Computer Science Journal of Moldova, vol.**17**, no.1(49), 2009

Computer modeling of multidimensional problems of gravitational gas dynamics on multiprocessor computers

Boris Rybakin, Natalia Shider

Abstract

This article deals with a method, based on total variation diminishing (TVD) scheme, for solving three-dimensional equations of gravitational gas dynamics. For this method a parallel algorithm of the decision is offered. Equations of this kind are a powerful approach to simulating astrophysical problems. Numerical schemes applied for their solving must provide high-resolution capturing of shocks, prevent spurious oscillations and specify the behavior of the matter in the neighborhood of small perturbations beyond shock fronts. Difference schemes have to combine the properties of high resolution in the regions of small perturbations and of monotonicity in the domains of steep gradients in order to satisfy such contradictory conditions.

1 Introduction

Modeling supernova explosions is referred to complex dynamic processes, requiring application of difference schemes of a high resolution. These have to describe the behavior of the matter in the neighborhood of the discontinuity at the maximum accuracy and to refer small perturbations far from shock fronts definitely. Conditions of such kind lead to the necessity of loss of dissipative (nonconservative) properties of numerical schemes and therefore to the apparition of large oscillations beyond shock fronts.

TVD, ENO, WENO, PPM schemes refer to the kind of schemes that satisfy all these necessary conditions and possess high resolution in regions of small perturbations combined with monotonicity in domains of steep gradients.

1.1 Magnetorotational mechanism of supernova explosion

In [1] a mechanism for the magnetorotational supernova explosion was analyzed. The basic concept of magnetorotational explosion consists in taking account on transition of rotating magnetic field energy into the radial kinetic energy of explosion. Various layers of the star rotate at the different angular velocities during the collapse. Differential rotation of this kind generates and enforces the magnetic field toroidal components. The growth of magnetic field intensity leads to the increase of pressure. Hence a compression shock wave appears in the neighborhood of the extreme magnetic pressure. It starts moving from the center toward the considerably fast falling density of the matter. For a rather short time this leads to the appearance of the fast magnetohydrodynamics (MHD) shock. When the shock wave reaches the surface of the collapsing star it throws out its matter. This emission may be interpreted as an explosion of supernova. Modeling magnetorotational supernova explosion in one-dimensional setting was examined, for example, in [2] and [3]. In one-dimensional case the star may be represented in the form of an infinite cylinder. The equations of ideal MHD with a self-gravitating substance in terms of Lagrangian coordinate system were considered.

The initial magnetic field had only the radial component. Differential rotation led to the appearance and increasing of toroidal component of the magnetic field. Modeling magnetorotational explosion of supernova in one-dimensional case illustrates that differential rotation of toroidal field leads to the apparition of the MHD shocks, moving towards the surface of the star. Modeling supernova explosion in two-dimensional case gives a more realistic flow pattern than in one-dimensional case. The first two-dimensional model of rotating star collapse was analyzed in [4]. The magnetic field magnitude considered in that work was unrealistic large and together with the differential rotating it led to the formation of axial emission.

Simulation of magnetorotational supernova explosion in three-dimensional case is considered in this work. Three-dimensional model of collapse is the most realistic one and does not have any restrictions connected with the assumptions, stated in 1D and 2D models. Three-dimensional models admit simulating of magnerotational supernova explosion in cases, when the axises of rotation do not coincide with the axises of dipole magnetic field (if dipole is taken as the initial value of magnetic field). If numerical schemes, elaborated for simulating two-dimensional cases, are utilized in three-dimensional case, then it will lead to big problems. The substance of the star compresses in the direction of φ in two-dimensional case. It is necessary to calculate hundreds and thousands of cycles of rotation for simulating the explosion of protoneutron star. The protoneutron star rotation occurs very differently. If in three-dimensional case Lagrangian mesh contains tetragonal elements, then it should be reorganized on every time step. But grid modification involves reinterpolation of the mesh functions with respect to mesh structure. Utilization of the rectangular Eulerian meshes allows to avoid this problem. A tree-dimensional model of collapsing star in rectangular coordinate system was proposed in [5].

1.2 TVD schemes

TVD-type schemes of the first and second order of accuracy are considered in this article. First order accurate difference schemes retain the property of monotonicity, but lead to the smearing of the shock fronts. Second order accurate nonlinear schemes with the diminishing of total variation allow to carry out calculations of high resolution and to prevent nonphysical oscillations beyond shock wave fronts. Schemes of this type are of different order of accuracy in the domains with steep and low gradients. Application of these schemes in tree-dimensional case produces especially good results while simulating collapsing stars.

Equations that govern hydrodynamic motion are conservation laws for mass, momentum, [6] and energy. The conservation form of hydrodynamic equations in terms of Eulerian coordinate system is the following:

$$\frac{\partial\rho}{\partial t} + \frac{\partial\rho}{\partial x_i}(\rho v_i) = 0, \tag{1}$$

$$\frac{\partial\rho v_i}{\partial t} + \frac{\partial}{\partial x_i}(\rho v_i v_j + P\delta_{ij}) = 0, \tag{2}$$

$$\frac{\partial e}{\partial t} + \frac{\partial}{\partial x_i}[(e+P)v_i] = 0. \tag{3}$$

The influence of gravitational field is omitted in equations (1) - (3) as well as the action of other sources of energy, for example, neutrino radiation. The equation of state may be written as follows

$$P = (\gamma - 1)\varepsilon, \tag{4}$$

here ρ is the density, v is the vector of speed and P is the pressure, besides that the total energy is $e = \frac{1}{2}\rho v^2 + \varepsilon$.

A TVD scheme was applied to the equations (1) - (3) in [7, 8]. A common restriction of oscillations is a nonlinear condition of stability. The discrete solution for TVD scheme may be defined in the following way

$$TV(u^t) = \sum_{i=1}^{N} |u^t_{i+1} - u^t_i| \tag{5}$$

as a measure of total amount of oscillations.

Thus using second order accurate fluxes $F^{(2)t}_{i+1/2}$ across cells boundaries a nonlinear TVD scheme may be presented in another way. Second order fluxes are derived from first order accurate fluxes $F^{(1)t}_{i+1/2}$ for the upwind scheme applying second order accurate correction. First order accurate flux is obtained, in turn, from the flux mean values. Second order accurate correction is introduced in order to bound spurious oscillations. Hence the number of oscillations on the current time step must not exceed the number of oscillations on the previous one. $TV(u_{i+1}) \leq TV(u_i)$.

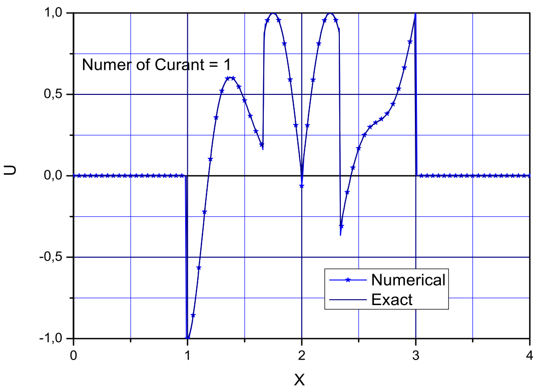

Figure 1. TVD scheme using **vanLeer** flux limiter (stared line) in comparison with analytical solution (solid line)

Different flux limiters are used in order to limit oscillations , specifically, **minmod**, **superbee**, **vanLeer**. The former limiter chooses the smallest absolute value from between the left and right corrections:

$$minmod(a,b) = \frac{1}{2}[sign(a) + sign(b)]\ min(|a|,|b|). \tag{6}$$

The **superbee** limiter choses between the larger correction and 2 times the smallest correction, whichever is smaller in magnitude

$$superbee(a,b) = \begin{cases} minmod(a,2b), & if|a| \geq |b|, \\ minmod(2a,b), & if|a| < |b|. \end{cases} \tag{7}$$

The **vanLeer** limiter is the most moderate of all limiters and finds a harmonic mean between left and right corrections

$$vanleer(a,b) = \frac{2ab}{a+b}.$$

The test, proposed in [7], was used for checking the obtained computer program

$$u_0 = \begin{cases} -x\sin(\frac{3}{2}\pi x^2, & -1 \le x < -\frac{1}{3}, \\ |\sin(2\pi x)|, & |x| < \frac{1}{3}, \\ 2x - 1 - \frac{1}{6}\sin(3\pi x), & \frac{1}{3} < x < 1. \end{cases} \quad (8)$$

The solution obtained by TVD scheme with the **vanLeer** limiter (stared line) is presented in Figure 1. The analytical solution (solid line) is included for comparison. The Courant – Friedrich's – Levy number CFL = 1. A close agreement between numerical and analytical solutions should be noted [8].

1.3 Equations of gravitational gas dynamics

The solution of equations of gravitational gas dynamics that describes the collapsing star may be written in the following way

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x_i}(\rho v_i) = 0, \quad (9)$$

$$\frac{\partial \rho v_i}{\partial t} + \frac{\partial}{\partial x_i}(\rho v_i v_j + P\delta_{ij}) = -\rho \frac{\partial \phi}{\partial x_i}, \quad (10)$$

$$\frac{\partial e}{\partial t} + \frac{\partial}{\partial x_i}[(e + P)v_i] = -\rho v_i \frac{\partial \phi}{\partial x_i}. \quad (11)$$

The value of gravitational potential ϕ is defined from Poisson equation: $\triangle\phi = 4\pi G\rho$. Equation of state is used in the form of (4). In the equations from above ρ - density, v - field of velocities, P - pressure, ε - specific internal energy, e - total energy:

$$e = \frac{1}{2}\rho v^2 + \varepsilon\ . \quad (12)$$

TVD scheme testing was accomplished for the Sedov-Taylor test-problem of point explosion. For this purpose computational domain was defined in the form of a cube with 128 cells. The cube domain is filled in with the medium of constant density ρ_1 while the pressure is a negligible quantity. A high energy deposition takes place at the moment

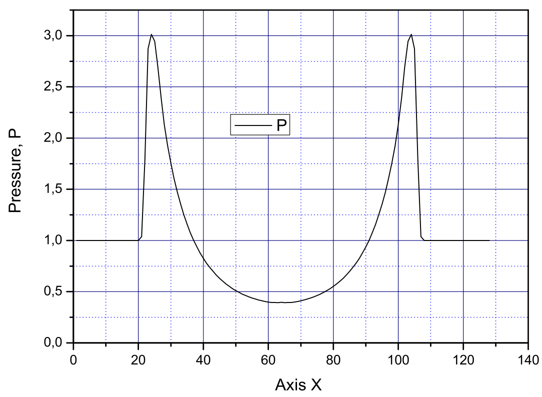

Figure 2. Pressure profile for the Sedov–Taylor test problem

t=0 in the center of the computational domain. Pressure profile is plotted in the Figure 2 at the moment t = t*. A good coincidence of numerical and analytical results has to be mentioned.

1.4 The main results

Let us consider the case of interaction of two shocks. Two sources of energy are placed in the center of a cube for this purpose. Instantaneous energy production takes place in the start time and the explosion is of the same yield as in the previous section. The complexity of this test consists in the necessity of an accurate computation of interaction of two shock waves. This test is more often used in astrophysical computations as a basis of supernova explosion simulating.

Pressure profile is plotted in the Figure 3 for the problem (1)-(3). The initial density and energy are respectively: $\rho_0 = 1.0$ and $E = 10^5$. The problem was solved for the case of rectangular coordinate system which is not invariant with respect to the rotation. However, the non

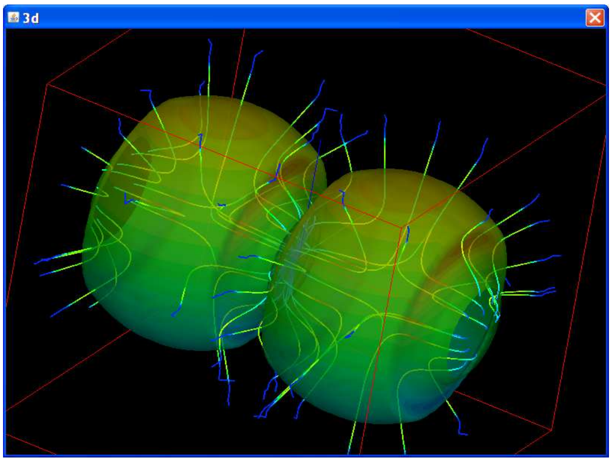

Figure 3. Pressure distribution for the case of two interacting shocks.

isotropic dispersion is not large. One can observe that the numerical solution has spherically symmetric form. Shock wave resolution is of two space cells dimension. Numerical scheme testing convinced us that TVD scheme may be used in solving supernova explosion problems.

Adaptive mesh techniques are currently being used for the improving of the accuracy of numerical calculations as well as of the algorithm efficiency. The methods of this type allow to reduce the computing time and to narrow the volume of employed memory. These techniques are especially effective in solving the problems of gas dynamics characterized by apparition of compression waves, shock waves and contact discontinuities. The use of adaptive meshes makes it possible to investigate the processes with a desirable degree of accuracy in complex geometry domains or steep gradients. AMR method allows to decrease the number of cells and therefore the time of computing. AMR technology is based on use of cells hierarchical structure. In this case every level of the hierarchy is referred to its level of spacial and time resolution. The possibility to add cells to a fixed place of computational domain locally and dynamically is the characteristic property of AMR methods. An algorithm for the refinement of the mesh on several levels with consecutively diminishing space steps is proposed in this work.

Nested meshes were used for solving three-dimensional Poisson equation, for example in [10, 11]. The density of collapsing star varies in many degrees. The density on the surface is not large but in the center the order of density increases up to $10^{14} g/cm^3$. Nested and refined meshes were built in order to take account on such enormous variation of density. In the center of computational domain a cube with the size of cells in 2^3 times smaller than the initial size of the cells was extracted. In the center of specified cube another cube was constructed with smaller dimension of cells. Dimensions of the nested cube were equal to M^3, here the value of M is varying from 64 up to 1024 cells. The solution of Poisson equation was found with the help of successive over-relaxation method. The density profile and the particles paths for two interacting shock waves are plotted in the Figure 4. Calculations were carried out on the 1024x1024x1024 mesh.

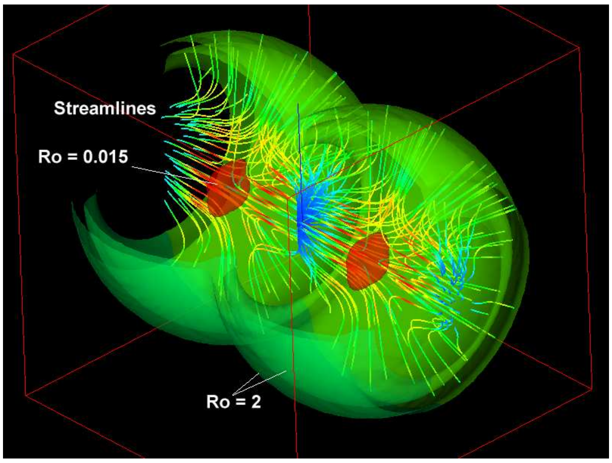

Figure 4. Distribution of the density and the particles paths for the 1024x1024x1024 mesh.

A parallel algorithm for solving Poisson hydrodynamic equations was constructed [11]. The algorithm efficiency is the highest one for 8-12 processors but for the greater number of processors the loss of efficiency is observed. Calculations have been performed on the high performance computing cluster of the Institute of Mathematics and Computer Science of the Academy of Sciences of Moldova.

