[9] S. Marcos and J. Sanchez, "Méthodes 'linéaires' haute résolution pour l'estimation de directions d'arrivée de sources. Performances asymptotiques et complexité," submitted to *Traitement du Signal*, 1996.

[10] P.H.M. Janssen and P. Stoica, "On the expectation of the product of four matrix–valued gaussian random variables," *IEEE Trans. on Automatic Control*, vol. AC-33, no. 9, pp. 867-870, Sep. 1988.

Study of Ambiguities in Array Manifold: A General Framework

A. FLIELLER[1], P. LARZABAL[1] and H. CLERGEOT[2]

[1] L.E.Si.R. E.N.S. Cachan 61, Av. du président Wilson 94235 CACHAN Cedex FRANCE
U.R.A. C.N.R.S. D1375 Tel . 33 1 47 40 27 09 Fax 33 1 47 40 27 08 larzabal@lesir.ens-cachan.fr
[2] L.T.S.M.M. Faculté de technologie de Guyane, Avenue d'estrées, BP 792, 97337 CAYENNE Cedex, Guyane Française

Abstract

In the performance evaluation of sources localisation methods, resolution is not the only criterion. Degradations may occur, due to parasite peaks in the spectrum, which may be connected to high sidelobes in the beam pattern or to ambiguities themselves. The aim of this paper is to study the presence of ambiguities in an array of given planar geometry. The ambiguity problem for an arbitrary array is examined and ambiguous situations are identified. We propose a general framework for the analysis and so we obtain a generalisation of results given in recent publications [2], [3] for rank one and two ambiguities. For rank $k \geq 3$ ambiguities, results focus on linear arrays, for which we derive original and synthetic results. Some interesting results are driven for non uniform linear arrays, including sparse linear arrays [4].

1. Introduction

In the performance evaluation of sources localisation methods, resolution is not the only criterion. Degradations may occur, due to parasite peaks in the spectrum, which may be connected to high sidelobes in the beam pattern (sometime referred as quasi-ambiguities) or to ambiguities themselves. These ambiguities arise when the array manifold intersects itself or when a manifold vector can be written as a linear combination of two or more manifold vectors [1]. The aim of this paper is to study the presence of ambiguities in array geometry.

The ambiguity problem for an arbitrary array will be examined and ambiguous situations will be identified. We propose a general framework for the analysis and so we obtain a generalisation of results given in recent publications [2], [3].

In section 2 notations and definitions of ambiguity are introduced. In section 3 a study of rank one ambiguous arrays is presented. Section 4 depicts the main results obtained for rank two ambiguous arrays. This study is made for arrays of arbitrary geometry. In section 5, the study is restricted to linear arrays for rank three ambiguities. In section 6 we derive original and synthetic results for rank $k \geq 3$ ambiguities restricted to linear arrays. Some interesting conclusion may be driven for non uniform linear arrays, including sparse linear arrays [4]. Section 7 includes some conclusions.

2. Problem formulation and definitions

Consider an array with M sensors receiving N narrowband signals impinging on the array from N different locations $\theta_1,...,\theta_N$. Note $A(\theta_1,...,\theta_N)=\left[\vec{a}(\theta_1),...,\vec{a}(\theta_N)\right]$, the matrix with columns the sources steering vectors called also the array manifold vectors.

The simultaneous localisation of N sources is only possible if the array manifold vectors $\vec{a}(\theta_1),...,\vec{a}(\theta_N)$ are linearly independent.

An array is said rank k ambiguous for a set of $k+1$ directions of arrival $\theta_1,...,\theta_{k+1}$ if matrix A is singular but rank k. This can be written:

$$\exists \alpha_1 \neq 0,...,\alpha_{k+1} \neq 0 \text{ so that } \alpha_1\vec{a}(\theta_1)+...+\alpha_{k+1}\vec{a}(\theta_{k+1})=\vec{0}$$
$$(\alpha_1,...,\alpha_{k+1}) \in C^{k+1} \quad (1)$$

3. Rank one ambiguities (for general arrays)

This case occurs when one array manifold vector $\vec{a}(\theta_1)$ can be written as a complex scalar multiple of another manifold vector $\vec{a}(\theta_2)$ where $\theta_1 \neq \theta_2$.

$$\exists(\alpha_1 \neq 0, \alpha_2 \neq 0) \in C^2, \text{ so that } \alpha_1\vec{a}(\theta_1)+\alpha_2\vec{a}(\theta_2)=\vec{0} \quad (2)$$

0-8186-7576-4/96 $5.00

In such case, the array cannot make the difference between two waves with bearings θ_1 or θ_2.
The wavefronts are supposed straight-line and on the same plane as the sensors. $\vec{k}_1$ and $\vec{k}_2$ being the ambiguous wave vectors for the array under consideration, the phase delay of signal n from sensor m to sensor one is :

$$\varphi_{mn} = \vec{k}_n . \vec{r}_m \qquad (3)$$

where $\vec{r}_m$ denotes the position of the m^{th} sensor in half wavelength. Equation (2) is then equivalent to the condition :

$$\alpha_1 e^{-j\varphi_{m1}} + \alpha_2 e^{-j\varphi_{m2}} = 0 \Leftrightarrow \exists n_m, \varphi_{m1} = \varphi_{m2} + 2n_m\pi \quad (4)$$
$$\text{for } m = 1,\ldots,M$$

The ambiguity condition can be written :

$$\exists p_m, \text{ integer } (\vec{k}_1 - \vec{k}_2)\vec{r}_m = 2p_m\pi \qquad (5)$$

with $|\vec{k}| = 2\pi/\lambda$ where λ stands for the wavelength. It can be given the following geometrical interpretation :

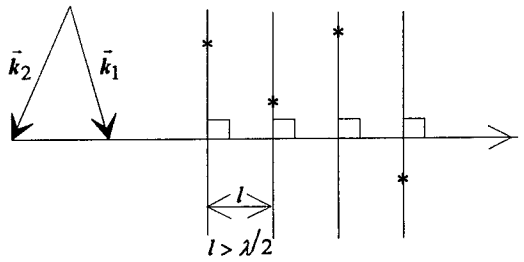


Fig. 1 : Stars represent some possible sensor positions for a rank one ambiguous array. The horizontal axis is defined by the vectors $\vec{k}_1$ and $\vec{k}_2$.

The consequence is that, for arrays of arbitrary geometry, rank 1 ambiguities can arise if all of its sensors are located on a set of parallel lines separated by a distance $l > \lambda/2$. In the case of a linear array this result refunds the classical Shannon condition. In the general case, it establishes conditions for ambiguity and then can give the ambiguous directions [5].

4. Rank two ambiguities (For general planar arrays)

This situation occurs when the array manifold line intersects a plane in more than two points. In such case, one manifold vector can be written as a linear combination of two others manifold vectors, which may be written:

$$\exists(\alpha_1,\alpha_2,\alpha_3) \in C^3 \ \alpha_1\vec{a}(\theta_1) + \alpha_2\vec{a}(\theta_2) + \alpha_3\vec{a}(\theta_3) = \vec{0}$$
$$(\alpha_1 = 1) \qquad (6)$$

with $\vec{a}(\theta_n) = \left[\ldots \quad e^{-j\varphi_{mn}} \quad \ldots\right]^T$ and $\varphi_{mn} = \vec{k}_n . \vec{r}_m$.
Sensor 1 is taken as a reference $\vec{r}_1 = \vec{0}$. Therefore for sensor 1, $\varphi_{11} = \varphi_{12} = \varphi_{13} = 0$.