1.5 Summary

A parallel algorithm and a code for three-dimensional gravitational gas dynamic equations were provided in this work. For this purpose a TVD scheme possessing high resolution in the regions of shock fronts and steep gradients was used. Numerical calculations obtained on the sequence of nested meshes have been presented. Calculations were implemented on the meshes from 64x64x64 to 1024x1024x1024 nodes up to 5 nesting levels. It was demonstrated that the algorithm is quite efficient for 8 – 12 processors.

The results of computer modeling obtained in this work were visualized with the help of HDVIS program [12].

This article has been written under the support of the grant RFFI-Moldova (Space Research Institute of the Academy of Sciences of Russia – Institute of Mathematics and Computer Science of the Academy of Sciences of Moldova) 08.820.06.40 RF.

References

[1] Bisnovatyi-Kogan G.S. 1970, Astron. Zh. 47, 813.

[2] Ardeljan N.V., Bisnovatyi-Kogan G.S., Kosmachevskii K.V., Moiseenko S.G. 2004, Astrophysics 47, 1

[3] Ardeljan N.V., Bisnovatyi-Kogan G.S., Moiseenko S.G., 2005, MNRAS 359, 333

[4] Leblanc L.M., Wilson J.R. 1970, ApJ. 161, 541.

[5] Bisnovatyi-Kogan G.S., Moiseenko S.G., Rybakin B.P., Secrieru G.V. *Modeling of explosive magnetorotational phenomena: from 2D to 3D.* Buletinul ACM, Matematica, N.2(42), 2007, pp. 113–122.

[6] Landau L. D., Lifshitz E. M. *Fluid Mechanics.* 1987, 733 .

[7] Harten A., J. Comp. Phys. **49**, (1983), pp. 357–365.

[8] Rybakin B. *Numerical methods for multiprocessor computers.* Chishinau, Publishing house oldSU, 338 .

[9] Hy Trac, Ue-Li Pen. *A Moving Frame Algorithm for High Mach Number Hydrodinamics.* arXiv:asrto-ph/0309599v2 24 Sep 2003.

[10] Rybakin B. *Construction of a Parallel TVD Method for the Solution of the III-D Hydrodynamical Equations.* Conference Mathematics and Information Technologies: Research And Education, Chisinau, October 1-4, 2008, pp. 41–42.

[11] Rybakin B. *Parallel implementation for solving three – dimensional Poisson equation on structured adaptive meshes in order to simulate supernova explosion.* International conference "Scientific Service in the INTERNET. Technologies of parallel computing", Abrau-Durso, september 18-23, 2007 , pp 73–75.

[12] Goreacheov V. D., Smirnov E. M., Balashov M. E. *Visualization of the numerical experiment for nonsteady flow modeling with large volume of data in the HDVIS system*, Research Service in Internet, Solution of large problems, Proceedings of Russian Scientific Conference (September 22-27, 2008, Novorossiysk), Publishing House of Moscow State University, Moscow 2008, pp. 55–59

Boris Rybakin, Natalia Shider,

Received June 8, 2009

Laboratory of Mathematical Modeling,
Institute of Mathematics and Computer Science,
Academy of Sciences of Moldova,
Moldova, Chishinau, Academy street 5,
E–mail: *rybakin@math.md*
natalia.shider@gmail.com

Computer Science Journal of Moldova, vol.**17**, no.1(49), 2009

Computational Experiments on the Tikhonov Regularization of the Total Least Squares Problem

Maziar Salahi* Hossein Zareamoghaddam

Abstract

In this paper we consider finding meaningful solutions of ill-conditioned overdetermined linear systems $Ax \approx b$, where A and b are both contaminated by noise. This kind of problems frequently arise in discretization of certain integral equations. One of the most popular approaches to find meaningful solutions of such systems is the so called total least squares problem. First we introduce this approach and then present three numerical algorithms to solve the resulting fractional minimization problem. In spite of the fact that the fractional minimization problem is not necessarily a convex problem, on all test problems we can get the global optimal solution. Extensive numerical experiments are reported to demonstrate the practical performance of the presented algorithms.

Keywords: Linear systems, Total least squares, Tikhonov regularization, Newton method, Bisection method.

1 Introduction

In this paper we aim to find meaningful solutions for the linear systems of the form

$$Ax \approx b, \tag{1}$$

where $A \in R^{m \times n}$, $b \in R^m$, $m \geq n$ are both contaminated by noise. This kind of systems frequently arise in discretization of certain integral equations [3].

* This author is a corresponding author

If A is ill-conditioned, a quite effective procedure to find a reasonably good solution for (1) is to use the regularized least squares approach. Perhaps the best known regularization technique is due to Tikhonov [6], which solves the following minimization problem rather than classical least squares one:

$$\min_{x \in R^n} \|Ax - b\|^2 + \rho \|x\|^2 , \tag{2}$$

where ρ is a positive constant.

It is worth mentioning that when $\rho = 0$ (the classical least squares approach), and A is ill-conditioned, then the solution of (2) might have large norm, while for positive ρ it is not the case. It is not easy to find the exact value of ρ, however there have been some studies on this subject [5]. It is obvious that (2) is a convex minimization problem and its optimal solutions should satisfy

$$(A^T A + \rho I)x = A^T b. \tag{3}$$

We may use existing efficient iterative algorithms like conjugate gradient methods to solve (3).

Another most popular approach to deal with such systems is the so called total least squares problem [1, 4]. This approach leads to a fractional nonconvex minimization problem. In this paper we present three efficient algorithms to solve it. Extensive numerical results are reported to show the efficiency of the discussed algorithms.

2 Total least squares problem

In this approach, one aims to find a feasible system by minimal changes in problem data i.e.,

$$\begin{aligned} \min_{x,E,r} \quad & \|E\|^2 + \|r\|^2 \\ & (A + E)x = b + r. \end{aligned} \tag{4}$$

The optimal E and r values are given in the following theorem.

Theorem 1. *The optimal E and r values of problem (4) are given by*

$$r^* = \frac{Ax^* - b}{1 + \|x^*\|^2}, \ E^* = -\frac{Ax^* - b}{1 + \|x^*\|^2} x^{*T}$$

where x^* is the optimal solution of

$$\min_{x \in R^n} \frac{\|Ax - b\|^2}{1 + \|x\|^2}. \tag{5}$$

Proof. The minimization problem in (4) can be written as two minimization problems as follows:

$$\begin{aligned} \min_{x \in R^n} \min_{E,r} \quad & \|E\|^2 + \|r\|^2 \\ & (A + E)x = b + r. \end{aligned}$$

Let us first consider the inner minimization problem. Obviously it is a convex optimization problem, therefore the KKT conditions are necessary and sufficient for optimality that follows:

$$\begin{aligned} & 2E^* + \lambda^* x^T = 0 \\ & 2r^* - \lambda^* = 0 \\ & (A + E^*)x - b - r^* = 0, \end{aligned} \tag{6}$$

where the vector λ^* denotes the lagrange multipliers. From the second equation of (6) we have $\lambda^* = 2r^*$ and subsequently from the first equation we have $E^* = -r^* x^T$. Finally, the last equation implies that

$$r^* = \frac{Ax - b}{1 + \|x\|^2},$$

and subsequently

$$E^* = -\frac{Ax - b}{1 + \|x\|^2} x^T.$$

Now the objective function of inner minimization problem becomes

$$\frac{\|Ax - b\|^2}{1 + \|x\|^2}.$$

Thus if x^* be an optimal solution of this problem, then the proof is completed. □

Therefore, by solving this minimization problem we have a modified linear system which is feasible. Since the original system is ill-conditioned, then the solution of (5) might be meaningless from practical point of view due to the large norm. Thus we can stabilize the solution by utilizing the Tikhonov regularization technique. The regularized problem becomes:

$$\min_{x \in R^n} f(x) := \frac{\|Ax - b\|^2}{1 + \|x\|^2} + \rho \|x\|^2, \tag{7}$$

where ρ is a nonnegative parameter. As it is obvious, problem (7) is not known to be convex or concave in general. In the sequel we present several numerical algorithms, which can help us to solve (7) up to global optimality. First let us derive the gradient and hessian of the objective function of (7) as follows:

$$\nabla f(x) = \frac{2A^T(Ax - b)}{1 + \|x\|^2} - \frac{2\|Ax - b\|^2 x}{(1 + \|x\|^2)^2} + 2\rho x$$

$$\nabla^2 f(x) = \frac{2A^T A}{1 + \|x\|^2} - \frac{4x(A^T(Ax - b))^T}{(1 + \|x\|^2)^2} + 2\rho I - \frac{4A^T(Ax - b)x^T}{(1 + \|x\|^2)^2} + \left(\frac{8xx^T}{(1 + \|x\|^2)^3} - \frac{2}{(1 + \|x\|^2)^2} I\right) \|Ax - b\|^2 .$$

The first approach which we utilize to tackle (7) numerically is the classical Newton method. During this process one might end up with an iterate when the hessian is singular or very close to singularity, but a slight perturbation of it usually resolves this bad behavior. Although the objective function of (7) is not known to be convex, but for most of the test problems we have considered, it yields a global solution as it will be shown in the next section. The structure of the algorithm is as follows:

Newton Based Algorithm

Inputs: An accuracy parameter $\epsilon > 0$;
A regularization parameter ρ;
A parameter δ usually 10^{-4};
A starting point $x_0 \in R^n$.
begin
i=0;
while $\|\nabla f(x_i)\| \geq \epsilon$
Find an appropriate α by Armijo line search and let
$x_{i+1} = x_i - \alpha(\nabla^2 f(x_i) + \delta I)^{-1}\nabla f(x_i)$.
i=i+1;
end
end

In the sequel we present another algorithm by an old idea due to Dinkelbach [2] which uses an equivalent formulation of the problem (7) to solve it. It is obvious that

$$\min_{x\in R^n}\{\frac{\|Ax-b\|^2}{1+\|x\|^2} + \rho\,\|x\|^2\} \leq t$$

is equivalent to

$$\min_{x\in R^n}\{\|Ax-b\|^2 - t(1+\|x\|^2) + \rho(\|x\|^2 + \|x\|^4)\} \leq 0. \tag{8}$$

Now let us define

$$\Phi(t) = \min_{x\in R^n}\{\|Ax-b\|^2 - t(1+\|x\|^2) + \rho(\|x\|^2 + \|x\|^4)\}.$$

Lemma 1. *The function $\Phi(t)$ is a strictly decreasing function.*

Proof. Let $t_1 < t_2$ and x_{t_1} be the point for which

$$\Phi(t_1) = \|Ax_{t_1} - b\|^2 - t_1(1+\|x_{t_1}\|^2) + \rho(\|x_{t_1}\|^2 + \|x_{t_1}\|^4).$$

Then we have

$$\Phi(t_1) > \|Ax_{t_1} - b\|^2 - t_2(1+\|x_{t_1}\|^2) + \rho(\|x_{t_1}\|^2 + \|x_{t_1}\|^4) \geq \Phi(t_2).$$

Therefore, $\Phi(t_1) > \Phi(t_2)$. □

We further have that $\Phi(0) > 0$ and

$$\Phi(\|b\|^2) \leq \|A0 - b\|^2 - \|b\|^2 (1 + \|0\|^2) + \rho(\|0\|^2 + \|0\|^4) = 0$$

Therefore function $\Phi(t)$ has a unique root in the interval $[0, \|b\|^2]$. Now our goal is to find this root. First in the Next lemma we prove that this gives us the global minimum of (7). Then explain how to find the root numerically.

Lemma 2. *The root of the function $\Phi(t)$ gives the global minimum of problem (7).*

Proof. Let t^* be the root of $\Phi(t)$. Then

$$\min_{x \in R^n} \{\|Ax - b\|^2 - t^*(1 + \|x\|^2) + \rho(\|x\|^2 + \|x\|^4)\} = 0.$$

Let x^* be the point on which this minimum happens. Then for any $x \in R^n$ one has

$$\|Ax - b\|^2 - t^*(1 + \|x\|^2) + \rho(\|x\|^2 + \|x\|^4) \geq 0,$$

or

$$\frac{\|Ax - b\|^2}{1 + \|x\|^2} + \rho \|x\|^2 \geq t^*.$$

This further implies that

$$\min_{x \in R^n} \{\frac{\|Ax - b\|^2}{1 + \|x\|^2} + \rho \|x\|^2\} \geq t^*,$$

but at least we know that equality holds when $x = x^*$. Thus x^* is the global minimum of (7). □

Now, to find the root of function $\Phi(t)$ we utilize the bisection algorithm to reduce the initial interval $[0, \|b\|^2]$ and also the classical Newton method to solve the corresponding minimization problem. Similar to the fractional case, here also we do not know whether the objective function is convex or not. But this simple procedure leads us to the global minimum for all test problems.

However, as we are aware, the bisection method is usually too slow, so the third approach which we consider is as follows. First we perform a few iterations of bisection algorithm, then crossover to formulation (7) rather than finding the root of function Φ. This combined algorithm finds the global solution much faster compared to both previous algorithms. In the next section we report extensive numerical testing which demonstrates the practical performance of the three presented algorithms.

3 Computational experiments

Test problems in Tables 1 and 2 are taken from [5] which contains ill-posed linear systems arising from certain integral equations, and problems on Table 3 are taken from University of Florida sparse matrix collection. The implementation of the algorithms are done in MATLAB 7.4 on a pentium M 1.7GHz laptop with 1 GB of memory. All test problems are square, however they can easily be made overdetermined by repeating some constraints with slightly different right hand side and still the same observations, which will be given in the sequel, hold.

For all test problems the coefficient matrices are either singular or very close to singularity. Moreover, for problems in Tables 1 and 2 we have the exact solution and for problems in Table 3 we consider all one vector as the exact solution. Furthermore the noisy system is generated by perturbing A and b by adding '$1e-3*randn(size(A))$' and '$1e-3*randn(size(b))$' respectively. Since the coefficient matrix is singular or very close to singularity, then either system $Ax=b$ is infeasible or its solution might have very large norm. Therefore the total least squares approach is utilized to find an appropriate feasible system with a meaningful solution.