The ambiguity condition (6) can thus be written:

$$1 + \alpha_2 + \alpha_3 = 0 \qquad (7)$$

This relation can be interpreted geometrically in the complex plan as a triangle which sides are the vectors associated with the complex numbers $1, \alpha_2, \alpha_3$.

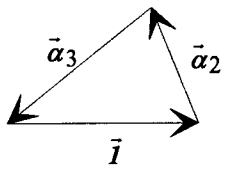


Fig. 2 : Interpretation of (7) in the complex plan.

For sensor m ambiguity condition (6) becomes:

$$e^{j\varphi_{m1}} + \alpha_2 e^{j\varphi_{m2}} + \alpha_3 e^{j\varphi_{m3}} = 0 \qquad (8)$$

In the complex plan the product by $e^{j\varphi}$ is a rotation. Thus the sides $\vec{1}, \vec{\alpha}_2, \vec{\alpha}_3$ turn respectively from angles $\varphi_{m1}, \varphi_{m2}, \varphi_{m3}$ and must reconstitute a triangle according to relation (8). The length of the sides of the triangle must be the same, therefore the triangles are deducted one from an other by an isometry. This isometry is a rotation or a rotation + a symmetry. Thus the triangles corresponding to the different values of m belong to two sub-families : the rotation family and the rotation + symmetry family.

The following results can then be derived [5] :

1) Any rank two ambiguous array may be splitten in two subarrays $a^1(\theta)$ and $a^2(\theta)$, where $a^1(\theta)$ and $a^2(\theta)$ are rank one ambiguous, for three directions θ_1, θ_2 and θ_3 i.e.: $\vec{a}^i(\theta_1) = \vec{a}^i(\theta_2) = \vec{a}^i(\theta_3)$.

2) As a consequence, the sensors for each subarray are located at the nodes of a two dimensional lattice for arbitrary θ_1, θ_2 and θ_3. The figure 3 illustrate one of these lattices.

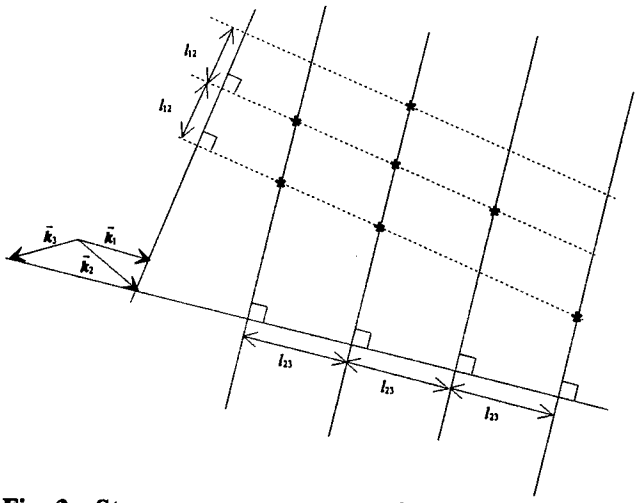


Fig. 3 : Stars represent some possible sensor positions for ambiguous subarrays (for directions θ_1, θ_2 and θ_3).

3) The second lattice corresponding to the second subarray is related by an arbitrary translation. This is a simpler demonstration and a generalisation of the theorem 1 given by Lo and Marple in reference [2].

5. Rank three ambiguities (For linear arrays)

By generalisation of the previous results, we infer that the sensors of a linear array can be splitten in three subarrays. In each subarray sensors are on a grid of spacing denoted a. The three grids are translated one from each other. For the first grid :

$$\vec{r}_m = aN_m\vec{v} \qquad (9)$$

where $\vec{v}$ is the unitary vector of the linear array.

Let us denote $\vec{k} = \frac{2\pi}{\lambda}\vec{u}$. If a is the greater common divisor of the inter sensor distances in a subarray, the ambiguity condition can be written [5] :

$$\exists n_{ij} \text{ integer, so that } \vec{v}\left(\vec{u}_i - \vec{u}_j\right) = n_{ij}\frac{\lambda}{a} \qquad (10)$$

Thus, all the sets of vectors $\vec{u}_1,\ldots,\vec{u}_4$ which can be projected on the grid of step λ/a are ambiguous. By arbitrary translation of this grid, an infinity of ambiguous direction sets can be obtained.

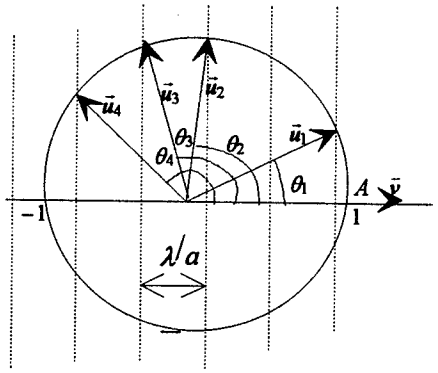


Fig. 4 : Rank three ambiguity for a linear array.

We refund here the notion of ambiguous generator set introduced by Proukakis and Manikas [3]. Their ambiguous generator set is the set of parallel lines crossing axis defined by $\vec{v}$ in the point A : (1,0).

It appears clearly on this figure that the condition for no rank three ambiguities is : $3(\lambda/a) > 2$.

Example of Proukakis and Manikas [3]

Fig 5 : Sensor positions on the array are in half wavelength.

In their example, three sources are located in : 0°, 55.582° and 82.505°. The considered array is a sparse linear array.

Two parasite peaks appear in the spectrum of MUSIC located in 107.719° and 137.657°. Because the array is ambiguous the MUSIC algorithm has provided five directions rather than three.

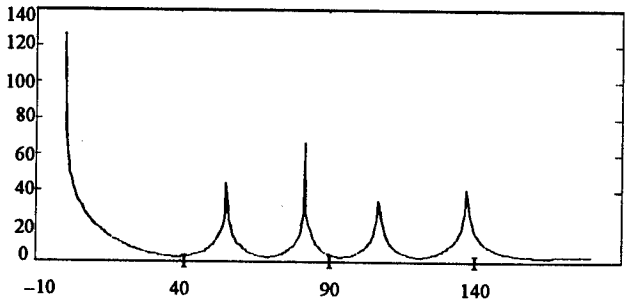


Fig. 6 : MUSIC spectrum obtained in Proukakis example.

This phenomena was not clearly explained. Application of the proposed study allows us to predict these ambiguous directions of arrival. By application of results of section 4, it is easy to see that this ambiguity is not a rank two ambiguity. Three subarrays can be find which

permit us to determine the ambiguous directions of arrival. The figure 7 depicts the splitting into subarrays which explain figure 6.

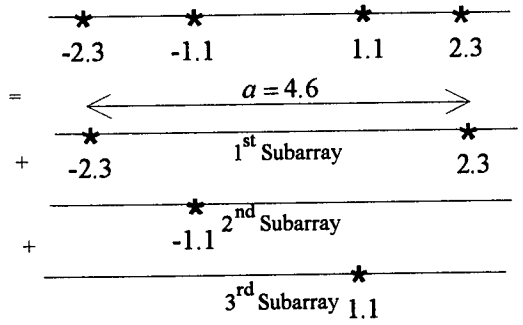


Fig. 7 : Splitting into subarrays.