In all tables x_s and x^* denote the exact and computed solution of problems, respectively and $\|Ax^*-b\|$ denotes the violation of the computed solution from the original system. The numbers in all parenthesis are for the classical Newton method, bisection method, and bisection-Newton method (crossover) respectively. For all test problems we have

Table 1. **Comparison of Newton, bisection-Newton and crossover algorithms**

problem	ρ	$\|Ax^* - b\|$	$\|x^*\|$	$\|x_s\|$	time
baart-100	0.001	(70.155, 0.0093, 0.0093)	(20.844, 1.2016, 1.2016)	1.2533	(0.65,0.8,0.45)
	0.1	(0.1690, 0.1690, 0.1690)	(1.0143, 1.0143, 1.0143)	1.2533	(0.5,0.9,0.5)
	1	(0.5223, 0.5223, 0.5223)	(0.7929, 0.7929, 0.7929)	1.2533	(0.18,0.95,0.55)
	10	(1.4873, 1.4873, 1.4873)	(0.4485, 0.4485, 0.4485)	1.2533	(0.16,1.2,0.65)
baart-500	0.001	(70.162, 0.0084, 0.0084)	(20.846, 1.2036, 1.2036)	1.2533	(13,18.5,9.6)
	0.1	(0.1678, 0.1678, 0.1678)	(1.0154, 1.0154, 1.0154)	1.2533	(8.9,21.2,13.4)
	1	(0.5216, 0.5216, 0.5216)	(0.7932, 0.7932, 0.7932)	1.2533	(3.2,21.5,14)
	10	(1.4869, 1.4869, 1.4869)	(0.4485, 0.4485, 0.4485)	1.2533	(3.2,23.6,9.6)
baart-1000	0.001	(70.1654, 0.0075, 0.0075)	(20.8443, 1.2068, 1.2068)	1.2533	(229,131,62)
	0.1	(0.167, 0.167, 0.167)	(1.0161, 1.0161, 1.0161)	1.2533	(25.7,153.2,69.8)
	1	(0.5212, 0.5212, 0.5212)	(0.7937, 0.7933, 0.7933)	1.2533	(22.4,155.4,72.5)
	10	(1.4867, 1.4867, 1.4867)	(0.4486, 0.4486, 0.4486)	1.2533	(19,163,77)
heat-100	0.001	(0.7048, 0.047, 0.047)	(0.7279, 1.739, 1.739)	2.4623	(0.13,0.9,0.3)
	0.1	(0.2636, 0.2636, 0.2636)	(0.6915, 0.6915, 0.6915)	2.4623	(0.1,0.27,0.12)
	1	(0.42081, 0.42081, 0.42081)	(0.1519, 0.1519, 0.1519)	2.4623	(0.1,0.26,0.1)
	10	(0.46329, 0.46329, 0.46329)	(0.0148, 0.0148, 0.0148)	2.4623	(0.1,0.25,0.12)
heat-500	0.001	(3.6139, 0.1831, 0.1831)	(7.5382, 3.3458, 3.3458)	5.5034	(5.1,20.4,10)
	0.1	(0.6448, 0.6448, 0.6448)	(1.3441, 1.3441, 1.3441)	5.5034	(3.9,23.3,9)
	1	0.8974, 0.8974, 0.8974)	(0.4768, 0.4768, 0.4768)	5.5034	(3.5,23,11)
	10	(1.0335, 1.0335, 1.0335)	(0.03618, 0.03618, 0.03618)	5.5034	(3.2,21,10)
heat-1000	0.001	(1.5452, 0.3286, 0.3286)	(2.1473, 4.3463, 4.3463)	7.7829	(30,151.5,63)
	0.1	(0.9633, 0.9633, 0.9633)	(1.7141, 1.7141, 1.7141)	7.7829	(34,181.5,63.5)
	1	(1.2513, 1.2513, 1.2513)	(0.7321, 0.7321, 0.7321)	7.7829	(28.2,165.8,65.2)
	10	(1.459, 1.459, 1.459)	(0.0578, 0.0578, 0.0578)	7.7829	(22.7,157,77)

Table 2. **Comparison of Newton, bisection-Newton and crossover algorithms**

problem	ρ	$\|Ax^* - b\|$	$\|x^*\|$	$\|x_s\|$	time
shaw-100	0.001	(0.5862, 0.5862, 0.5862)	(9.3802, 9.3802, 9.3802)	9.982	(0.18,0.28,0.17)
	0.1	(6.7763, 6.7763, 6.7763)	(6.093, 6.093, 6.093)	9.982	(0.1,0.3,0.15)
	1	(12.128, 12.128, 12.128)	(3.985, 3.985, 3.985)	9.982	(0.1,0.33,0.18)
	10	(16.704, 16.704, 16.704)	(2.3207, 2.3207, 2.3207)	9.982	(0.1,0.35,0.15)
shaw-500	0.001	(224.4, 3.9035, 3.9035)	(58.085, 19.253, 19.253)	22.32	(6,25.9,14.2)
	0.1	(23.265, 23.265, 23.265)	(10.37, 10.37, 10.37)	22.32	(2.7,27.6,17.3)
	1	(33.95, 33.95, 33.95)	(6.409, 6.409, 6.409)	22.32	(3.2,27.8,18.7)
	10	(41.504, 41.504, 41.504)	(3.715, 3.715, 3.715)	22.32	(3.6,31,0.14.2)
shaw-1000	0.001	(270.21, 8.1311, 8.1311)	(66.423, 25.695, 25.695)	31.566	(58.3,172.1,95.5)
	0.1	(37.78, 37.78, 37.78)	(12.82, 12.82, 12.82)	31.566	(22.5,193,118.6)
	1	(51.57, 51.57, 51.57)	(7.7873, 7.7873, 7.7873)	31.566	(25.3,199.5,125.5)
	10	(60.833, 60.833, 60833)	(4.5, 4.5, 4.5)	31.566	(28.60,207.6,101.7)
spike-100	0.001	(1.2306, 1.2306, 1.2306)	(14.121, 14.121, 14.121)	29.017	(0.2,0.3,0.15)
	0.1	(17.211, 17.211, 17.211)	(11.664, 11.664, 11.664)	29.017	(0.1,0.33,0.2)
	1	(43.986, 43.986, 43.986)	(8.4666, 8.4666, 8.4666)	29.017	(0.1,0.34,0.23)
	10	(72.392, 72.392, 72.392)	(5.4189, 5.4189, 5.4189)	29.017	(0.1,0.39,0.23)
spike-500	0.001	(1.3441, 1.3441, 1.3441)	(22.706, 22.706, 22.706)	34.67	(17,25.9,14.8)
	0.1	(33.24, 33.24, 33.24)	(21.207, 21.207, 21.207)	34.670	(6.3,32,19)
	1	(138.54, 138.54, 138.54)	(17.723, 17.723, 17.723)	34.670	(9.1,29.5,20.5)
	10	(319.12, 319.12, 319.12)	(12.607, 12.607, 12.607)	34.670	(4.4,29.3,18.4)
spike-1000	0.001	(1.9066, 1.9066)	(30.877, 30.877, 30.877)	40.645	(580,210,124)
	0.1	(51.536, 51.632, 51.536)	(29.536, 29.536, 29.536)	40.645	(26,205.9,123.7)
	1	(252.78, 252.78, 252.78)	(25.833, 25.833, 25.833)	40.645	(48.3,215,156.2)
	10	(678.16, 678.16, 678.16)	(19.205, 19.205, 19.205)	40.645	(20,220.2,152)

Table 3. **Comparison of Newton, bisection-Newton and crossover algorithms**

problem	ρ	$\|Ax^* - b\|$	$\|x^*\|$	$\|x_s\|$	time
can187	0.001	(0.4187, 0.4187, 0.4187)	(13.6, 13.6, 13.6)	13.675	(0.31,1.4,0.85)
	0.1	(16.399, 16.399, 16.399)	(11.495, 11.495, 11.495)	13.675	(0.14,0.9,0.5)
	1	(42.695, 42.695, 42.695)	(8.236, 8.236, 8.236)	13.675	(0.15,1,0.7)
	10	(67.983, 67.983, 67.983)	(5.159, 5.159, 5.159)	13.675	(0.12,0.95,0.6)
can268	0.001	(1.1281, 1.1281, 1.1281)	(16.014, 16.014, 16.014)	16.371	(8,2.4,1.4)
	0.1	(19.266, 19.266, 19.266)	(13.777, 13.777, 13.777)	16.371.	(0.35,2.7,1.8)
	1	(61.559, 61.559, 61.559)	(10.6, 10.6, 10.6)	16.371	(0.67,2.2,1.4)
	10	(112.79, 112.79, 112.79)	(6.92, 6.92, 6.92)	16.371	(0.4,2.4,1.5)
can445	0.001	(1.2475, 1.2475, 1.2475)	(20.892, 20.892, 20.892)	21.095	(6.3,7,4.3)
	0.1	(39.62, 39.62, 39.62)	(16.336, 16.336, 16.336)	21.095	(1,6.3,3.7)
	1	(84.947, 84.947, 84.947)	(11.118, 11.118, 11.118)	21.095	(1.2,5.8,3.8)
	10	(122.60, 122.60, 122.60)	(6.8154, 6.8154, 6.8154)	21.095	(1.5,7.5,4.7)
cavity07	0.001	(7.0198, 7.0198, 7.0198)	(22.403, 22.043, 22.403)	34.38	(16.7,103,65)
	0.1	(25.558, 25.558, 22.558)	(10.563, 10.563, 10.563)	34.38	(16.4,105.6,64.9)
	1	(36.13, 36.13, 36.13)	(6.7589, 6.7589, 6.7589)	34.38	(20.7,74.6,51.3)
	10	(47.42, 47.42, 47.42)	(4.1722, 4.1722, 4.1722)	34.38	(20.5,123.8,87.5)
ex21	0.001	(8.3144, 0.09386, 0.09386)	(12.006, 6.7722, 6.7722)	25.612	(7.6,21,11)
	0.1	(2.0441, 2.0441, 2.0441)	(2.619, 2.619, 2.619)	25.612	(5,22.4,11)
	1	(2.4770, 2.4770, 2.4770)	(1.3508, 1.3508, 1.3508)	25.612	(6,18.9,7.5)
	10	(2.8593, 2.8593, 2.8593)	(0.3089, 0.3089, 0.3089)	25.612	(5,24.3,11.9)
dwt869	0.001	(3.3016, 3.3016, 3.3016)	(28.916, 28.916, 28.916)	29.479	(7.6,43.7,27)
	0.1	(72.326, 72.326, 72.326)	(20.557, 20.557, 20.577)	29.479	(6.8,47.3,30.5)
	1	(133.26, 133.26, 133.26)	(13.49, 13.49, 13.49)	29.479	(6.6,28.5,18)
	10	(179.94, 179.94, 179.94)	(8.1496, 8.1496, 8.1496)	29.479	(7.5,51,34.3)
gd01a	0.001	(1.0654, 1.0654, 1.0654)	(12.541, 12.541, 12.541)	30.871	(11,278.4,27.7)
	0.1	(12.323, 12.323, 12.323)	(8.6939, 8.6939, 8.6939)	30.871	(9.4,370,99)
	1	(24.636, 24.636, 24.636)	(6.0131, 6.0131, 6.0131)	30.871	(15.2,265.9,210)
	10	(36.958, 36.958, 36.958)	(3.7135, 3.7135, 3.7135)	30.871	(11.7,420,30.9)
gd00a	0.001	(1.4398, 1.4398, 1.4398)	(12.69, 12.69, 12.69)	18.762	(1.1,10,1.2)
	0.1	(10.985, 10.985, 10.958)	(7.2358, 7.2358, 7.2358)	18.762	(0.8,11.6,2.2)
	1	(16.768, 16.768, 16.768)	(4.5229, 4.5229, 4.5229)	18.762	(1.4,13,7)
	10	(21.304, 21.304, 21.304)	(2.6177, 2.6177, 2.6177)	18.762	(0.9,14,2)

used $10 * ones(n, 1)$ as the starting point[1] with four different values of the ρ parameter. Having prior information of the solution also indeed is suggested to be incorporated as the starting point selection procedure.

As our computational results show, the classical Newton method solves all problems for $\rho = 0.1, 1, 10$ faster than the other two approaches, however it fails for many problems when $\rho = 0.001$. It is worth to note that by changing the starting point to for example $100 * ones(n, 1)$ Newton method solves some of the failed problems. However, the other two approaches successfully solve all problems for all ρ values up to global optimality. Therefore, based on these computational results we may conclude that for smaller ρ values the later two approaches are preferred to Newton algorithm, specially the crossover approach, otherwise the Newton algorithm seems to find the global solution much faster.

4 Conclusions

In this paper, first we have introduced the total least squares problem to deal with approximate feasible linear systems. Then three numerical algorithms are presented to solve the resulting fractional minimization problem. Finally, several numerical examples are presented to demonstrate the practical efficiency of the presented algorithms.

References

[1] A. Beck, A. Ben-Tal. *On the solution of the Tikhonov regularization of the total least squares problem.* SIAM J. On Opt., vol. 17 (2006), pp. 98–118.

[2] W. Dinkelbach. *On nonlinear fractional programming.* Management Sci., vol. 131 (1967), pp. 492–498.

[1]For problem 'spike1000' with $\rho = 0.001$ we could not solve it even by bisection method, however by $100 * ones(n, 1)$ as the starting point the solution is given in Table 2.

[3] H.W. Engl. *Regularization methods for the stable solutions of inverse problems.* Surveys Math. Indust., vol. 3 (1993), pp. 71–143.

[4] G.H. Golub, P.C. Hansen, D.P. O'Leary. *Tikhonov regularization and total least squares.* SIAM J. Matrix Anal. Appli., vol. 21 (1999), pp. 185–194.

[5] P.C. Hansen. *Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems,* Numerical Algorithms, vol. 6 (1994), pp. 1–35.

[6] A.N. Tikhonov. *Solution of incorrectly formulated problems and regularization method.* Soveit Math. Dokl., vol. 4 (1963), pp. 1036–1038.

Maziar Salahi, Hossein Zaremoghaddam, Received December 22, 2008

Maziar Salahi,
Department of Mathematics,
Faculty of Sciences,
University of Guilan,
Rasht, Namjoo Street, Guilan, P.O.Box 1914
E–mail: *salahim@guilan.ac.ir*

Hossein Zareamoghaddam,
Department of Mathematics,
Faculty of Sciences,
University of Guilan,
Rasht, Namjoo Street, Guilan, P.O.Box 1914
E–mail: *zareamoghaddam@yahoo.com*

Computer Science Journal of Moldova, vol.17, no.1(49), 2009

Conceptual issues in development of telemedicine in the Republic of Moldova

I. Ababii, C. Gaindric, O. Lozan, I. Brinister

Abstract

The article discusses a concept of development of telemedicine in the Republic of Moldova that determines the role and place of telemedicine in the structure of health services and sets priority development directions, considering the processes of development of the Republic of Moldova and its objective of integration with the European Union.

In the article we refer to telemedicine as to the use of information technology to deliver medical services and information from one location to another [1].

1 Background

The process of information society development in the Republic of Moldova bases on the National e-Strategy for Building an Information Society "Electronic Moldova", approved by the Government Decision nr. 255 from March 9, 2005. In the framework of the Strategy, e-Health chapter refers to the need of telemedicine implementation in health care system.

During last years, a strong growth in utilization of ICT in health care system is observed. A range of management information systems for primary care, health care insurance administration, a National Automated Information System "State register of medicines", management information systems for blood transfusion service (including the National register of donors and the informational register of blood products), and for monitoring and evaluation of the National programme on control and prevention of tuberculosis were developed.

Automation of medical information and data circulation is being on an initial stage of development.

Presently, pilot telemedical projects are implemented in the area of perinatology and neurology. Telemedical videoconferencing and Web-based education support are applied in distant learning programmes for health professionals. There is some trial experience in conducting international telemedical consultations.

Recently, a telemedical link for consultation of neurology patients was implemented between the Neurology and Neurosurgery Institute in Chisinau, and neurology department in the municipal hospital in Balti. Using a videoconferencing link and a hospital management suite implemented in the Institute, neurology specialists from Balti can conduct real-time consultations with high class specialists of a leading republican institution. The implemented neurology telemedical system is expected to considerably increase the quality of care provided to neurology patients, especially in emergency cases, and become a lighthouse experience for other health care institutions in Moldova.

At the international level, implementation and utilization of telemedical services frames into a general effort of modernization of health services with the purpose of improving quality and accessibility of them.

Telemedicine optimizes utilization of health care systems resources and reduces the costs of treatment and care by improving communication among providers and access to health information for patients. Telemedical distant learning environments help to improve training of doctors and nurses, and educate and manage patients with chronic conditions. Calculations suggest that presently, about 0.1% of potential telemedicine demand is met in developing countries [2].

A EU report covering wider area of e-Health [3] outlined that the economic impact of all ten sites participated in the Study on Economic Impact of e-Health was positive, and on average it took four years to reach the level of benefits prevailing over costs. Utilization of e-Health solutions, including telemedical ones, grew, in some cases, exponentially, and showed a steady growth over a longer period of time, in others. While the economic impact of e-Health benefits surged by ten times between 1994 and 2004 from €20 million to €200 million per year, the associated costs remained stable, and did not exceed €100 million per year. In all e-Health applications a prominent impact was observed over the effect of the application on timeliness, effectiveness and efficiency of health service.