In this example we take $a = 4.6$, which is the array aperture in half wavelength. Therefore $\frac{\lambda}{a} = \frac{1}{2.3}$. The second (respectively the third) subarray corresponds to the unique sensor number two (res. number three). The ambiguous directions can all be predicted with the following construction :

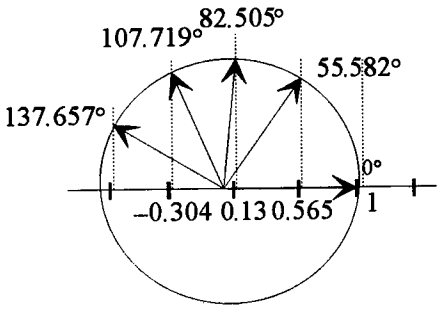


Fig. 8 : Graphic determination of the two rank three ambiguous directions for the linear array of Proukakis example.

There are two rank three ambiguity. The predicted directions of arrivals are exactly those obtained in the MUSIC spectrum.

6. Ambiguities of linear arrays (any rank)

By generalisation of the rank two case, we infer the following result : a rank k ambiguous general array can be splitten into k rank one ambiguous subarrays, for $(k+1)$ simultaneous directions. We will now focus on linear arrays for which we can obtain a general result.

A linear array is rank k ambiguous if it can be decomposed in k subarrays (which may be reduced to one sensor) with spacing $a \geq k\,\lambda/2$. The corresponding sets of ambiguous directions $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_{k+1}$ may be obtained by the following geometrical construction (given here for $k = 4$).

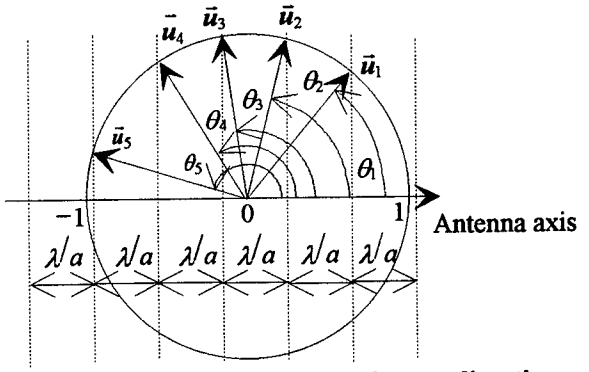


Fig. 9 : Determination of the ambiguous directions of arrival for a linear array.

The projections of vectors $\vec{u}_i$ on the antenna direction must be on a grid with spacing λ/a. Every translated grid provides also ambiguous directions.

7. Conclusion

We propose a general framework to study ambiguities for general arrays and give some properties of ambiguous array geometry. This study generalises results previously obtained in the literature.

8. References

[1] R.Schmidt, "A signal subspace approach to multiple emitters location and spectral estimation," Ph.D. dissertation, Stanford Univ., Palo Alto, CA, 1981

[2] J.T.H.Lo and L.Marple, "Observability Conditions for Multiple Signal Direction Finding and Array Sensor Localization," *IEEE Trans. Signal Processing*, vol. 40, n°11, pp. 2641-2650, Nov.92

[3] C.Proukakis and A.Manikas, "Study of ambiguities of linear arrays," in *Proc. ICASSP'94*, vol. 4, pp. 549-552, Adelaide, Australia, April 1994

[4] S.Haykin, J.P.Reilly, V.Kezys and E.Vertatschitsch, "Some aspect of array signal processing," in *Proc. IEE F*, 139, No.1, Feb 1992, pp. 1-26

[5] A.Flieller, "Mise en oeuvre des méthodes à haute résolution en traitement d'antenne : Autocalibration robuste et géométrie de réseaux," Ph.D. dissertation, Univ. Paris VI, Paris, 1995

Subspace Tracking via Rigid Body Dynamics

Daniel R. Fuhrmann, Anuj Srivastava, and Hojin Moon

Electronic Systems and Signals Research Laboratory
Department of Electrical Engineering
Washington University
St. Louis MO 63130

Abstract

The problem of estimating or tracking the time-varying principal components of a data covariance is considered. We assert that the incorporation of some notion of subspace motion or dynamics will make possible the application of subspace-based direction-finding or beamforming algorithms in scenarios which otherwise would be considered data-starved. An ordinary differential equation for simple uniform motion in the space of projection matrices is developed. This dynamical model is then used along with the artificial assumption of subspace sphericalization in a Gaussian data model, from which the cost function for maximum-likelihood estimation of subspace motion parameters is derived. Approaches to computing these subspace parameters in the one-dimensional case are proposed.

1. Introduction

For many array processing problems of interest, in both military applications and in commercial applications such as wireless mobile communications, one wishes to perform subspace-based high-resolution direction-finding and beamforming in an time-varying environment. Much has been reported in the past years in the subspace tracking problem, that of adaptively determining the principal components or subspace of a data covariance which is evolving in time. The state-of-the-art up to 1991 is described in an excellent survey paper by Comon and Golub [1], and several other promising algorithms have been proposed since then, e.g. [2-4].

In all of the algorithms reported in the literature, only the simplest of models is used to describe the motion or rotation of the subspace under consideration. This model is in some sense 1st-order Markov, by which we mean that one's best estimate of the subspace several time units into the future is simply the current value of the subspace. In these algorithms, one does not take into account any predictive value that the observed *motion* of the estimated subspace may have over time.

This lack of a predictive dynamic model for the subspace comes primarily from the fact that a subspace cannot easily be equated with a vector quantity moving about in a finite-dimensional vector space, described by a conventional state-space model. If this were the case, then a straightforward application of Kalman filtering would be an obvious approach to the problem. What we seek is a model which naturally describes the evolution of a M-dimensional subspace of $\mathbb{C}^N$, which is not an element of a Euclidean space. Given such a model, we could then consider how to use the predictive capability of such a model in a subspace tracker.

Our study of the subspace tracking problem involves two parts: a dynamical model for time-varying subspaces, and a data model which relates the observations to the desired underlying parameters. These topics are addressed in Sections 2 and 3, respectively. The general estimation problem posed at the end of Section 3 remains open. A special case in which the dimension of the subspace is 1 is of interest and is treated in Section 4. Results and conclusions follow.

2. Dynamical Model for Subspaces

Consider $X_{N,M}$ to be the space of M-dimensional subspaces of $\mathbb{C}^N$. $X \in X_{N,M}$ is a subspace in $\mathbb{C}^N$ uniquely specified by $\mathbf{P}$, an $N \times N$ projection matrix of rank M. $\mathbf{P}$ has the two properties that it is Hermitian ($\mathbf{P} = \mathbf{P}^{\mathrm{H}}$) and idempotent ($\mathbf{P}^2 = \mathbf{P}$). The set of all $N \times N$ rank-M projection matrices form a connected manifold which is denoted $P_{N,M}$.

We seek a natural dynamical description on $X_{N,M}$, or equivalently on the set of coordinate descriptions $P_{N,M}$. Subspace X can be thought of as an M-dimensional complex plane extending to infinity in all directions and containing the origin, hence its motion is basically rotational.

0-8186-7576-4/96 $5.00

In order to describe its motion, we borrow ideas from the rigid body dynamics of objects undergoing rotations with respect to the origin of some coordinate system (such as an airplane), and extend them to higher dimensions and complex spaces.

Consider a rigid body with the origin of its coordinate system fixed, rotating freely about this origin. At any time t, the orientation of this body is described by a unitary $\mathbf{Q}(t)$, sometimes called a rotation matrix. If $\mathbf{x}$ is the coordinate vector for some point on the body in its nominal orientation, then $\mathbf{y}(t) = \mathbf{Q}(t)\mathbf{x}$ gives the coordinate vector for this same point in the rotated body.

A natural differential equation for time-varying rotations is

$$\dot{\mathbf{Q}}(t) = \mathbf{Q}(t)\mathbf{A}(t) \tag{2.1}$$

where $\mathbf{A}(t)$ is a skew-symmetric matrix. To see that this is the case, note that

$$\mathbf{Q}(t)\mathbf{Q}^{\mathrm{T}}(t) = \mathbf{I} \tag{2.2}$$

and hence

$$\dot{\mathbf{Q}}(t)\mathbf{Q}^{\mathrm{T}}(t) + \mathbf{Q}(t)\dot{\mathbf{Q}}^{\mathrm{T}}(t) = \mathbf{0} \quad . \tag{2.3}$$

Defining

$$\mathbf{A}(t) = -\dot{\mathbf{Q}}^{\mathrm{T}}(t)\mathbf{Q}(t) \tag{2.4}$$

(2.1) follows immediately, and furthermore from (2.3) we have that

$$\mathbf{A}^{\mathrm{T}}(t) = -\mathbf{A}(t) \quad . \tag{2.5}$$

In the special case where $\mathbf{A}$ is a constant skew-symmetric matrix, we have the closed-form solution to (2.1) as

$$\mathbf{Q}(t) = \mathbf{Q}(0)e^{\mathbf{A}t} \quad . \tag{2.6}$$

In higher-dimensional complex spaces $\mathbb{C}^N$, an analogous result holds. We can write

$$\dot{\mathbf{Q}}(t) = \mathbf{Q}(t)\mathbf{A}(t) \tag{2.7}$$

where $\mathbf{A}$ is skew-Hermitian ($\mathbf{A}^{\mathrm{H}} = -\mathbf{A}$) and given by

$$\mathbf{A}(t) = -\dot{\mathbf{Q}}^{\mathrm{H}}(t)\mathbf{Q}(t) \quad . \tag{2.8}$$

Skew-Hermitian matrices have the property that the eigenvalues are purely imaginary, and eigenvectors corresponding to different eigenvalues are orthogonal. If $\mathbf{A}$ is held constant, then the closed-form solution to (2.7) is

$$\mathbf{Q}(t) = \mathbf{Q}(0)e^{\mathbf{A}t} \quad . \tag{2.9}$$

This dynamical model for time-varying rotations $\mathbf{Q}(t)$ is not our objective here, although it will be relevant. Rather, we seek a description of the motion of a point in $X_{N,M}$, which perhaps can be visualized as a "rigid body" consisting of an infinite sheet fixed at the origin and rotating under the influence of various torques and forces. We choose the projection matrix $\mathbf{P}$ as a unique coordinate description for a point in $X_{N,M}$ and thus consider the flow on $P_{N,M}$.

The use of dynamical models for $\mathbf{P}(t)$ was introduced by Dowling and DeGroat in [5]. There a differential equation for $\mathbf{P}(t)$ was proposed, not for the purpose of developing a subspace tracker, but rather for establishing the convergence properties of a previously developed algorithm [2]. The proposed ODE was a Riccati equation of the form

$$\dot{\mathbf{P}}(t) = \mathbf{P}(t)\mathbf{R} + \mathbf{R}\mathbf{P}(t) - 2\mathbf{P}(t)\mathbf{R}\mathbf{P}(t) \tag{2.10}$$

where $\mathbf{R}$ is a positive definite Hermitian matrix. The global attractor for this ODE is the projection operator for the true signal subspace of $\mathbf{R}$.

We propose here a simpler ODE for $\mathbf{P}(t)$ which describes something akin to "uniform motion" on $P_{M,N}$. It is

$$\dot{\mathbf{P}}(t) = \mathbf{P}(t)\mathbf{A}(t) - \mathbf{A}(t)\mathbf{P}(t) \tag{2.11}$$

for skew-Hermitian $\mathbf{A}(t)$. To see that (2.11) generates a flow on $P_{M,N}$, note that the two defining properties of projection matrices require that

$$\dot{\mathbf{P}}^{\mathrm{H}}(t) = \dot{\mathbf{P}}(t) \tag{2.12}$$

and

$$\mathbf{P}(t)\dot{\mathbf{P}}(t) + \dot{\mathbf{P}}(t)\mathbf{P}(t) = \dot{\mathbf{P}}(t) \tag{2.13}$$

both of which are satisfied by (2.11) provided that $\mathbf{A}$ is skew-Hermitian. Furthermore, since $\mathbf{P}(t)$ is continuous for bounded $\mathbf{A}$, it cannot change rank at any time t.

If $\mathbf{A}$ is held constant, the solution to (2.11) is

$$\mathbf{P}(t) = \mathbf{Q}^{\mathrm{H}}(t)\mathbf{P}(0)\mathbf{Q}(t) \tag{2.14}$$

where

$$\mathbf{Q}(t) = e^{\mathbf{A}t} \tag{2.15}$$

by the previous arguments of this section. We take (2.11) to be the simplest non-trivial model for uniform motion in the space $P_{N,M}$, and consider the flow to be somewhat analogous to constant velocity motion along a straight line for points in a Euclidean space. With this analogy in mind, we now consider the problem of estimating the parameters of this uniform motion, given observations from a particular stochastic model.

3. Data Model

The usual model often considered in the application of subspace-based algorithms such as MUSIC or ESPRIT, is one in which the data vector $\mathbf{x} \in \mathbb{C}^N$ is complex Gaussian with mean 0 and covariance $\mathbf{R}$. The shorthand notation for this is

$$\mathbf{x} \sim \mathrm{CN}(0, \mathbf{R}) \quad . \tag{3.1}$$

The eigenvalues of $\mathbf{R}$ can be split into two classes, with the eigenvectors of $\mathbf{R}$ corresponding to the larger eigenvalues spanning the *signal subspace*, and the eigenvectors corresponding to the smaller eigenvalues spanning the *noise subspace*. The maximum-likelihood estimates of these two subspaces are found by *principal component analysis*: form a sample covariance matrix from the data vectors, and compute an eigendecomposition to determine the signal and noise subspaces.

It is interesting to note that the eigenvalues of $\mathbf{R}$ are nuisance parameters if all that is desired are the two subspaces. Furthermore, the ML estimates of the subspaces do not change if one introduces an artificial constraint that all of the signal subspace eigenvalues are equal (σ_s^2) and that all of the noise subspace eigenvalues are equal (σ_n^2), with $\sigma_s^2 > \sigma_n^2$. DeGroat and Dowling [2,5] refer to this assumption as *subspace sphericalization*, and several computational advantages follow from it. Under this model the covariance $\mathbf{R}$ is given by

$$\mathbf{R} = \sigma_s^2 \mathbf{P}_s + \sigma_n^2 \mathbf{P}_n \tag{3.2}$$

where $\mathbf{P}_s$ and $\mathbf{P}_n$ are projection matrices for the signal and noise subspaces, respectively, and

$$\mathbf{P}_n = \mathbf{I} - \mathbf{P}_s \quad . \tag{3.3}$$

For time-varying subspaces, we propose the following data model which incorporates the simple dynamic model of Section 2 and subspace sphericalization. Let

$$\mathbf{x}(k) \sim \mathrm{CN}(0, \mathbf{R}(k)) \quad k = 1 \cdots K \tag{3.4}$$

where

$$\mathbf{R}(k) = \sigma_s^2 \mathbf{P}_s(kT) + \sigma_n^2 \mathbf{P}_n(kT) \quad . \tag{3.5}$$

$\mathbf{P}_s(t)$ is evolving in time according to the differential equation

$$\dot{\mathbf{P}}(t) = \mathbf{P}(t)\mathbf{A} - \mathbf{A}\mathbf{P}(t) \tag{3.6}$$

with $\mathbf{A}$ held constant. From the results of the previous section we have

$$\mathbf{P}_s(kT) = e^{-\mathbf{A}kT} \mathbf{P}_0 e^{\mathbf{A}kT} \quad . \tag{3.7}$$

The problem of tracking $\mathbf{P}_s(t)$ reduces, under this model, to that of estimating $\mathbf{P}_0$ and $\mathbf{A}$ (or $\mathbf{Q} = e^{\mathbf{A}T}$).