WHO continuously supports the efforts of the member states in developing telemedicine by providing assistance in identification of priorities, elaboration of e-Health and telemedical policy documents, consolidation of the legal, normative and ethical base in the area of health information utilization, dissemination of best practices and facilitation of implementation for National technical programmes.

In the Regulation COM (2008) 689 from 4.11.2008 on telemedicine for the benefit of patients, healthcare systems and society, the European Commission underlines the importance of telemedicine, and, for better implementation of telemedical services offers its member-states large facilities for building the confidence and acceptance of telemedical services, introduction of legal clarity, solution of technical issues of compatibility and standardization, and facilitation of market relation in the area.

During 2009-2011, EU member-states are required to complete baseline evaluations and develop national regulation regarding access to telemedical services, including accreditation, jurisdiction and professional accountability, service reimbursement methods, confidentiality and security of data.

Best international telemedical practices include both telemedical services improving access of patients to health services in hard-to-reach areas and in critical conditions, and telemedical services targeting improvement of quality of health care services, optimization of health care system resources utilization, reduction of individual expenditures of patients in highly populated areas with a better access to ICT.

2 Baseline scenario

Despite the progress in health sector reforms, there is a range of drawbacks and problems related to accessibility and quality of health services, and to organization of health services for the population:

- distribution of health care personnel over the territory of the Republic of Moldova is non-uniform;
- there is a disparity in the level of professional training of health care personnel from rural and urban areas;
- emergency service, due to lack of mobile consulting and diagnostics tools, looses critical time to save lives of patients;
- the quality and volume of health care services provided can not be fully managed under existing practices of health services provision;
- private expenditure for patients from rural areas for visiting municipal and republican health care facilities, are much larger than for inhabitants of municipalities;
- systems for management of chronic conditions, remote monitoring and home care are underdeveloped;
- management of emergency situations, natural disasters and man-caused catastrophes is weak.

The professional level of health care personnel does not correspond to the growing requirements in health care system, and distant professional consulting and training resources are limited.

Continuous professional training of health personnel presently uses the methods that do not fully ensure continuity of professional training, and imply in training process additional considerable side costs (extended absence from work place, cost of travel and accommodation of health care personnel during the training etc.)

ICT infrastructure in health care system is underdeveloped limiting the possibilities for optimization of health data circulation. Existing capacities of health care system provide weak continuity of health information among different care levels.

Access to ICT in professional activity, and IT knowledge among medical personnel is subambient. The level of awareness and acceptance of telemedical services is limited.

Existing instruments for informing the population regarding health, disease prevention and promotion of health lifestyle are not ample for improvement of the population's health, if we consider growing non-communicable and communicable morbidity and mortality, and prevalence of pernicious behaviour and habits observed.

There is a disparity between offer and demand for health services on-line in Moldova. More than a third of Internet users in the Republic of Moldova [4] are pushed to look for health information in the Internet on foreign health information resources. Presence of national health care institutions on the Internet does not exceed 5% [5], and national useful health information resources are limited.

Cooperation in the area among different stakeholders is inadequate and does not allow effective coordination of the integrated development of telemedicine.

International professional and scientific integration of health care institutions and doctors from the Republic of Moldova is insufficient, both on crossborder cooperation between health care facilities and health care personnel, and in accessing international sources of medical data for application in professional and scientific activity.

3 Telemedicine development path

Telemedical services represent a range of secured processes of obtaining, transmission, reception, processing, storing and analysis of medical data and information with further formulation of a diagnosis and recommendations for treatment, or direct provision of health service, or distant learning in health care, using ICT available.

The term telemedicine refers to:

- Telemedical consultations - medical consultation with the purpose of remote diagnosing or treatment by the mean of ICT. A complete clinic case or separate clinic data case be the object of a telemedical consultation [6].

- Telemonitoring/telemetry - a range of telemedical services to remotely monitor and manage health of a patient [7].

- Telemedical distant learning - distant learning using telemedical links (videoconferencing, Web, etc.), including professional training of medical personnel, educating patients and the population.

Introduction of telemedical consultation, tools for patient monitoring, distant learning for medical personnel, patients and the population is expected to contribute to solving of a range of issues in health care sector.

Considering the need for an integrated approach to issues of telemedicine development, social importance and economic impact of them, the issue of infrastructure development for telemedical services should be tackled as well, while developing telemedicine projects and implementing them in clinical practice. Following directions and measures are suggested for development:

- implementation of telemedical consultative services in different level of care (primary, hospital, emergency care);
- creation of remote monitoring and home care services based on ICT for elderly, convalescent patients, for patients with chronic conditions, disabilities, pregnant women and young mothers;
- introduction of modern distant learning methods with application of telemedical technologies as videoconferencing and Web to training of medical personnel;
- stimulation of public health care institutions to increasing connectivity for institutions, equipping the institutions with productive modern and compatible IT equipment and medical appliances with the capacity to obtain, stock, transmit, receive and analyze digital data and images within national telemedical network;
- creation of telemedical service for emergency situations and disasters;
- wider application of Internet and other communication services in public health programmes;

- creation and maintenance of web-pages of health care institutions, through a common effort of the Ministry of Health and health care facilities;

- elaboration of standards and ethical code of conduct for exchange of medical information and provision of telemedical services in the Internet.

Facilitation of international cooperation in the area of telemedicine is especially important. Participation to regional and international telemedical networks and projects, joining international professional associations, facilitation of exchange programmes in telemedicine contributes to knowledge transfer, and improves access of doctors to reliable international health data and medical information.

Information technology underlying telemedicine, offers new possibilities for collaborative work of health professionals from different countries with proficient communication and health data exchange. Accumulated experience shows that development of international cooperation between clinicians by the mean of telemedicine contributes to growth of the quality and accessibility of health care services with simultaneous reduction of costs in many cases.

It can be observed that development directions refer to different domains of health care and ICT. Proper management of coordination and monitoring of the development of telemedical support for health care system is important for optimal utilization of available resources, and maintenance of continuity and integrity of telemedical services. Therewith, efficient coordination helps to better accumulate national experience and knowledge in the area of telemedicine.

Introduction of new market condition in health care increased opportunities for investment attraction, including private, in health care system. Application of private-public partnerships can be one of the means for attracting investment for telemedical services by health care institutions.

4 Evaluation of the impact of telemedicine development for the Republic of Moldova

Implementation of telemedicine in the Republic of Moldova will contribute to:

- approximation of high quality health services to the patients home, including rural and isolated areas;
- improvement of the quality of professional medical training at all levels of care;
- improved informing of the population regarding health, and accessibility of useful public health information.

International experience suggests that along with growing utilization rate of a telemedical service, the cost of the service decreases in comparison with the costs of a similar traditional health care services and training methods. By anticipating higher utilization levels for a telemedical services, an advanced paypack period on investment into telemedicine can be achieved.

It is expected that successful implementation of telemedical consultations will contribute to financial savings in health care system, through extending the range cost-effective health services available at primary and secondary level, and optimization of patient pathways in system.

Telemedical monitoring and homecare services should contribute to prevention of hospitalization and reductionist duration with concomitant savings.

Economic impact of telemedicine in Moldova can be demonstrated on the example of the emergency service. The process of emergency solicitation involves a telephone probe that reduces the number of unjustified solicitations.

Evaluation of the application of videoconferencing to continuous education of medical professionals demonstrated high appreciation of this method of training and its quality. The economic efficiency of telemedical training was found much higher compared to traditional education and training means.

Rapt interest to telemedicine in developed countries is conditioned by the social impact of telemedicine on accessibility and quality of health care services and strengthening health and quality of life for the population.

Telemedical consultations contribute to levelling distribution of health care personnel in the country, and bring specialized health care closer to patients home. Private expenditure of patients on costly travel to republican institutions goes down correspondingly. Application of telemedical consultations can improve access to health services of a range of target groups (poor, pregnant women and young mothers, patients with chronic conditions, disabled, prisoners, other socially vulnerable groups).

Telemedical remote monitoring improves the quality of supervision and treatment of patients in outpatient conditions, prevention and early detection of complications and emergency conditions. Telemedical homecare prevents institutionalization of persons with special needs and improve the quality of their lives in communities.

Improving the quality of health care services is indispensable from the continuous professional training of health care personnel. Introduction of new forms of professional training with application of ICT will contribute to the growth of competences among health care personnel that will help to improve the quality of health care services they provide.

Provision of the access to reliable health information for the population and patients will contribute to increase of personal responsibility for health, accompanying reduction of health risks, and improvement of the national health indicators.

Wide introduction of telemedical services will also influence the transparency of relations between health care provider and patients, simultaneously reducing corruption in health care system.

The possibility of extending the access to cost-effective medical service in rural or isolated areas is an important argument for development of telemedicine in the Republic of Moldova, where more than half of the population lives in rural areas.

Stimulating participation of health institutions from the Republic of Moldova to international telemedical networks should improve mobility both for patients and health professionals from Moldova and other countries, regarding obtaining required health service or getting a consultation of a health profession anywhere, including abroad.

In case the country fails on effective coordination of development processing in telemedicine, telemedical applications will continue developing in a sporadic way, a fact proved by early experiences in many countries. The costs of isolated telemedical projects jumps high, but the results obtained can be non-satisfactory. Most of patients will be deprived from the opportunity to benefit from cost-effective ICT facilities that improve accessibility and quality of health services.

Without wider application of ICT in clinical practice, the costs of health care services will be strongly dependent on the continuous growth of health care personnel reimbursement expenditure, both medical and non-medical.

Utilization of existing methods of continuous professional training will continue training of weak health care personnel, maintaining the inadequacy of health care system capacities.

The process of European integration of the Republic of Moldova will be affected by the incapacity of the country to correspond to EU community requirements to the process of maintaining population's health.

Through wider implementation and utilization of telemedicine, Republic of Moldova will make an important step to realization of the human right on qualitative health services in the necessary place at the right time.

References

[1] WHO. Regional Office for the Eastern Mediterranean. *What is e-health?*. http://www.emro.who.int/HIS/ehealth/AboutEhealth.htm, 2005.

[2] WOOTTON, Richard. *Telemedicine support for the developing world*, Journal of Telemedicine & Telecare. 14(3):109–14, 2008.

[3] STROETMANN Karl A. et al. *eHealth is Worth it. The economic benefits of implemented eHealth solutions at ten European sites*, Office for Official Publications of the European Communities, Luxemburg, 2006.

[4] "BUSINESS INTELLIGENT SERVICES" SRL. *The aspects of ICT usage by users*, Chisinau, 2007, p. 23.

[5] "BUSINESS INTELLIGENT SERVICES". *Web sites presence and content in Republic of Moldova*, Chisinau, 2006, p. 9.

[6] NORRIS, A.C. *Essentials of Telemedicine and Telecare*, John Whiley and Sons, Ltd, 2002.

[7] COM(2008) 689 *Communication from the Commission to the European Parliament, the Council, the European Economic And Social Committeeand the Committee of the regions on telemedicine for the benefit of patients, healthcare systems and society*, Brussels, 4.11.2008.

List of acronyms

1. **EU** - European Union
2. **ICT** - Information and Communication Technology
3. **IT** - Information Technology
4. **WHO** - World Health Organization

I. Ababii, C. Gaindric, O. Lozan, I. Brinister   Received May 20, 2009

I. Ababii, O. Lozan, I. Brinister,
State University of Medicine and Pharmacy N.Testemitanu
E–mail: *nicolae@mededu.moldline.net*,
olozan@usmf.md, *ybrinister@hotmail.com*

C. Gaindric,
Institute of Mathematics and Computer Science,
Academy of Sciences of Moldova,
str. Academiei 5, 2028, Chisinau, Moldova.
E–mail: *gaindric@math.md*

Computer Science Journal of Moldova, vol.**17**, no.1(49), 2009

About one algorithm of C^2 interpolation using quartic splines

Igor Verlan

Abstract

The problem of C^2 interpolation of a discrete set of data on the interval [a,b] representing the function f using quartic splines is investigated. An explicit scheme of interpolation is obtained using different quartic splines on even and odd subintervals of interpolation.

Mathematical Subject Classification: 41A05, 41A15

Keywords and phrases: interpolation, quartic splines, explicit interpolation.

1 Introduction

Let us suppose that the mesh $\Delta : a = x_0 < x_1 < ... < x_n = b$ is given and $f_i = f(x_i), i = 0(1)n$ are the corresponding data points. The problem of the construction of an interpolation function $S \in C^2[a,b]$ is considered. It is well known (e.g. [1]) that cubic splines may be used in order to solve this problem. In this case you have to solve a tri-diagonal system of linear algebraic equations, which is diagonally dominant one. Well, but in the case of very large set of data you might have problems with capacity of your computer in order to solve this problem. In the case when additional data are available and you have to solve the problem again, it may become critical. In the case of two-dimensional interpolation it is much more difficult to overcome these problems. In what follows quartic splines are considered in order to solve this problem.

2 Algorithm of interpolation using quartic splines

In what follows the next notations are used: $m_i = S'(x_i)$, $M_i = S''(x_i)$, $h_i = x_{i+1} - x_i$, $t = (x - x_i)/h_i$, $\delta_i^{(1)} = (f_{i+1} - f_i)/h_i$.

The following three cases are considered.

a) Let us introduce splines as follows

$$\begin{aligned} S(x) &= f_i + (f_{i+1} - f_i)t + h_i^2 M_i(t^4 - 3t^3 + 3t^2 - t)/6 - \\ &\quad -h_i^2 M_{i+1}(t^4 - 3t^3 + 2t)/6 \end{aligned} \quad (1)$$

For derivatives we have

$$\begin{aligned} S'(x) &= \delta_i^{(1)} + h_i M_i(4t^3 - 9t^2 + 6t - 1)/6 - \\ &\quad -h_i M_{i+1}(4t^3 - 9t^2 + 2)/6 \end{aligned} \quad (2)$$

and

$$S\text{"}(x) = M_i(2t^2 - 3t + 1) - M_{i+1}(2t^2 - 3t) \quad (3)$$

From (1) it follows immediately that interpolation conditions are fulfilled. From (3) it follows that the second derivative is the continuous one at the knots of the mesh.

From (2) for the first derivative at the knots of the mesh we obtain

$$S'(x_i+) = \delta_i^{(1)} - h_i M_i/6 - h_i M_{i+1}/3 \quad (4)$$

and

$$S'(x_i-) = \delta_{i-1}^{(1)} + h_{i-1} M_i/2 \quad (5)$$

From the requirement of continuity of the first derivative at the knots of the mesh the following system of equations is obtained:

$$(3h_{i-1} + h_i)M_i/6 + h_i M_{i+1}/3 = \delta_i^{(2)}, i = 1(1)n - 1,$$

where $\delta_i^{(2)} = \delta_i^{(1)} - \delta_{i-1}^{(1)}$, $i = 1(1)n - 1$.

As it can be seen, the system presented above is the undetermined one. In this case end conditions are required. But, it should be mentioned, for example, that if we have end conditions $M_0 = f"(a)$ and $M_n = f"(b)$, in the sistem given above the value of M_0 is not present.