The probability density function for the observed data is

$$f_{\mathbf{x}} = \prod_{k=1}^{K} \pi^{-NK} [\sigma_s^{2M} \sigma_n^{2(N-M)}]^{-K} e^{-\mathbf{x}^{\mathrm{H}}(k)\mathbf{R}^{-1}(k)\mathbf{x}(k)} \tag{3.8}$$

where

$$\mathbf{R}^{-1}(k) = \sigma_s^{-2} \mathbf{P}_s(kT) + \sigma_n^{-2} \mathbf{P}_n(kT) \quad . \tag{3.9}$$

Straightforward manipulation of (3.8-3.9) reveals that the ML estimates of the parameters $\mathbf{P}_0$ and $\mathbf{A}$ are not functions of σ_s^2 and σ_n^2, and that they are found via maximization of the cost function

$$J = \sum_{k=1}^{K} \mathbf{x}^{\mathrm{H}}(k) e^{-\mathbf{A}kT} \mathbf{P}_0 e^{\mathbf{A}kT} \mathbf{x}(k) \quad . \tag{3.10}$$

The interpretation of (3.10) is intuitive: we seek a set of rotations $e^{\mathbf{A}kT}$, which, when applied in sequence to the data vectors $\mathbf{x}(k)$, would move them as closely as possible back to a single M-dimensional subspace described by $\mathbf{P}_0$.

At the time of this writing we do not have a general closed form solution for the maximization of J, and consider it to be an open research problem.

4. The One-Dimensional Case

The special case $M = 1$ deserves special attention. When $M = 1$, each data vector $\mathbf{x}(k)$ can support a crude estimate of $\mathbf{P}(k)$, independent of the other data, and these estimates can be used in various *ad hoc* ways to approximate the matrix $\mathbf{Q} = e^{\mathbf{A}T}$ which relates them all.

For the sake of argument, let us suppose that each data vector $\mathbf{x}(k)$ lies in the range of the one-dimensional projection $\mathbf{P}(k)$. Define the normalized observation

$$\mathbf{u}(k) = \mathbf{x}(k)/|\mathbf{x}(k)| \tag{4.1}$$

and then

$$\mathbf{P}(k) = \mathbf{u}(k)\mathbf{u}^{\mathrm{H}}(k) \quad . \tag{4.2}$$

From (2.14) we have the recurrence relation

$$\mathbf{P}(k) = \mathbf{Q}^{\mathrm{H}} \mathbf{P}(k-1) \mathbf{Q} \quad . \tag{4.3}$$

Note that (4.3) is not equivalent to the statement $\mathbf{u}(k) = \mathbf{Q}^{\mathrm{H}}\mathbf{u}(k-1)$, since we only know that $\mathbf{u}(k)$ lies in the range of $\mathbf{P}(k)$. However, we can say that

$$\mathbf{u}(k)d(k) = \mathbf{Q}^{\mathrm{H}}\mathbf{u}(k-1) \tag{4.4}$$

where $d(k)$ is an unknown complex scalar of the form

$$d(k) = e^{j\theta_k} \quad . \tag{4.5}$$

One promising approach to the estimation of the $\mathbf{Q}$, which can be interpreted as the small rotation applied to the subspace of interest at each time step, is to find parameters $d(k)$ and $\mathbf{Q}$ which make (4.4) approximately true for $k = 2 \cdots K$. Define

$$\mathbf{A} = [\mathbf{u}(2) \mid \cdots \mid \mathbf{u}(K)] \tag{4.6}$$

and

$$\mathbf{B} = [\mathbf{u}(1) \mid \cdots \mid \mathbf{u}(K-1)] \quad . \tag{4.7}$$

Then in matrix form (4.4) becomes

$$\mathbf{AD} = \mathbf{Q}^{\mathrm{H}}\mathbf{B} \tag{4.8}$$

where

$$\mathbf{D} = diag(d(1) \cdots d(K)) \quad . \tag{4.9}$$

Given sufficient data, (4.8) will not be exactly true, but we can make it approximately true by defining an appropriate cost function and minimizing it with respect to the unknown parameters $\mathbf{D}$ and $\mathbf{Q}$. The squared-error cost function is

$$H = tr(\mathbf{AD} - \mathbf{Q}^{\mathrm{H}}\mathbf{B})(\mathbf{AD} - \mathbf{Q}^{\mathrm{H}}\mathbf{B})^{\mathrm{H}} \quad . \tag{4.10}$$

There are closed-form solutions for the two problems of minimizing H w.r.t $\mathbf{Q}$ holding $\mathbf{D}$ fixed, and vice versa.

Fix $\mathbf{D}$ in (4.10), and let

$$\tilde{\mathbf{A}} = \mathbf{AD} \quad . \tag{4.11}$$

The cost function becomes

$$H = tr(\mathbf{B} - \mathbf{Q}\tilde{\mathbf{A}})(\mathbf{B} - \mathbf{Q}\tilde{\mathbf{A}})^{\mathrm{H}} \quad . \tag{4.12}$$

Expanding, we have

$$H = tr(\mathbf{B}\mathbf{B}^{\mathrm{H}} - \mathbf{Q}\tilde{\mathbf{A}}\mathbf{B}^{\mathrm{H}} - \mathbf{B}\tilde{\mathbf{A}}^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^{\mathrm{H}}) \tag{4.13}$$

and thus H is minimized when we maximize

$$H' = \mathrm{Re}\,\{tr\mathbf{Q}\tilde{\mathbf{A}}\mathbf{B}^{\mathrm{H}}\} \quad . \tag{4.14}$$

Let the singular value decomposition (SVD) of $\tilde{\mathbf{A}}\mathbf{B}^{\mathrm{H}}$ be given by

$$\tilde{\mathbf{A}}\mathbf{B}^{\mathrm{H}} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{H}} \quad . \tag{4.15}$$

Then

$$H' = \mathrm{Re}\,\{tr\mathbf{Q}\mathbf{U}\Sigma\mathbf{V}^{H}\} \quad . \tag{4.16}$$

H' is upper-bounded by $tr\Sigma$ and this bound can be met with equality when

$$\mathbf{Q} = \mathbf{V}\mathbf{U}^{\mathrm{H}} \quad . \tag{4.17}$$

This gives a constructive method for finding the optimal $\mathbf{Q}$ when $\mathbf{D}$ is fixed.

Likewise, when $\mathbf{Q}$ is fixed, let

$$\tilde{\mathbf{B}} = \mathbf{Q}^{\mathrm{H}}\mathbf{B} \tag{4.18}$$

and solve for $\mathbf{D}$ using least-squares. This leads to K decoupled problems because of the diagonal constraint on $\mathbf{D}$. Define $\mathbf{a}_k$ and $\mathbf{b}_k$ to be the kth column of $\mathbf{A}$ and $\tilde{\mathbf{B}}$, respectively. Then the least-squares solution for d_k is given by $d_k = e^{-j\theta_k}$ where

$$\theta_k = arg(\mathbf{a}_k^{\mathrm{H}}\mathbf{b}_k) \quad . \tag{4.19}$$

5. Conclusion

We have proposed the use of a dynamic predictive model for subspaces for application in the subspace tracking problem. A model for simple uniform motion in the space of projection matrices has been developed. This model, combined with a complex Gaussian data model, leads to a maximum-likelihood estimation problem where the parameters of interest are the subspace motion parameters.

Successful incorporation of subspace dynamics into the tracking problem will extend the range of useful applications of subspace methods in high-resolution direction-finding and beamforming. The work described in this paper represents a preliminary investigation of this potentially important problem.

References

[1] P. Comon and G. Golub, "Tracking a few extreme singular values and vectors in signal processing", *Proc. IEEE*, vol. 78, no. 8, pp. 1327-1343, August 1990.

[2] R. DeGroat, "Noniterative subspace tracking", *IEEE Trans. Signal Processing*, vol. 40, no. 3, pp. 571-577, March 1992.

[3] G. Stewart, "An updating algorithm for subspace tracking", *IEEE Trans. Signal Processing*, vol. 40, no. 6, pp. 1535-1541, June 1992.

[4] B. Yang, "Projection approximation subspace tracking", *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 95-107, January 1995.

[5] E. Dowling and R. DeGroat, "Adaptation dynamics of the spherical subspace tracker", *IEEE Trans. Signal Processing*, vol. 40, no. 10, pp. 2599-2602, October 1992.

Blind Multi-Channel System Identification and Deconvolution: Performance Bounds *

Daniel Yellin and Benjamin Friedlander
Department of Electrical & Computer Engineering
University of California
Davis, CA 95616

Abstract

We consider the problem of estimating the parameters of an unknown multi-input multi-output linear system, and the related problem of deconvolving and recovering its inputs, using only observations of the system outputs. We derive simple closed-form asymptotic expressions for the Cramer-Rao lower bound (CRLB) for the system parameters, as well as lower bounds on the signal reconstruction performance. These show that the identification/deconvolution performance depend on the accuracy with which the scale and the location parameters of the input probability density functions can be identified from observation of the input signals. It is also shown that the CRLB possesses a block diagonal structure, indicating that the general multi-channel deconvolution problem is decoupled into two independent *simpler sub-problems: The signal separation problem where the unknown system is deconvolved to a diagonal one, and the remaining* independent *single-channel deconvolution problems associated with the equalization of each of its diagonal elements.*

1 Introduction

In many applications, observations are made on the outputs of an unknown multi-input multi-output (MIMO) linear system, from which it is necessary to identify the system and recover its inputs. A classical example is the problem of separating several speakers using multiple microphone measurements. The unknown system in this case represents the acoustic media which couples the speakers to the microphones, including all of its multipath and reverberation effects. Another example, receiving growing attention recently, is the problem associated with the recovery of data communication signals that share the same frequency band. Multiple receivers are typically used to decouple and reconstruct the original transmitted information from its superimposed and distorted observations. Here, the system to be identified and deconvolved is the MIMO communication channel which links the information sources to the receivers. Similar problems can be found in diverse fields of engineering and applied science including radar/sonar array processing, seismic exploration, radio astronomy, econometrics, and more.

*This work was supported by the Office of Naval Research under contract no. N00014-95-1-0912, and by the University of California MICRO program and Applied Signal Technology, Inc.

2 Problem Formulation

We consider the two closely related problems of multi-channel system identification and deconvolution, in which we observe the outputs $y_1(t), \ldots, y_N(t)$ of an unknown $N \times N$ stable linear time invariant (LTI) system $\mathcal{H}$, whose (unobserved) inputs are $s_1(t), \ldots, s_N(t)$ and whose frequency response is:

$$\mathcal{H}(\omega) = \begin{bmatrix} H_{11}(\omega) & \ldots & H_{1N}(\omega) \\ \vdots & & \vdots \\ H_{N1}(\omega) & \ldots & H_{NN}(\omega) \end{bmatrix} \quad (1)$$

Thus, the entry $H_{ij}(\omega)$ of $\mathcal{H}(\omega)$ denotes the frequency response of the SISO system that couples input $s_j(t)$ to output $y_i(t)$.

Based on observation of $y_1(t), \ldots, y_N(t)$ we want to identify $\mathcal{H}$, and/or deconvolve and recover its inputs using an $N \times N$ reconstruction system $\mathcal{G}$, whose inputs are $y_1(t), \ldots, y_N(t)$ and whose outputs are $\widehat{s}_1(t), \ldots, \widehat{s}_N(t)$.

Let $\mathcal{A}$ denote the combined system relating the reconstructed signals $\widehat{s}_1(t), \ldots, \widehat{s}_N(t)$ to the input signals $s_1(t), \ldots, s_N(t)$. Then,

$$\mathcal{A}(\omega) = \mathcal{G}(\omega)\,\mathcal{H}(\omega) \quad (2)$$

where $\mathcal{A}(\omega)$ and $\mathcal{G}(\omega)$ are the frequency responses of the systems $\mathcal{A}$ and $\mathcal{G}$, respectively, and $\mathcal{H}(\omega)$ is given by (1).

0-8186-7576-4/96 $5.00 © 1996 IEEE

Assuming that $\mathcal{H}$ is an invertible system, we want to set

$$\mathcal{G}(\omega) \;=\; \mathcal{H}(\omega)^{-1} \qquad \forall\omega \tag{3}$$

which according to (2) implies that $\mathcal{A}$ is a unity transformation, in which case $\widehat{s}_i(t) = s_i(t)$ and the inputs are exactly recovered.

The following two assumptions will be used throughout:

Assumption I The input signals are sample functions (realizations) from mutually independent stochastic processes.

Assumption II Each of the input signals constitute a sequence of independent and identically distributed (i.i.d.) random variables.

Assumptions I and II are sufficient (although in some special cases not necessary) for the identification of $\mathcal{H}$ and the recovery of its inputs. In the following sections, we study how accurately these two tasks can be performed.

3 The Cramer-Rao Lower Bound

Derivation of the *exact* Cramer-Rao lower bound (CRLB) for the problem at hand is generally intractable. We will develop therefore the asymptotic bound only, in which case the "end effects" can be ignored. A similar approach was used in e.g. [2] [5] [8] [10] [11] for various SISO identification/deconvolution problems. Furthermore, in order to simplify the analysis and to gain several important insights, we first compute the CRLB with respect to the components of the combined system $\mathcal{A}$, as if they are free adjustable parameters. We then use the relation in (2) to derive the bound with respect to the components of $\mathcal{H}$. As we shall see in Section 4, the accuracy in which the system $\mathcal{A}$ can be identified is by itself an important issue, as it governs the deconvolution performance.

3.1 The CRLB for $\mathcal{A}$

Let quantities with over-bar denote the true parameter values. Thus, $\bar{\mathcal{H}}$ stands for the system that actually generated $y_1(t), \ldots, y_N(t)$. We set $\mathcal{G}$ to be the inverse of $\bar{\mathcal{H}}$. Therefore, we shall compute the CRLB for the system $\mathcal{A}$ whose true value, $\bar{\mathcal{A}}$, is the $N \times N$ unity system.