If the reprezentation of the spline via the first derivatives is used we have

$$\begin{aligned} S(x) &= f_i + (f_{i+1} - f_i)(2t^4 - 6t^3 + 5t^2) + \\ &\quad + h_i m_i(-t^4 + 3t^3 - 3t^2 + t) + h_i m_{i+1}(-t^4 + 3t^3 - 2t^2), \quad (6) \end{aligned}$$

$$\begin{aligned} S'(x) &= \delta_i^{(1)}(8t^3 - 18t^2 + 10t) + m_i(-4t^3 + 9t^2 - 6t + 1) + \\ &\quad + m_{i+1}(-4t^3 + 9t^2 - 4t), \quad (7) \end{aligned}$$

$$\begin{aligned} S"(x) &= \frac{\delta_i^1}{h_i}(24t^2 - 36t + 10) + \frac{m_i}{h_i}(-12t^2 + 18t - 6) + \\ &\quad + \frac{m_{i+1}}{h_i}(-12t^2 + 18t - 4). \quad (8) \end{aligned}$$

From (6) it follows that interpolation conditions are fullfilled and from (7) it follows that the first derivative is the continuous one at the knots of the mesh.

From (8) it follows

$$S"(x_i+) = 10\frac{\delta_i^{(1)}}{h_i} - 6\frac{m_i}{h_i} - 4\frac{m_{i+1}}{h_i}$$

and

$$S"(x_i-) = -2\frac{\delta_{i-1}^{(1)}}{h_{i-1}} + 2\frac{m_i}{h_{i-1}}.$$

From the requirement of continuity of the second derivative the following system of equations is obtained:

$$(h_i + 3h_{i-1})m_i + 2h_{i-1}m_{i+1} = 5h_{i-1}\delta_i^{(1)} + h_i\delta_{i-1}^{(1)}, i = \overline{1, n-1},$$

which is the undetermined one and end conditions are required.

b) Let's consider now the splines in the following form:

$$\begin{aligned} S(x) &= f_i + (f_{i+1} - f_i)t + h_i^2 M_i(-t^4 + t^3 + 3t^2 - 3t)/6 + \\ &+ h_i^2 M_{i+1}(t^4 - t^3)/6. \end{aligned} \tag{9}$$

For derivatives we have

$$S'(x) = \delta_i^{(1)} + h_i M_i(-4t^3 + 3t^2 + 6t - 3)/6 + h_i M_{i+1}(4t^3 - 3t^2)/6$$

$$S"(x) = M_i(-2t^2 + t + 1) + M_{i+1}(2t^2 - t).$$

As in the previous case the interpolation conditions are hold and the second derivative is continuous at the knots of the mesh.

In this case at the knots of the mesh for the first derivative we have

$$S'(x_i+) = \delta_i^{(1)} - h_i M_i/2 \tag{10}$$

and

$$S'(x_i-) = \delta_{i-1}^{(1)} + h_{i-1} M_{i-1}/3 + h_{i-1} M_i/6. \tag{11}$$

From (10) and (11) the corresponding system of linear algebraic equations which ensure the continuity of the first derivative of the spline at the knots of the mesh is obtained:

$$h_{i-1} M_{i-1}/3 + (h_{i-1} + 3h_i) M_i/6 = \delta_i^{(2)}, i = 1(1)n-1.$$

As in the previous case, if the representation via the first derivatives of the spline is used, we have

$$\begin{aligned} S(x) &= f_i + (f_{i+1} - f_i)(-2t^4 + 2t^3 + t^2) + \\ &+ h_i m_i(t^4 - t^3 - t^2 + t) + h_i m_{i+1}(t^4 - t^3), \end{aligned} \tag{12}$$

$$\begin{aligned} S'(x) &= \delta_i^{(1)}(-8t^3 + 6t^2 + 2t) + m_i(4t^3 - 3t^2 - 2t + 1) + \\ &+ m_{i+1}(4t^3 - 3t^2), \end{aligned} \tag{13}$$

$$S"(x) = \frac{\delta_i^{(1)}}{h_i}(24t^2+12t+2)+\frac{m_i}{h_i}(12t^2-6t-2)+ \\ +\frac{m_{i+1}}{h_i}(12t^2-6t). \quad (14)$$

At the knots of the mesh in this case

$$S"(x_i+) = 2\frac{\delta_i^{(1)}}{h_i} - 2\frac{m_i}{h_i},$$

$$S"(x_i-) = -10\frac{\delta_{i-1}^{(1)}}{h_{i-1}} + 4\frac{m_{i-1}}{h_{i-1}} + 6\frac{m_i}{h_{i-1}}.$$

So, the next system of equations results in

$$2h_i m_{i-1} + (3h_i + h_{i-1})m_i = 5h_i\delta_{i-1} + h_{i-1}\delta_i, i = \overline{1, n-1}.$$

c) Let us consider now a scheme of interpolation, when splines (1) and (6) are used alternatively, namely on odd subintervals the splines (1) are used and on even subintervals the splines (6), respectively. As a result, at the odd knots of the mesh from (5) and (7) the following condition of continuity of the first derivative is obtained:

$$M_i = 2\delta_i^{(2)}/(h_{i-1}+h_i) \quad (15)$$

and for even knots of the mesh we get

$$2h_{i-1}M_{i-1} + (h_{i-1}+h_i)M_i + 2h_iM_{i+1} = 6\delta_i^{(2)}. \quad (16)$$

Substituting in (16) expressions which follow from (14) for M_{i-1} and M_{i+1} we get the next formulae for the second derivative at the even knots of the mesh

$$M_i = 6(\delta_i^{(2)} - 2h_{i-1}\delta_{i-1}^{(2)}/(3(h_{i-2}+h_{i-1})) - \\ -2h_i\delta_{i+1}^{(2)}/(3(h_i+h_{i+1})))/(h_{i-1}+h_i). \quad (17)$$

So, we obtain an explicit scheme of interpolation.

If the representation via the first derivatives is used we have

$$S\text{"}(x_i-) = -2\frac{\delta_{i-1}^{(1)}}{h_{i-1}} + 2\frac{m_i}{h_i}, \tag{18}$$

$$S\text{"}(x_i+) = 2\frac{\delta_i^{(1)}}{h_i} - 2\frac{m_i}{h_i}. \tag{19}$$

From requirement of continuity of the second derivative of the spline at the knots of the mesh it follows:

$$m_i = \frac{h_{i-1}\delta_i^{(1)}}{h_{i-1}+h_i} + \frac{h_i\delta_{i-1}^{(1)}}{h_{i-1}+h_i}. \tag{20}$$

Let's consider the knots $i+1$. In this case

$$S\text{"}(x_{i+1}-) = -10\frac{\delta_i^{(1)}}{h_i} + 4\frac{m_i}{h_i} + 6\frac{m_{i+1}}{h_i}, \tag{21}$$

$$S\text{"}(x_{i+1}+) = 10\frac{\delta_{i+1}^{(1)}}{h_{i+1}} - 6\frac{m_{i+1}}{h_{i+1}} - 4\frac{m_{i+1}}{h_{i+1}}. \tag{22}$$

Then we have the following:

$$2h_{i+1}m_i + 3(h_i+h_{i+1})m_{i+1} + 2h_im_{i+2} = 5h_i\delta_{i+1}^{(1)} + 5h_{i+1}\delta_i^{(1)}. \tag{23}$$

Substituting formulae for m_i and m_{i+2} which are obtained from (20) in (23) we get

$$\begin{aligned} m_{i+1} = \frac{1}{3(h_i+h_{i+1})}\Big[&-\frac{2h_ih_{i+1}}{h_{i+1}+h_{i+2}}\delta_{i+2}^{(1)} + (5h_i - \frac{2h_ih_{i+2}}{h_{i+1}+h_{i+2}})\delta_{i+1}^{(1)} + \\ &+(5h_{i+1} - \frac{2h_{i-1}h_{i+1}}{h_{i-1}+h_i})\delta_i^{(1)} - \frac{2h_{i-1}h_{i+1}}{h_{i-1}+h_i}\delta_{i-1}^{(1)}\Big] \end{aligned} \tag{24}$$

and an explicit scheme of interpolation is obtained when representation of spline via the first derivative is used.

3 Remarks on errors of approximation.

We'll consider the case of uniform mesh with step h. Then the formula (15) has the form

$$M_i = (f_{i-1} - 2f_i + f_{i+1})/h^2 \tag{25}$$

and the formula (17), respectively,

$$M_i = (-f_{i-2} + 5f_{i-1} - 8f_i + 5f_{i+1} - f_{i+2})/h^2. \tag{26}$$

If the reprezentation via the first derivatives of the spline at the knots of the mesh is used we have

$$m_i = \frac{f_{i+1} - f_i}{2h}, \tag{27}$$

which follows from (20) and

$$m_{i+1} = \frac{-f_{i+3} + 5f_{i+2} - 5f_i + f_{i-1}}{6h}, \tag{28}$$

which follows from (24).

The set of functions f, which have absolutely continuous derivatives of order $r-1$ on the interval $[a,b]$ and which have derivatives of order r from $L_\infty[a,b]$, is denoted by $W^r_\infty[a,b]$. The norm in this case is defined as follows:

$$\|f(x)\|_\infty = ess\sup|f(x)|\,, x \in [a,b].$$

Lemma 1 *Let us suppose that* $f(x) \in W^3_\infty[a,b]$. *Then the following estimates are valid for regular mesh:*

$$|m_i - f'_i| \le \frac{h^2}{6}\left\|f^{(3)}(x)\right\|_\infty \tag{29}$$

at the odd knots and

$$|m_i - f'_i| \le \frac{13h^2}{18}\left\|f^{(3)}(x)\right\|_\infty \tag{30}$$

for the even knots.

Proof. Let's consider the case (29).

We have
$$|m_i - f_i'| = \left|\frac{f_{i+1} - f_{i-1}}{2h} - f_i'\right|.$$

Substituting f_{i+1} and f_{i-1} by the corresponding Taylor series expansions at the point x_i with the remainder term in the integral form, after necessary transformations we get

$$|m_i - f_i'| = \frac{1}{4h}\left|\int\limits_{x_i}^{x_{i+1}} (x_{i+1} - v)^2 f^{(3)}(v)dv - \int\limits_{x_i}^{x_{i-1}} (x_{i-1} - v)^2 f^{(3)}(v)dv\right|.$$

Using the Hölder inequality in the last relation and computing integrals the presented above estimation follows immediately.

Let's consider now the case (30). We have

$$|m_i - f_i'| = \left|\frac{-f_{i+2} + 5f_{i+1} - 5f_{i-1} + f_{i-2}}{6h} - f_i'\right|.$$

Using the corresponding Taylor series expansions for f_{i-2}, f_{i-1}, f_{i+1}, f_{i+2} with the remainder term in the integral form we get

$$|m_i - f_i'| = \left|\frac{1}{12h}\left(-\int\limits_{x_i}^{x_{i+2}} (x_{i+2} - v)^2 f^{(3)}(v)dv + \right.\right.$$

$$+5\int\limits_{x_i}^{x_{i+1}} (x_{i+1} - v)^2 f^{(3)}(v)dv - 5\int\limits_{x_i}^{x_{i-1}} (x_{i-1} - v)^2 f^{(3)}(v)dv +$$

$$\left.\left. + \int\limits_{x_i}^{x_{i-2}} (x_{i-2} - v)^2 f^{(3)}(v)dv\right)\right|.$$

From the last relation using the Hölder inequality and computing integrals we get
$$|m_i - f_i'| \le \frac{13h^2}{18}\left\|f^{(3)}(x)\right\|_\infty.$$

So, the lemma is proved.

Lemma 2 *Let us suppose that $f(x) \in W^3_\infty[a,b]$. Then the following estimates are valid for regular mesh:*

$$|M_i - f"_i| \leq \frac{h}{3}\left\|f^{(3)}(x)\right\|_\infty$$

at the odd knots and

$$|M_i - f"_i| \leq \frac{13h}{3}\left\|f^{(3)}(x)\right\|_\infty$$

at the even knots.

The proof of the lemma 2 is the analogous one as for lemma 1.
Let us introduce now Hermite splines

$$\begin{aligned} H(x) &= f_i + (f_{i+1} - f_i)(2t^4 - 6t^3 + 5t^2) + \\ &+ h_i f'_i(-t^4 + 3t^3 - 3t^2 + t) + h_i f'_{i+1}(-t^4 + 3t^3 - 2t^2) \end{aligned} \quad (31)$$

and

$$\begin{aligned} H(x) &= f_i + (f_{i+1} - f_i)(-2t^4 + 2t^3 + t^2) + \\ &+ h_i f'_i(t^4 - t^3 - t^2 + t) + h_i f'_{i+1}(t^4 - t^3). \end{aligned} \quad (32)$$

Lemma 3 *Let us suppose that $f(x) \in W^3_\infty[a,b]$. Then for regular mesh:*

$$\left\|H^{(k)}(x) - f^{(k)}(x)\right\|_\infty = O(h^{3-k}), k = 0, 1, 2.$$

Proof. Let's consider the remainder term

R(x)=H(x)-f(x).

For the case (31), substituting Taylor series expansions for f_i, f_{i+1}, f'_i, f'_{i+1} at the point $x = x_i + th$ with remainder term in the integral form after necessary transformations we obtain

$$R(x) = \int\limits_x^{x_i} \Big[(x_i - v)^2 (\frac{1}{2} - t^4 + 3t^3 - \frac{5t^2}{2}) +$$

$$+h(x_i - v)(-t^4 + 3t^3 - 3t^2 + t)\Big] f^{(3)}(v)dv+$$

$$+\int\limits_{x}^{x_{i+1}} \Big[(x_{i+1} - v)^2(t^4 - 3t^3 + \frac{5t^2}{2})+$$

$$+h(x_{i+1} - v)(-t^4 + 3t^3 - 2t^2)\Big] f^{(3)}(v)dv.$$

Substituting in the previous relation $v - x_i = \tau h$ we get

$$R(x) = h^3 \int\limits_0^t \psi_1(t,\tau) f^{(3)}(x_i + \tau h)d\tau + h^3 \int\limits_t^1 \psi_2(t,\tau) f^{(3)}(x_i + \tau h)d\tau,$$

where

$$\psi_1(t,\tau) = \tau \left[-t^4 + 3t^3 - 3t^2 + t - \tau(\frac{1}{2} - t^4 + 3t^3 - \frac{5t^2}{2})\right]$$

and

$$\psi_2(t,\tau) = (1-\tau)\left[(1-\tau)(t^4 - 3t^3 + \frac{5t^2}{2}) - t^4 + 3t^3 - 2t^2\right].$$

From the above it follows that $R(x) = O(h^3)$.

Consider the case (32).

$$R(x) = \frac{h^3}{2} \int\limits_0^t \psi_1(t,\tau) f^{(3)}(x_i + \tau h)d\tau + \frac{h^3}{2} \int\limits_t^1 \psi_2(t,\tau) f^{(3)}(x_i + \tau h)d\tau,$$

where

$$\psi_1(t,\tau) = \tau(t^4 - t^3) - \tau^2(1 + 2t^4 - 2t^3 - t^2)$$

and

$$\psi_2(t,\tau) = (1-\tau)^2(-2t^4 + 2t^3 + t^2) - (1-\tau)(t^4 - t^3),$$

from where it follows that $R(x) = O(h^3)$.

Similarly, for derivatives corresponding estimates are obtained.

Now we are in position to state:

Theorem 1 *If $f(x) \in W^3_\infty[a,b]$ then for regular mesh*

$$\left\| S^{(k)}(x) - f^{(k)}(x) \right\|_\infty = O(h^{3-k}), k = 0, 1, 2.$$

The proof of the theorem follows from the identity

$$R(x) = S(x) - H(x) + H(x) - f(x),$$

and from Lemma 1 and Lemma 3.

4 Conclusions.

So, in the presented paper an explicit scheme of interpolation using quartic splines is obtained. The order of approximation by the proposed algorithm is the same as by the one for cubic splines. The presented algorithm can be extended for bidimensional case.

References

[1] Ahlberg J.H., Nielson E.N., Walsh J.L., *The theory of splines and their applications.* New York: Academic Press 1967.