Suppose that T observations of the signals $\widehat{s}_1(t), \ldots, \widehat{s}_N(t)$ are available. Then, the log-likelihood of the observed data is asymptotically given by:

$$\begin{aligned} &-\log[\,|\det\{\mathbf{A}\}|\,] + \sum_{t=1}^{T}\log[\,\mathcal{P}_{S_1}(\widehat{v}_1(t))\,] \\ &\qquad + \ldots + \sum_{t=1}^{T}\log[\,\mathcal{P}_{S_N}(\widehat{v}_N(t))\,] \end{aligned} \tag{4}$$

where $\mathbf{A}$ is an $NT \times NT$ block matrix whose i,j block is the $T \times T$ Toeplitz matrix containing the unit sample response coefficients of A_{ij}, $\widehat{v}_1(t), \ldots, \widehat{v}_N(t)$ are the outputs of the $N \times N$ system whose frequency response is $\mathcal{A}(\omega)^{-1}$ and whose inputs are $\widehat{s}_1(t), \ldots, \widehat{s}_N(t)$, and $\mathcal{P}_{S_i}(x)$ stands for the pdf of $s_i(t)$ which is assumed to be strictly positive and differentiable, for all x.

Time Domain Parameterization

Suppose that the unknown parameters are the unit sample response coefficients $a_{ij}(\tau)$, for $i,j \in \{1,\ldots,N\}$ and $\tau \in \{0, \pm 1, \ldots, \pm K\}$. The non-negative integer K is assumed to be much smaller than T, so that the total number of unknowns is small compared to the data block size.

Define the log-likelihood gradient $\ell_{ij}(\tau)$ as the derivative of (4) with respect to $a_{ij}(\tau)$ at $\mathcal{A} = \bar{\mathcal{A}}$, where $a_{ij}(\tau)$ is the unit sample response of A_{ij}. Then,

$$\ell_{ii}(\tau) \;=\; -T\,\left(\delta(\tau) + \widehat{R}_{z_i s_i}(\tau)\right) \tag{5}$$

$$\ell_{ij}(\tau) \;=\; -T\,\widehat{R}_{z_i s_j}(\tau) \qquad i \neq j \tag{6}$$

where $\delta(\,\cdot\,)$ denotes the Kronecker delta function, and

$$z_i(t) \;=\; \frac{\mathcal{P}_{S_i}(s_i(t))'}{\mathcal{P}_{S_i}(s_i(t))} \tag{7}$$

$$\widehat{R}_{z_i s_j}(\tau) \;=\; \frac{1}{T}\sum_{t=1}^{T} z_i(t)\, s_j(t-\tau) \tag{8}$$

Thus, the non-zero elements of the Fisher Information matrix (FIM) are given by:

$$\begin{aligned} E\{\ell_{ii}(\tau_1)\ell_{ii}(\tau_2)\} \;&=\; T\,[\,\mathrm{cum}\{z_i, z_i, s_i, s_i\}\,\delta(\tau_1)\,\delta(\tau_2) \\ &+ \quad \mathcal{L}_i\,\delta(\tau_1-\tau_2) \;+\; \delta(\tau_1+\tau_2)\,] \end{aligned} \tag{9}$$

$$E\{\ell_{ij}(\tau_1)\ell_{ij}(\tau_2)\} \;=\; T\,\frac{\mathrm{Var}\{s_j\}}{\mathrm{Var}\{s_i\}}\,\mathcal{L}_i\,\delta(\tau_1-\tau_2) \tag{10}$$

$$E\{\ell_{ij}(\tau_1)\ell_{ji}(\tau_2)\} \;=\; T\,\delta(\tau_1+\tau_2) \tag{11}$$

where cum{ } denotes the joint cumulant of the random variables in the brackets, and

$$\mathcal{L}_i \;=\; \mathrm{Var}\{z_i\}\mathrm{Var}\{s_i\} \tag{12}$$

which is known as Fisher's information for the location parameter (FIL).

All other elements of the FIM are zero, indicating lack of statistical correlation between the estimates of the diagonal and the off-diagonal elements of $\bar{\mathcal{A}}$. Furthermore, the estimates of $\bar{a}_{ii}(\tau)$ are uncorrelated with the estimates of $\bar{a}_{jj}(\tau)$ for $i \neq j$. Therefore, at least asymptotically, the general $N \times N$ identification/deconvolution problem can be decoupled into two *independent* simpler sub-problems: The signal separation problem associated with the identification of

the $\bar{a}_{ij}(\tau)$s with $i \neq j$, and the remaining N *independent* single-channel problems in which the $\bar{a}_{ii}(\tau)$s are specified. Furthermore, the $N \times N$ signal separation problem further decouples into $\binom{N}{2}$ *independent* pairwise separation problems.

The CRLB on the error variance of any unbiased estimate $\widehat{a}_{ij}(\tau)$ of $\bar{a}_{ij}(\tau)$ is given by the inverse of the FIM,

$$\mathrm{Var}\{\widehat{a}_{ij}(\tau)\} \geq \frac{1}{T}\frac{\mathrm{Var}\{s_i\}}{\mathrm{Var}\{s_j\}}\frac{\mathcal{L}_j}{\mathcal{L}_i\mathcal{L}_j - 1} \qquad (13)$$

where for $i = j$ and $\tau = 0$, the term $\frac{\mathcal{L}_i}{\mathcal{L}_i^2-1}$ should be replaced by $\frac{1}{\mathcal{S}_i}$, with $\mathcal{S}_i$ being the Fisher information for the scale parameter (FIS),

$$\mathcal{S}_i = \mathrm{cum}\{z_i, z_i, s_i, s_i\} + \mathcal{L}_i + 1 \qquad (14)$$

The simple structure of the bound in (13), indicates that essentially, it is the FIL's (or FIS's) that govern the estimation accuracy of $\widehat{a}_{ij}(\tau)$. Thus, the identification of the combined system $\mathcal{A}$, is strongly related to the basic problem of estimating the scale and location parameters of the input distributions based on direct observation of the input signals. In fact, if $\mathrm{Var}\{s_i(t)\} = 1$ and $\mathcal{L}_i \gg 1$ $i \in \{1, \ldots, N\}$, then for all combinations of i, j and τ except $i = j$, $\tau = 0$, the RHS of (13) coincides with the CRLB for the estimation of the location parameter of $\mathcal{P}_{S_i}(x)$ given T independent realizations. Similarly, for $i = j$ and $\tau = 0$, the RHS of (13) coincides with the CRLB for the estimation of the scale parameter of $\mathcal{P}_{S_i}(x)$.

Due to the block diagonal nature of the FIM, the RHS of (13) with $i = j$ is the CRLB for estimating $\bar{a}_{ii}(\tau)$ given that all the other components of the system $\bar{\mathcal{A}}$ are *known a-priori*. Therefore, it coincides with the result of [2], where the CRLB for the identification of a SISO system from observation of its output was derived.

Similarly, the RHS of (13) with $i \neq j$ coincides with the CRLB for estimating $\bar{a}_{ij}(\tau)$ given that all the other components of $\bar{\mathcal{A}}$ except to $\bar{a}_{ji}(\tau)$ are *known a-priori*. Thus, lack of precise knowledge of the $\bar{a}_{ii}(\tau)$'s do not affect the estimation of the $\bar{a}_{ij}(\tau)$'s with $i \neq j$, and vice-versa.

Frequency Domain Parameterization

Next consider the frequency domain formulation in which the unknown parameters are set to be $A_{ij}(\omega) = \sum_\tau a_{ij}(\tau)\, e^{-\sqrt{-1}\omega\tau}$ at the DFT frequencies

$$\omega \in \{2\pi B k \ ; \ k = 0, 1, \ldots, \frac{1}{B} - 1\},$$

where we assume that $\frac{1}{B}$ is an integer and that $1 \gg B \gg \frac{1}{T}$. As before, the overall number of unknown parameters is small compared to the data block size.

By analogy to the time domain formulation, one can show that the CRLB is given by:

$$\mathrm{Var}\{\widehat{A}_{ij}(\omega)\} \geq \frac{1}{BT}\frac{\mathrm{Var}\{s_i\}}{\mathrm{Var}\{s_j\}}\frac{\mathcal{L}_j}{\mathcal{L}_i\mathcal{L}_j - 1} \qquad (15)$$

where for $i = j$ and $\omega = 0, \pm\pi$, the term $\frac{\mathcal{L}_i}{\mathcal{L}_i^2-1}$ should be replaced by $\frac{1}{\mathcal{L}_i+1}$.

Observe that (15) is identical, up to the factor of $\frac{1}{B}$, to the CRLB for the time domain parameterization given in (13). Therefore, the time and frequency domain parameterizations are essentially equivalent. However, due to the windowing operation used in the frequency domain parameterization, the data block is effectively smaller in that case (by a factor of $\frac{1}{B}$).

Note also that if $s_1(t), \ldots, s_N(t)$ are jointly Gaussian, then $\mathcal{L}_i = 1$ and the RHS of (15) is infinite indicating that the problem can not be solved in this case. The reason being that in the Gaussian case all the available information about the unknown system is contained in the first and second-order statistics of the observed signals. These statistics are "blind" to unitary transformations on the data signals. Therefore, the system may only be identified up to an arbitrary unitary transformation.

Furthermore, consider for simplicity the case $N = 2$ and suppose that one of the inputs, say $s_1(t)$, is non-Gaussian. Then, the performance is the worst when the other signal $s_2(t)$ is Gaussian. The variances of $\widehat{A}_{12}(\omega)$ and $\widehat{A}_{21}(\omega)$ are the highest in this case, and the variance of $\widehat{A}_{22}(\omega)$ is infinite in accordance with the well known fact that a SISO system driven by a Gaussian process can not be identified/deconvolved. To verify this note that $\mathcal{L}_2 \geq 1$ with equality if and only if $s_2(t)$ is Gaussian (see e.g. [2] [10] [11]), and the RHS of (15) is always a monotone decreasing function of $\mathcal{L}_2$.

3.2 The CRLB for $\mathcal{H}$

Recall that we were originally interested in the CRLB with respect to the components of $\bar{\mathcal{H}}$. Since $\bar{\mathcal{H}}$ and $\bar{\mathcal{A}}$ are related through the linear transformation $\mathcal{G}$, we can translate the results above to a bound on the components of $\bar{\mathcal{H}}$. With the frequency domain parameterization we obtain simple closed form expressions. Using the relation in (2), which holds approximately for the DFT frequencies, the asymptotic CRLB on the estimate $\widehat{H}_{ij}(\omega)$ of $\bar{H}_{ij}(\omega)$ is:

$$\mathrm{Var}\{\widehat{H}_{ij}(\omega)\} \geq \frac{1}{BT}\sum_{k=1}^{N}|\bar{H}_{ik}(\omega)|^2\frac{\mathrm{Var}\{s_k\}}{\mathrm{Var}\{s_j\}}\frac{\mathcal{L}_j}{\mathcal{L}_k\mathcal{L}_j - 1} \qquad (16)$$

where for $k = j$ and $\omega = 0, \pm\pi$ the term $\frac{\mathcal{L}_j}{\mathcal{L}_j^2-1}$ should be replaced by $\frac{1}{\mathcal{L}_j+1}$.

4 Signal Reconstruction

In the previous section we addressed the system identification problem, and examined how accurately the systems $\mathcal{A}$ and $\mathcal{H}$ can be identified. Next we consider the signal deconvolution issue, and determine how precisely the inputs can be recovered.

A common measure of signal reconstruction is the interference-to-signal power ratio at each of the reconstruction filter outputs. Invoking (15) and the model assumptions, it is not difficult to verify that the interference-to-signal at the ith output terminal is bounded by:

$$(\mathrm{I/S})_i \geq \frac{N-1}{BT}\frac{1}{\mathcal{L}_i} \tag{17}$$

Note that the expression in (17) does not depend on the pre-processing interference-to-signal ratios nor on the unknown system $\bar{\mathcal{H}}$. It only depends on the basic amount of information contained in the input signals with respect to their location parameters.

Another useful measure of signal reconstruction is the mean square restoration error (MSE), defined as:

$$MSE_i = E\{[s_i(t) - \hat{s}_i(t)]^2\} \tag{18}$$

Once again, invoking the results of the previous section, it can be shown that MSE_i is bounded by:

$$MSE_i \geq \frac{N}{BT}\frac{\mathrm{Var}\{s_i(t)\}}{\mathcal{L}_i} \tag{19}$$

Finally, we note that the bounds in (17) and (19) hold also for the case where the system $\bar{\mathcal{H}}$ has more outputs than inputs. In such a case, one may use several different sets of N outputs of the system $\bar{\mathcal{H}}$ to generate different sets of reconstructed signals. Then, average over the different reconstructed signal sets, in an attempt to improve the performance. However, the set of averaged reconstructed signals is related to the input signal through some equivalent $N \times N$ system, that has the *same* true value $\bar{\mathcal{A}}$. Thus, the reconstruction performance remain intact.

Of course, if additive noises are present, then the above procedure will indeed improve performance as the noise contributions will be averaged out. However, for low level of noise, this improvement is expected to be small. We therefore conclude that for sufficiently high SNR, there is not much point in trying to increase the number of available data sensors beyond the minimum required.

References

[1] Y. Bar-Ness, J. Carlin, and M. Steinberger, "Bootstrapping adaptive cross-pol canceler for satellite communications", *IEEE International Conference on Communications*, pp. 4F.5.1–4F.5.5, Philadelphia, PA, June 1982.

[2] S. Bellini, and F. Rocca, "Asymptotically efficient blind deconvolution," *Signal Processing*, vol 20, No. 1, pp 193-209, 1990.

[3] J. Cardoso, and A. Souloumiac "An efficient technique for the blind separation of complex sources," *Proc. IEEE SP Workshop on Higher-Order Statistics*, Lake Tahoe, CA, USA June 1993.

[4] P. Comon, "Independent component analysis, a new concept ? ", *Signal Processing*, vol. 36, No. 3, April 1994.

[5] B. Friedlander, "On the computation of the Cramer-Rao bound for ARMA parameter estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-32, No. 4, pp. 721-727, August 1984.

[6] C. Jutten, and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol 24, No. 1, pp 1-10, July 1991.

[7] J. L. Lacoume, and P. Ruiz, "Separation of independent sources from correlated inputs," *IEEE Trans. Signal Processing*, Vol. 40, No 12, pp 3074-3078, December 1992.

[8] L. Ljung, *System identification: Theory for the user.* Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[9] E. Moreaum, and O. Macchi, "High order contrasts for self-adaptive source separation," *Int. Journal of Adaptive Control and Signal Processing.* To appear.

[10] D. Sengupta and S. Kay, "Efficient estimation of parameters for non-Gaussian autoregressive processes," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 37, pp. 785-794, June 1989.

[11] O. Shalvi, and E. Weinstein, "Maximum likelihood and lower bounds in system identification with non-Gaussian inputs, " *IEEE Trans. on IT.* Vol-40, No. 2, pp. 328-339, March 1994.

[12] D. Yellin, and E. Weinstein, "Multi-channel signal separation: Methods and analysis," *IEEE Trans. Signal Processing* vol. 44, No. 11, pp. 106-118, January 1996.

[13] D. Yellin, and B. Friedlander, "Multi-channel system identification and deconvolution: Performance bounds," Submitted to the *IEEE Trans. on Information Theory.*

Index of Authors

IEEE COMPUTER SOCIETY
50 YEARS OF SERVICE • 1946-1996
http://www.computer.org

Press Activities Board