Igor Verlan

Received April 23, 2009

Institute of Mathematics and Computer Science,
Academy of Sciences of Moldova,
str. Academiei 5, 2028, Chisinau, Moldova.

Department of Applied Mathematics,
Moldova State University,
str. A. Mateevici, 60.
E–mail: *iverlan@rambler.ru*

Computer Science Journal of Moldova, vol.**17**, no.1(49), 2009

Postoptimal analysis of one lexicographic combinatorial problem with non-linear criteria

Vladimir A. Emelichev, Olga V. Karelkina

Abstract

In this article we consider a multicriteria combinatorial problem with ordered MINMIN criteria. We obtain necessary and sufficient conditions of that type of stability to the initial data perturbations for which all lexicographic optima of the original problem are preserved and occurrence of the new ones is allowed.

Mathematics subject classification: 90C27, 90C29, 90C31

Keywords and phrases: multicriteria combinatorial problem, lexicographic set, quasi-stability, binary relations, perturbing matrix

Vector (multicriteria) discrete optimization problems may arise as a result of formalization of object-oriented behavior of a human being in various fields of human activity such as e.g. technical system design, planning and management, business administration, environmental analysis and etc. As far as accuracy of input data is not-guaranteed, frequently, even in the well formalized problems, the reliability of the results (solutions) may be questionable. The data inaccuracy may happen due to various factors, among them the most typical ones are measurement and calculation errors, mathematical model inadequacy and many other. Therefore, it seems to be very natural to define classes of optimization problems for which small perturbations of input data are not significant. This research continues the series of works devoted to the above-mentioned topic [1–5]. We study different aspects of stability to the initial data perturbations for the lexicographic combinatorial problem with MINMIN criteria. In the paper, we formulate and prove necessary and sufficient conditions of quasi-stability of the problem.

This type of stability characterizes the case where all optimal solutions remain optimal under small changes of input data.

Let us consider n-criteria trajectory problem, i.e. problem is given on a system T of non-empty subsets (trajectories) of the set $N_m = \{1, 2, \ldots, m\}$ with sub-criteria of the MINMIN form

$$f_i(t, A) = \min_{j \in t} a_{ij} \to \min_{t \in T}, \ i \in N_n,$$

where $A = [a_{ij}] \in \mathbf{R}^{n \times m}, \ n \geq 1, \ m \geq 2, \ |T| > 1$.

Under n-criterial trajectory problem $Z^n(A)$ we understand the problem of finding the lexicographic set (the set of lexicographic optimal trajectories):

$$L^n(A) = \{t \in T : \ \forall t' \in T \ \ (t \underset{A}{\not\succ} t')\},$$

where $\underset{A}{\not\succ}$ as usual is a negation of the binary lexicographic relation $\underset{A}{\succ}$ defined on the set of trajectories $T \subseteq 2^{N_m}$ by the formula:

$$t \underset{A}{\succ} t' \ \Leftrightarrow \ \exists p \in N_n \ (f_p(t, A) > f_p(t', A) \ \& \ p =$$

$$= \min\{k \in N_n : f_k(t, A) \neq f_k(t', A)\}).$$

It is easy to see that the set $L^n(A)$ is non-empty for any matrix $A \in \mathbf{R}^{n \times m}$ as the subset of the Pareto set.

Note, that many classical combinatorial extreme problems on graphs (traveling salesman problem, spanning tree problem, matching problem, etc.), various problems of scheduling theory and boolean programming problems [6–8] are included into the scheme of the scalar (singlecriterion) problems (with linear, bottleneck, $\sum$-MINMAX, and $\sum$-MINMIN criteria).

By definition, put $\overline{L^n}(A) = T \setminus L^n(A)$.

The following properties are obvious.

Corollary 1. *If* $t \underset{A}{\succ} t'$, *then* $t \in \overline{L^n}(A)$.

Corollary 2. *If* $t \underset{A}{\succ} t'$, *then* $t' \underset{A}{\not\succ} t$.

It is also known (see, e.g., [9]) that the lexicographic set $L^n(A)$ may be defined as a result of solving the sequence of n scalar problems

$$L_i^n(A) = \operatorname{Arg\,min}\{f_i(t, A) : \ t \in L_{i-1}^n(A)\}, \qquad i \in N_n, \tag{1}$$

where $L_0^n(A) = T$, $\operatorname{Arg\,min}\{\cdot\}$ is the set of all optimal trajectories for corresponding minimization problem. Hence, the following inclusions

$$T \supseteq L_1^n(A) \supseteq L_2^n(A) \supseteq \ldots \supseteq L_n^n(A) = L^n(A) \tag{2}$$

are true.

Following [1–5], the problem $Z^n(A)$ is quasi-stable if the formula

$$\exists \varepsilon > 0 \ \ \forall A' \in \Omega(\varepsilon) \ \ (L^n(A) \subseteq L^n(A + A'))$$

is valid. Here

$$\Omega(\varepsilon) = \{A' \in \mathbf{R}^{n \times m} \quad : \quad ||A'|| < \varepsilon\}$$

is a set of perturbing matrices

$$||A'|| = \max\{|a'_{ij}| : \ (i, j) \in N_n \times N_m\}, \ A' = [a'_{ij}].$$

Thus, quasi-stability characterizes the case when all trajectories from lexicographic set preserve a property of optimality for sufficiently small initial data perturbations. Therefore, quasi-stability may be interpreted as the discrete analogue of Hausdorff lower semicontinuity [10] at a point A of the many-valued optimal mapping

$$L^n : \ \mathbf{R}^{n \times m} \rightarrow 2^T.$$

We define binary relations for any non-empty set $I \subseteq N_n$ on the set of trajectories T for the problem $Z^n(A)$

$$t \underset{I,A}{\geq} t' \ \Leftrightarrow \ \forall i \in I \ (f_i(t, A) \geq f_i(t, A)),$$

$$t \underset{I,A}{>} t' \ \Leftrightarrow \ \forall i \in I \ (f_i(t, A) > f_i(t, A)),$$

$$t \underset{I,A}{\vdash} t' \;\Leftrightarrow\; \forall i \in I \; (N_i(t,A) \supseteq N_i(t',A)),$$

where $N_i(t,A) = \text{Argmin}\{a_{ij} : j \in t\}$, i. e. $N_i(t,A) = \{j \in t : a_{ij} = f_i(t,A)\}$.

The following properties are obvious.

Corollary 3. *If* $t \underset{I,A}{\vdash} t'$, *then there exists a number* $\varepsilon > 0$ *such that for any perturbing matrix* $A' \in \Omega(\varepsilon)$ *the relation*

$$t' \underset{I,A+A'}{\geq} t$$

holds.

Corollary 4. *If* $t \underset{N_n,A}{\geq} t'$, *then* $t' \underset{A}{\succeq} t$.

Consequently applying the properties 3, 4 and using continuity of the functions $f_i(t,A), i \in N_n$ on the set of parameters $\mathbf{R}^m$, we deduce the following properties.

Corollary 5. *If* $t \underset{N_n,A}{\vdash} t'$,*then*

$$\exists \varepsilon > 0 \;\; \forall A' \in \Omega(\varepsilon) \; (t \underset{A+A'}{\succeq} t').$$

Corollary 6. *If any of the following conclusions:*

$$(i) \; t \underset{1,A}{>} t',$$

$$(ii) \; \exists k \in N_{n-1} \; (t' \underset{N_k,A}{\vdash} t \;\&\; t \underset{k+1,A}{>} t'),$$

holds for trajectories t *and* t', *then the formula*

$$\exists \varepsilon > 0 \;\; \forall A' \in \Omega(\varepsilon) \; (t \underset{A+A'}{\succ} t')$$

is true.

Denote

$$U^n(A) = \{t \in L^n(A) : \forall i \in N_n \;\; \forall t' \in L_i^n(A) \; (t \underset{i,A}{\vdash} t')\}.$$

Next property follows directly from the previous definition.

Corollary 7. *If* $t \in U^n(A)$ *and* $t' \in L^n(A)$, *then* $t \underset{N_n,A}{\vdash} t'$.

In order to prove the quasi-stability criteria we need a series of lemmas.

Lemma 1. *If* $t \in U^n(A)$ *and* $t' \in T$, *then*

$$\exists \varepsilon > 0 \ \ \forall A' \in \Omega(\varepsilon) \ \ (t \underset{A+A'}{\succeq} t') \tag{3}$$

Proof. Let $t \in U^n(A)$. We consider two possible cases for trajectory t'.

Case 1: $t' \in L_1^n(A)$. Suppose that $t' \in L^n(A)$. Then by virtue of the property 7 the relation

$$t \underset{N_n,A}{\vdash} t'$$

holds. Hence, taking into account the property 5, we get (3).

Now let $t' \in L_1^n(A) \setminus L^n(A)$. Thus, there exists an index $k = k(t') \in N_n \setminus \{1\}$, such that $t' \notin L_k^n(A)$ and $t' \in L_i^n(A)$ for $i \in N_{k-1}$. Therefore, we obtain

$$t \underset{N_{k-1},A}{\vdash} t' \ \text{ and } \ t' \underset{k,A}{>} t.$$

Making use of this facts and property 6, we conclude that the formula

$$\exists \varepsilon > 0 \ \ \forall A' \in \Omega(\varepsilon) \ \ (t' \underset{A+A'}{\succ} t)$$

is true. Therefore, due to the property 2, we obtain (3).

Case 2: $t' \in T \setminus L_1^n(A)$. Thus,

$$t' \underset{1,A}{>} t.$$

Therefore, in view of the properties 2 and 6 the formula (3) is true.

Lemma 1 is thus proved.

Lemma 2. *If* $t \in L^n(A) \setminus U^n(A)$, *then the formula*

$$\exists t^0 \in T \ \ \forall \varepsilon > 0 \ \ \exists A^0 \in \Omega(\varepsilon) \ (t \underset{A+A^0}{\succ} t^0) \tag{4}$$

is true.

Proof. Since $t \notin U^n(A)$, then there exist $k \in N_n$ and $t^0 \in L_k^n(A)$ such that $N_k(t, A) \not\supseteq N_k(t^0, A)$ and $t \in L_k^n(A)$ (by virtue of $t \in L^n(A)$). Hence $f_k(t, A) = f_k(t^0, A) = a_{kp}$, if $p \in N_k(t^0, A) \setminus N_k(t, A)$. Therefore, let us assume $\varepsilon > 0$ and construct elements of a perturbing matrix $A^0 = [a_{ij}^0] \in \mathbf{R}^{n \times m}$ according to the rule

$$a_{ij}^0 = \begin{cases} -\alpha, & \text{if } i = k,\ j = p, \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < \alpha < \varepsilon$, in view of $p \in N_k(t^0, A) \setminus N_k(t, A)$ we conclude that the relations

$$f_k(t^0, A + A^0) = \min\{a_{kj} + a_{kj}^0 : j \in t^0\} = a_{kp} - \alpha < a_{kp} =$$

$$= f_k(t, A) = f_k(t, A + A^0),$$

$$f_i(t^0, A + A_0) = f_i(t^0, A) = f_i(t, A) = f_i(t, A + A_0),\ i \in N_{k-1}$$

hold true. Hence,

$$t \underset{A+A^0}{\succ} t^0,$$

i.e. formula (4) is true.

Lemma 2 is thus proved.

Now let us formulate quasi-stability criterion for the concerned problem.

Theorem. *The vector problem $Z^n(A)$, $n \geq 1$, is quasi-stable if and only if the formula*

$$\forall t \in L^n(A) \quad \forall i \in N_n \quad \forall t' \in L_i^n(A) \quad (t \underset{i,A}{\vdash} t') \tag{5}$$

is true.

Proof. Sufficiency. Let the formula (5) holds true and $t \in L^n(A)$. Then $t \in U^n(A)$ and, therefore, due to Lemma 1 we find that

$$\forall t' \in T \quad \exists \varepsilon(t') > 0 \quad \forall A' \in \Omega(\varepsilon(t'))\ (t \underset{A+A'}{\not\succ} t').$$

Hence, by putting $\varepsilon(t) = \min\{\varepsilon(t') : t' \in T\}$, it is easy to see that for any trajectory $t \in L^n(A)$ and for any perturbing matrix $A' \in \Omega(\varepsilon(t))$

the inclusion $t \in L^n(A + A')$ is true. Therefore, if $\varepsilon^* = \min\{\varepsilon(t) : t \in L^n(A)\}$, we obtain

$$\exists \varepsilon^* > 0 \ \ \forall A' \in \Omega(\varepsilon^*) \ \ (L^n(A) \subseteq L^n(A + A')).$$

Thus, the problem $Z^n(A)$ is quasi-stable.

Necessity. We assume that, on the contrary, the problem $Z^n(A)$ is quasi-stable, but the formula (5) is not true. Then there exists trajectory $t \in L^n(A) \setminus U^n(A)$, for which on account of Lemma 2 and property 1 the formula

$$\forall \varepsilon > 0 \ \ \exists A^0 \in \Omega(\varepsilon) \ (t \in \overline{L^n}(A + A^0))$$

is true. Hence, we conclude

$$\forall \varepsilon > 0 \ \ \exists A^0 \in \Omega(\varepsilon) \ (L^n(A) \not\subseteq L^n(A + A^0)),$$

This is contradiction to the quasi-stability of the problem $Z^n(A)$.

Theorem is proved.

Let us give two examples which illustrate stated result.

Example 1. Let $n = 2$, $m = 4$, $T = \{t^1, t^2, t^3\}$, $t^1 = \{1, 2, 4\}$, $t^2 = \{1, 4\}$, $t^3 = \{1, 2\}$,

$$A = \begin{pmatrix} 1 & 3 & 1 & 2 \\ 3 & 2 & 2 & 1 \end{pmatrix}.$$

Thus,

$$f_1(t^1, A) = f_1(t^2, A) = f_1(t^3, A) = 1.$$

Therefore $L_1^2(A) = \{t^1, t^2, t^3\} = T$. Moreover, we have

$$f_2(t^1, A) = f_2(t^2, A) = 1, \quad f_2(t^3, A) = 2.$$

Hence, we get lexicographic set $L^2(A) = L_2^2(A) = \{t^1, t^2\}$.

Further we find the sets

$$N_1(t^1, A) = N_1(t^2, A) = N_1(t^3, A) = \{1\},$$

$$N_2(t^1, A) = N_2(t^2, A) = \{4\}.$$

Therefore, the formula

$$\forall t \in L^2(A) \;\; \forall i \in N_2 \;\; \forall t' \in L_i^2(A) \;\; (N_i(t, A) = N_i(t', A))$$

is true. Consequently, in virtue of the theorem the problem $Z^2(A)$ is quasi-stable.

Example 2. Let $n = 2$, $m = 4$, $T = \{t^1, t^2, t^3\}$, $t^1 = \{1, 2, 4\}$, $t^2 = \{1, 2\}$, $t^3 = \{1, 3, 4\}$,

$$A = \begin{pmatrix} 1 & 3 & 1 & 2 \\ 3 & 2 & 2 & 1 \end{pmatrix}.$$

Then

$$f_1(t^1, A) = f_1(t^2, A) = f_1(t^3, A) = 1.$$

Thus, $L_1^2(A) = \{t^1, t^2, t^3\} = T$. Then, we have

$$f_2(t^1, A) = f_2(t^3, A) = 1, \quad f_2(t^2, A) = 2.$$

Hence we get the lexicographic set $L^2(A) = L_2^2(A) = \{t^1, t^3\}$.

Having found the sets

$$\{1\} = N_1(t^1, A) = N_1(t^2, A) \not\supseteq N_1(t^3, A) = \{1, 3\},$$

$$N_2(t^1, A) = N_2(t^3, A) = \{4\},$$

we conclude that conditions of the theorem don't hold. Therefore, the problem $Z^2(A)$ isn't quasi-stable.

Corollary 1. *A sufficient condition for the problem $Z^n(A)$ to be quasi-stable is equality $|L_1^n(A)| = 1$.*

Let us give an example illustrating that the equality $|L_1^n(A)| = 1$ isn't necessary condition for the problem to be quasi-stable.

Example 3. Let $n = 2$, $m = 3$, $T = \{t^1, t^2\}$, $t^1 = \{1, 2, 3\}$, $t^2 = \{1, 2\}$,

$$A = \begin{pmatrix} 3 & 2 & 4 \\ 3 & 5 & 5 \end{pmatrix}.$$

Then

$$f_1(t^1, A) = f_1(t^2, A) = 2.$$

Therefore $L_1^2(A) = \{t^1, t^2\} = T$. Moreover

$$f_2(t^1, A) = f_2(t^2, A) = 3.$$

Hence, the lexicographic set is $L^2(A) = L_2^2(A) = \{t^1, t^2\}$. Thus, $|L_1^2(A)| = 2$.

Further, having found the sets

$$N_1(t^1, A) = N_1(t^2, A) = \{2\},$$

$$N_2(t^1, A) = N_2(t^2, A) = \{1\},$$

we conclude that the formula

$$\forall t \in L^2(A) \;\; \forall i \in N_2 \;\; \forall t' \in L_i^2(A) \;\; (N_i(t, A) = N_i(t', A))$$

is true. Therefore, by theorem, the problem $Z^2(A)$ is quasi-stable but $|L_1^2(A)| > 1$.

Corollary 2. *The formula*

$$\forall t, t' \in L^n(A) \;\; (N_1(t, A) = N_1(t', A)) \tag{6}$$

is necessary condition for the problem $Z^n(A), n \geq 1$ to be quasi-stable.

It is obvious that the formula (6) is simultaneously a sufficient condition for quasi-stability of the problem $Z^1(C)$ in scalar case ($n = 1$).

References

[1] Emelichev V.A., Berdysheva R.A. *On the radii of steadiness, quasi-steadiness, and stability of a vector trajectory problem on lexicographic optimization.* Discrete Math. Appl., 1998, **8**, N 2, p. 135–142.

[2] Berdysheva R.A., Emelichev V.A. *Some forms of stability of a combinatorial problem of lexicographic optimization.* Izv. Vyssh. Uchebn. Zaved. Mat., 1998, N 12, p. 11–21 (in Russian).

[3] Emelichev V.A., Girlich E., Nikulin Yu.V., Podkopaev D.P. *Stability and regularization of vector problem of integer linear programming.* Optimization, 2002, **51**, N 4, p. 645–676.

[4] Emelichev V.A., Kuzmin K.G., Leonovich A. M. *Stability in the combinatorial vector optimization problems* Automatic and Remote Control, 2004, **65**, N 2, p. 227–240.

[5] Emelichev V.A., Kuzmin K.G. *Stability criteria in vector combinatorial bottleneck problems in terms of binary relations.* Cybernetics and Systems Analysis, 2008, **44**, N 3, p. 397–404.

[6] Sergienko I.V., Kozeratskaya L. N., Lebedeva T. T. *Investigation of stability and parametric analysis of discrete optimization problems.* Kiev, Navukova Dumka, 1995 (in Russian).

[7] Emelichev V.A., Kravtsov M.K. *On combinatorial vector optimization problems.* Discrete Math. Appl., 1995, **5**, N 2, p. 93–106.

[8] Sotskov Yu.N., Leontev V.K., Gordeev E.N. *Some concepts of stability analysis in combinatorial optimization.* Discrete Appl. Math., 1995, **58**, N 2, p. 169–190.

[9] Ehrgott M. *Multicriteria Optimization (2nd edn).* Berlin-Heidelberg, Springer, 2005.

[10] Tanino T. *Sensitivity analysis in multiobjective optimization. Journal of Optimization Theory and Applications.* 1988, **56**, N 3, p. 479–499.

Vladimir A. Emelichev, Olga V. Karelkina,
Received October 15, 2008

Belarussian State University
ave. Independence, 4,
Minsk, 220030, Belarus
E–mail: *emelichev@bsu.by*, *olga.karelkina@gmail.com*

Computer Science Journal of Moldova, vol.**17**, no.1(49), 2009

Structured knowledge management techniques for the development of interactive and adaptive decision support system

Iulian Secrieru

Abstract

The phase of knowledge acquisition and formalization is being considered as the key one for the development of decision support system (DSS). The main problem at this stage is to find a knowledge representation (KR) and a supporting reasoning system that can make the inferences your application needs. The main criterion of choice is what kind of inference the developers prefer and is more appropriate to the problem domain. However, already at the stage of the user interface creation there arise tension between KR and end-user "needs" (preferences and habits). So, multiple representations of the acquired and structured knowledge and management techniques could provide a solution for decision support system inference and interface requirements satisfaction.

Keywords: Knowledge representation, knowledge management techniques, decision support system, interactive and adaptive interface.

1 Introduction

The problem, associated with the physicians (doctors) diagnostic activities, acquire a special relevance in modern circumstances. First of all it is connected with the fact that the doctors have to work with weakly structured and formalized information. Besides, the volume of information is in a continuous growth thanks to the appearance of new methods of examination of patients.

The ultrasound investigation domain is not an exception. The appearance of new ultrasound devices or the improvement of the old scanners doesn't simplify but complicates the physician's diagnostic thinking, because he has to analyze a much larger number of diagnostic data, which typically reduces the accuracy and increases the time of determining the diagnosis.

The phase of knowledge acquisition and formalization is being considered as the key one for the development of decision support system (DSS). In order to obtain a well-structured description of the problem domain, the developers are forced to choose a "rigid" scheme of its representation. The problem domain is quite often represented as a decision tree or semantic net.

The development of the user's interface for the DSS based on the decision tree can lead to various problems and inconveniences.

The main lack is the fact that the interface does not correspond to the daily work and habits of the end-user.

The discrepancy of the user's interface of the decision support medical system with the form of the doctor's diagnostic thinking may become the reason of different mistakes or it may lead to the rejection of the user to utilize it in medical practice.

The principles and techniques of structured knowledge manipulation and management aiming to create an interactive and adaptive user interface for a decision support system in the ultrasound investigation domain, is being described in this article.

2 Knowledge Representation Schemes

The fundamental goal of KR is to represent knowledge in a manner as to facilitate drawing conclusions (inferencing) by decision support or another computer-aided systems.

We distinguish two approaches - single and hybrid KR schemes.

First we focus on the most popular single KR schemes.

Semantic nets, decision trees and their descendants (*frames* or *schemes*) [1] represent knowledge in the form of graph (or hierarchy). Nodes in graph represent concepts and the edges represent relations between the concepts. Nodes in a frame hierarchy also represent concepts, but they have internal structure that describes the corresponding concept via a set of attributes. All of these KR schemes are very natural and well suited for representing structural and relational knowledge. They can also make efficient inferences for small to medium graphs (hierarchies). However, it is difficult to represent heuristic knowledge, uncertain knowledge, and make inferences from partial inputs. Also, explanations are not provided and knowledge updates are difficult. *Conceptual graphs* are similar to semantic nets, whereas ontologies [2] refer to a representation scheme similar to frames, but more restrictive.

Symbolic rules are one of the most popular KR methods [1]. They represent general domain knowledge in the form of IF-THEN rules: if <conditions> then <conclusion>, where the term <conditions> represents the conditions of a rule, whereas the term <conclusion> represents its conclusion. The conditions are connected with one or more logical operators such as "and", "or", and "not". The inference engine uses the knowledge in the rule base as well as facts about the problem at hand to draw conclusions. Typically, facts are provided by the user during inference. There are two main inference approaches: backward chaining (guided by the conclusions) and forward chaining (guided by the input data). The explanation module provides explanations regarding the drawn conclusions. Rules are natural (easy to comprehend) and rule-base updates (removing/inserting rules) can be easily made. In addition, heuristic knowledge is naturally represented by rules. Efficiency of the inference process depends on the length of the inference chains. Additionally, conclusions cannot be derived if some of the inputs are unknown. Finally, pure rules cannot represent uncertain or vague knowledge and are not suitable for representing structural and relational knowledge.

Fuzzy rules (fuzzy logic) are good at representing imprecise and fuzzy terms, like "low" and "high". Fuzzy logic extends traditional logic and sets membership by defining membership functions over the range [0.0,1.0], where 0.0 denotes absolute falseness and 1.0 - absolute truth [3]. Given the above, fuzzy rules are good for representing vagueness. However, fuzzy rules are not as natural as symbolic rules, that complicates the knowledge acquisition and the updates processes. Inference is more complicated and less natural than in simple rule-based reasoning. Provision of explanations is feasible, but not all reasoning steps can be explained.

Case-based representations [4] store a large set of past cases with their solutions in the case base and use them whenever a similar new case has to be dealt with. A case-based system performs inference in four stages: (1) retrieve, (2) reuse, (3) revise, and (4) retain. In the retrieval stage, the stored case(s) most relevant to the new case is (are) retrieved. Similarity measures and indexing schemes are used in this context. In the reuse stage, the retrieved case is combined with the new case to create a solution. The revise stage validates the correctness of the proposed solution. Finally, the retain stage decides on retention (or not) of the new case. Cases are usually easy to obtain. Cases are natural. Explanations cannot be provided in a straightforward way as in rule-based systems. Even if some of the inputs are not known, conclusions can be reached through similarity to stored cases. Updates can be easily made. However, the efficiency of the inference process depends on the size of the case base. Finally, cases are not suitable for representing structural, uncertain, and heuristic knowledge.

Neural networks represent a totally different approach to artificial intelligence, known as connectionist [5]. A neural network consists of many simple interconnected processing units called neurons. Neural networks are very efficient in producing conclusions, since inference is based on numerical calculations, and can reach conclusions based on partially known inputs due to their generalization ability. On the other hand, neural networks lack naturalness of representation, that is, the encompassed knowledge is incomprehensible, and explanations for the reached conclusions cannot be provided. It is also difficult to make structural updates to specific parts of the network. Neural networks do not possess inherent mechanisms for representing structural, relational, and uncertain knowledge. Heuristic knowledge can be represented to some degree via supervised training.

Belief networks (or *probabilistic nets*) [6] are graphs, where nodes represent statistical concepts and links represent mainly causal relations between them. Each link is assigned a probability which represents how certain is it that the concept, where the link departs from, causes the concept, where the link ultimately arrives. Belief nets are good for representing causal relations between concepts. Also, they can represent heuristic knowledge. Furthermore, they can represent uncertain knowledge through the probabilities and make relatively efficient inferences (via computations of probabilities propagation). However, estimation of probabilities is difficult, making the knowledge acquisition process a problem. For the same reason, it is difficult to make updates. Also, explanations are difficult to produce, since the inference steps cannot be easily followed by humans. Furthermore, their naturalness is reduced.

Hybrid schemes are integrations of two or more single KR schemes. We mention the most popular ones.

Connectionist rule-based representations [5] combine neural networks with rule-based representation. The knowledge base is a network whose nodes correspond to domain concepts. Dependency information regarding the concepts is used to create links among nodes. The network's weights are calculated through a training process using a set of training patterns. In addition to the knowledge base, connectionist rule-based systems also consist of an inference engine and an explanation mechanism. Compared to neural networks, they offer more natural representation and can provide some type of explanation. Naturalness is enhanced due to the fact that most of the nodes correspond to domain concepts.

Another approach in hybrid knowledge representation is the *integrations of rule-based reasoning with case-based reasoning* [7]. Compared to "pure" case-based reasoning, their key advantage is the improvement in the performance of the inference engine and the ability to represent heuristic and relational knowledge. Furthermore, the synergism of rules and cases can cover up deficiencies of rules (improved knowledge acquisition) and also enable partial input inferences. The existence of rules in such hybrid schemes makes updates more difficult than "pure" case-based representations. Also, explanations can be provided but not as easily as in pure rule-based representations, given that similarity functions are still present.

There are various ways to integrate neural networks and fuzzy logic [8]. Such integrations are the fuzzy neural networks and the hybrid neuro-fuzzy representations. Fuzzy neural networks retain the basic properties and architectures of neural networks and "fuzzify" some of their elements. In a hybrid neuro-fuzzy system, both fuzzy techniques and neural networks play key role. Each does its own job in serving different functions in the system. Hybrid neuro-fuzzy systems seem to satisfy KR requirements to a greater degree than fuzzy neural networks. This hybrid approach enables the representation of incomplete, imprecise, and vague information and also exploits the generalization capability of neural networks.

Neurules are type of hybrid rules integrating symbolic rules with neurocomputing [9, 10]. In contrast to other hybrid approaches, the constructed knowledge base retains the modularity of rules, since it consists of autonomous units (neurules), and also retains their naturalness in a great degree, since neurules look much like symbolic rules. Neurules can be constructed either from symbolic rules [9], thus exploiting existing symbolic rule bases, or empirical data [10].

A conclusion that can be drawn is that there is no single or hybrid schemes that satisfy end-users preferences and/or all the requirements of decision support systems developers. So, taking into account only the system requirements on the knowledge acquisition and modeling stages, one can say that semantic nets, decision trees, frames, description logics are more suitable for representing knowledge in the domain model.

3 Knowledge base of the decision support system SONARES.

SONARES is a knowledge based system in the ultrasound investigation domain. Experts are its main source of knowledge. Expert knowledge was obtained in result of "knowledge engineer – expert group" communication and stored as a pyramid of meta-concepts. The common work of knowledge engineer with the experts revealed that in the ultrasound investigation domain the reasoning based on meta-concepts (facts) and knowledge representation in the form of a hierarchy (pyramid) totally corresponds to the expert's thinking and reasoning.

The semantic rules scheme was chosen as a model of acquired knowledge representation. Based on the principles of semantic rules scheme the knowledge base of the decision support system SONARES has been established. It consists of a pyramid of meta-concepts, and of a set of rules created on its basis.

As a result of 23 common working sessions of knowledge engineer with experts there was received a pyramid of knowledge (decision tree) which consists of 335 facts (9 root nodes with a maximum deep level equal to 9) and 54 rules [11, 12 , 13]. This knowledge represents formalized description of the ultrasound investigation process of gallbladder.

4 The user interface based on the decision tree scheme

Let's analyze the following example.

Suppose that during the stage of knowledge acquisition the developers have identified a group of 3 mutually exclusive facts: F1 = <gallbladder volume, normal>, F2 = <gallbladder volume, enlarged>, F3 = <gallbladder volume, reduced>. That is, there was identified the attribute A1=<GALLBLADDER VOLUME> with three possible values: V1="normal", V2="enlarged", V3="reduced". Due to the sources of knowledge, these three facts can be represented as a decision tree in three different ways (see Figure 1).

If we take into consideration all possible options for the interchangeability of the facts for each of these ways of representation, we'll get 18 different decision trees to describe these 3 facts.

Thus, the interface creation on the base of one of the 18 decision trees can be unusual for certain users, whose process of reasoning is described by one of the remaining 17 options. In this case, the created interface won't correspond to their daily work.

Besides, the user interface based on the decision tree scheme is

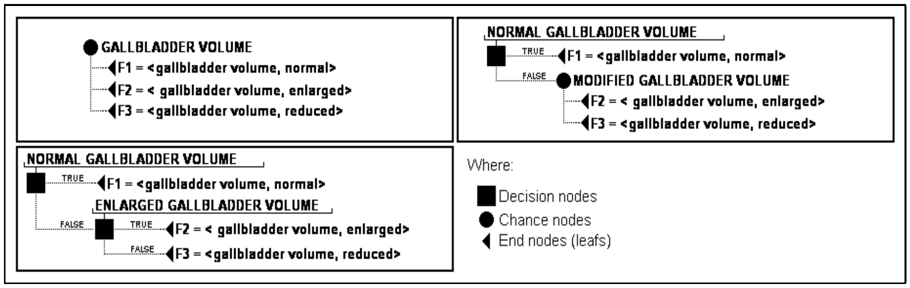

Figure 1. Decision trees for description of 3 exclusive facts.

inefficient because:

- unjustifiedly limits the end-user actions
- requires frequent appeals to the knowledge base
- requires a lot of screen space

It is obvious, in order to organize an effective dialog with the DSS users, it is necessary to develop an alternative representation scheme for the knowledge base represented in form of decision tree.

5 Alternative representation of the knowledge base as a means of effective user interface for the DSS.

Note that we are not talking about replacing the decision tree as a means of knowledge representation scheme at the stage of knowledge acquisition. Since, at this stage, the use of such representations is reasonable, and in some cases it is the necessary and the only right decision. Especially in cases when developers want to receive a well-structured knowledge base and have to deal with poorly formalized domain. In these cases it is meaningful and effective to use the decision tree scheme even as a means of visualization the knowledge base. [14, 15].

The aim of creation of the alternative representation scheme of the knowledge base is the organization on it's bases the effective dialog with the DSS users and the elimination of the deficiencies, inherent in the user interface based on decision tree scheme.

The source of information for alternative representation scheme is the DSS knowledge base, described as a decision tree.

In our case, the DSS SONARES knowledge base is represented in the form of a decision tree which consists of 335 facts that can describe any situation in the domain of gallbladder ultrasound investigation. There were formulated 54 rules on its base, which correspond to the anomalies and pathologies of this problem domain. It is necessary to propose such representation of this acquired knowledge in order to have the opportunities to realize the user interface of DSS SONARES with the following features:

<u>The interface must be simple and understandable.</u> The dialog with the end-users should take place in its usual rhythm and form, and should not be necessary required the change of his reasoning.

<u>The interface should correspond to the user's daily work and preferences.</u> The users must have the possibility to influence the dialog form.

<u>The interface must be "transparent".</u> The solution proposed by DSS should be easily verified.

<u>The user dialog should not have a linear structure.</u> The user should always have the opportunity to return to the appointed step back.

<u>The interface should be interactive.</u> It must change in dependence on the user available time to make a decision. In addition, the interface must conform to the basic forms of user's diagnostic thinking.

<u>The interface should not unnecessarily restrict the end-user's actions.</u>

<u>The interface should be oriented on a restricted screen space and a limited decision making time.</u> The medical DSS users often have to use them in an emergency or a network mode.

The most common form of communication and information transfer is the dialog. Therefore, the organization of the DSS user interface as an ordered set of questions is justified.

The facts of which the decision tree consist, in fact constitute a meta-knowledge, through which we can describe some situation in the selected problem domain and are involved in the inference to determine solutions. That is, these are the answers to those questions that we should address to the user in order to help him with the decision making. It is obvious that in the interface based on a decision tree, the questions themselves are missing (determined by the structure of the decision tree). It doesn't correspond to the usual way of the users reasoning, because for each of the fact from the decision tree representation the user is forced to formulate a question. A more common variant for him could be the option to answer to the specific questions by selecting from a list of all possible answers.

The essence of the proposed new representation approach is the separation of knowledges to those ones that can be used in the inference and those which are used only in the interface.

At the first stage of the creation of the alternative representation of the knowledge base there were determined those facts of the decision tree, which are involved in the inference. For each of them there has been formulated a question concerning the existence or not of this fact. For example, for the fact F1 = <gallbladder volume, normal> there was formulated the question Q1 = "Is the volume of gallbladder a normal one?", for the fact F2 = <gallbladder volume, enlarged> - the question Q2 = "Is the volume of gall bladder enlarged?", and for F3 = <gallbladder volume, reduced> - Q3 = "Is the volume of gallbladder reduced?".

As a result, 203 questions were formulated.

Answering to some of these questions, the user can describe the case of gallbladder ultrasound investigation domain, in which he needs assistance of DSS SONARES.

In terms of these questions all of 54 pathologies and anomalies of this domain are described. That is, each pathology or anomaly from the gallbladder ultrasound investigation domain can by described by the vector (Q1.value,Q2.value,...Qn.value), where Qi.value — is the answer to the question Qi, n - total number of questions, in our case n=203. Under these conditions, the whole diagnostic knowledge base (the information about all pathologies and anomalies, that is necessary for the inference to make a decision) can be represented in the form of decision making matrix [Pi,Qj.value], in our case i=1..54, j=1..203.

Concerning the matrix representation of diagnostic knowledge base, the proposed approach was named *alternative matrix representation* of KB.

On the second stage, we saved all existing relationships between facts. That is, we elaborated an interconnection system between all formulated questions.

There are two types of relationships between facts in the decision tree.

The first one indicates the position of given fact in the knowledge base hierarchy.

Let's analyze the subtree F4-F5-F6, where F4=<gallbladder form, abnormal (abnormality of conformation)>, F5=<gallbladder twist, present>, F6=<gallbladder twist form, circular>. There is a hierarchical relationship between the facts F4-F5, which indicates that it makes sense to show the fact F5 only in the case when the fact F4 is determined. The same relationship exists between the facts F4-F6 and F5-F6. They are not taken into account during the inference process, however, they are of great importance for the determination of the opportunity of a fact visualization. In our case, this information helps us to determine the opportunity of a question visualization and organize the dialog with the DSS user.

The second type of relationships indicates the existence of interdependence between facts.

For example, the above mentioned facts F1, F2, F3 are mutually exclusive (it means If F1 = TRUE then (F2 = FALSE) & (F3 = FALSE), If F2 = TRUE then (F1 = FALSE) & (F3 = FALSE), If F3 = TRUE then (F1 = FALSE) & (F2 = FALSE)). These relations does not depend on the form of visualization of the facts or the whole user interface, but form the basis of the system knowledge base and inference.

The separation of the existing relationships between questions in two groups, those that can be used in inference and those which are used only in the interface, allows us to create a high-quality interactive interface based on individual characteristics and habits of the end-user. This is achieved because the user can define himself the subject and the form of dialog (by changing the visualization relationships between questions), without any fear to influence the inference.

Some of the questions can be grouped. For example, the questions Q1 = “Is the volume of gall bladder a normal one?”, Q2 = “Is the volume of gall bladder enlarged?”, and Q3 = “Is the volume of gall bladder reduced?” could be grouped into the group, which describes the volume of the gallbladder. Now, if the user wants to visualize all the questions related to the volume of the gallbladder, he may do so through the visualization of the group.

Additionally, the questions association into the group will allow to diversify the form of dialog.

The resulting relational database is the alternative representation of DSS SONARES knowledge base.

6 Conclusions

Realization of the described approach has shown that the creation of alternative matrix representation of DSS knowledge base requires additional time for its creation, but it is justified, if we want to be able to organize an effective interactive dialog with the user. In addition, the user interface based on a matrix representation is simple, understandable and transparent, fully corresponds to the daily activities and habits of the user, not unreasonably restrict the user actions.

This approach allows realization of different versions of the interface with the restricted screen space and limited time of the decision making (for systems used in emergency cases).

Table 1 compares the KR schemes discussed in the previous sections with the proposed approach. Symbol ”-” means ”unsatisfactory”; ”±” - ”average”; ”+” - ”good”; and ”V” - ”very good”.

Table 1. Comparison of knowledge representation schemes and approaches (data source [16])

	USER'S REQUIREMENTS									SYSTEM REQUIREMENTS					
	Naturalness	Ease knowledge update	Efficient inference	Explanations	Ease knowledge acquisition	Partial input inferences	Adaptive inference	Interactive interface	End-user's preferences update	Structural knowledge	Relational knowledge	Uncertain knowledge	Vague knowledge	Heuristic knowledge	New knowledge discovery
Semantic nets / decision trees / frames	V	±	V	-	+	-	±	±	±	V	V	-	-	-	±
Symbolic rules	V	V	+	V	±	-	+	+	±	-	±	-	-	V	±
Case-based representations	V	V	+	+	V	+	±	±	-	-	+	-	-	-	±
Belief networks	±	-	V	-	±	-	-	±	-	+	V	V	±	±	±
Neural networks	-	-	V	-	V	V	+	+	±	-	±	-	-	±	+
Fuzzy rules	+	-	+	-	±	-	±	±	-	-	±	±	V	V	±
Connectionist expert system	±	±	V	±	V	V	±	±	±	-	±	-	-	±	±
Neuro-fuzzy representations	±	-	+	-	+	±	±	±	±	-	±	±	V	+	±
Cases and rules	V	+	+	+	+	+	+	+	±	-	+	-	-	+	±
Neurules	+	+	V	V	V	V	+	+	+	-	±	-	-	V	±
Semantic nets/decision trees & symbolic rules in form of "matrix" representation	V	+	V	V	+	±	+	+	+	V	V	-	-	V	+

Also there have been identified some additional advantages of using a matrix representation of the DSS knowledge base.

1. Matrix representation of knowledge base allows to organize an interactive interface according to the type of user's diagnostic thinking. We realized a version of user interface with adaptive support for inductive reasoning ability.

2. Matrix representation of knowledge base has a cognitive value. It can be used as a means of visualization and detection of weakly described sub-domain in the problem domain in general and in the knowledge base in particular.

3. In matrix representation of the knowledge base every decision is described by the vector (Q1.value, Q2.value, ... Qn.value). By calculating the correlation coefficient between the vectors, the solutions can be grouped by various criteria. This will allow better knowledge formalization of problem domain.

Acknowledgments
This research is supported partially by Moldova State Program and the Science & Technology Center in Ukraine (STCU) grant #4035.

References

[1] Negnevitsky, M. *Artificial intelligence: A guide to intelligent systems.* Reading, MA: Addison Wesley, 2002.

[2] Staab, S., Studer, R. (Eds.). *Handbook on Ontologies.* Berlin: Springer-Verlag, 2004.

[3] Dubois, D., Pride, H., Yager, R. (Eds.). *Fuzzy rules in knowledge-based systems.* In Reading in fuzzy sets for intelligent systems. San Francisco: Morgan Kaufmann, 1993.

[4] Leake, D.B. (Ed.) *Case-based reasoning: Experiences, lessons & future directions.* Menlo Park, CA: AAAI Press/MIT Press, 1996.

[5] Gallant, S.I. (1993). *Neural network learning and expert systems.* Cambridge, MA: MIT Press, 1993.

[6] Russel, S., Norvig, P., *Artificial intelligence. A modern approach.* Englewood Cliffs, NJ: Prentice Hall, 2003.

[7] Golding, A.R,. Rosenbloom, P .S., *Improving accuracy by combining rule-based and case-based reasoning.* Artificial intelligence, 87, pp. 215–254, 1996.

[8] Nauck, D., Klawonn, F., Kruse, R. *Foundations of neuro-fuzzy systems.* New York: John Wiley & Sons, 1997.

[9] Hatzilygeroudis, I., Prentzas, J., *Neurules: Improving the performance of symbolic rules.* International Journal on Artificial Intelligence Tools, 9, pp. 113–130, 2000.

[10] Hatzilygeroudis, I., Prentzas, J., *Constructing modular hybrid rule bases for expert systems.* International Journal on Artificial Intelligence Tools, 10, pp. 87–105, 2001.

[11] Popcova, O., Secrieru, Iu., Sologub, D., Jantuan, E., Papanaga, V., *Decision Support System for Ultrasound Diagnostics,* Proceedings of the 1st International Conference of Young Scientists Computer Science & Engineering 2006, Lviv, Ukraine, pp. 30–31, 2006.

[12] Burtseva, L., Cojocaru, S., Gaindric, C., Jantuan, E., Popcova, O., Secrieru, I., Sologub, D., *SONARES – A decision support system in ultrasound investigations,* Computer Science Journal of Moldova, nr. 2 (44), vol. 15, Kishinev, pp. 153–177, 2007.

[13] Secrieru, Iu., Popcova, O., Puiu, S., Sologub, D.,. *Knowledge Structure Modeling in Ultrasound Investigation Domain.* International Conference BIT+ "Informational Technologies-2007". VIIth edition, Chişinău, Moldova, pp. 59–60, 2007.

[14] Secrieru, Iu., Popcova, O., Puiu, S., Sologub D., *Knowledge Structure Modeling in Ultrasound Investigation Domain.* Proceedings

of the 5th European Conference on Intelligent Systems and Technologies (ECIT 2008), Iasi, Romania, CD, Performantica. pp. 1–7, 2008.

[15] Popcova, O., Puiu, S., Secrieru, Iu., *Knowledge structurization and formalization for gallbladder ultrasound investigation domain.* Proceedings of the II International Conference on Advanced Information and Telemedicine Technologies for Health (AITTH 2008), Minsk, Belarus, pp. 98–102, ISBN 978-985-6744-45-0, 2008.

[16] Hatzilygeroudis, I., Prentzas, J., *Knowledge representation in intelligent educational systems.*, Web-based intelligent e-learning systems, Information Science Publishing, pp. 175–192, 2006

Iulian Secrieru
Received May 8, 2009

Institute of Mathematics and Computer Science
of Academy of Sciences of Moldova,
MD2028, Moldova, Chisinau,
str.Academiei 5, off. 337
E–mail: *secrieru@math.md*

Computer Science Journal of Moldova, vol.**17**, no.1(49), 2009

On Covering Approximation Subspaces*

Xun Ge

Abstract

Let $(U';\mathcal{C}')$ be a subspace of a covering approximation space $(U;\mathcal{C})$ and $X \subset U'$. In this paper, we show that $\overline{\mathcal{C}'}(X) = \overline{\mathcal{C}}(X)\bigcap U'$ and $B'(X) \subset B(X)\bigcap U'$. Also, $\underline{\mathcal{C}}(X) = \underline{\mathcal{C}'}(X)\bigcap \underline{\mathcal{C}}(U')$ iff $(U;\mathcal{C})$ has Property Multiplication. Furthermore, some connections between outer (resp. inner) definable subsets in $(U;\mathcal{C})$ and outer (resp. inner) definable subsets in $(U';\mathcal{C}')$ are established. These results answer a question on covering approximation subspace posed by J. Li, and are helpful to obtain further applications of Pawlak rough set theory in pattern recognition and artificial intelligence.

Keywords: Rough set; covering approximation subspace; covering approximation operator; definable; outer definable; inner definable.

1 Introduction

In order to extract useful information hidden in voluminous data, many methods in addition to classical logic have been proposed. Pawlak rough-set theory, which was proposed by Z. Pawlak in [11], plays an important role in applications of these methods. Their usefulness has been demonstrated by many successful applications in pattern recognition and artificial intelligence (see [4, 5, 6, 8, 10, 11, 12, 13, 14, 16, 19, 24, 28], for example). In the past years, Pawlak rough-set theory have been extended from Pawlak approximation spaces to covering approximation spaces (see [1, 2, 3, 7, 17, 18, 20, 21, 22, 23, 25, 26, 27, 28, 29], for example).

* This paper is supported by Natural Science Foundation of P. R. China (No.10571151 and 10671173)

Definition 1.1. *Let U be a finite set (a universe of discourse), $\mathcal{C}$ be a cover of U and $X \subset U$. Put*

$$\underline{\mathcal{C}}(X) = \bigcup\{K : K \in \mathcal{C} \bigwedge K \subset X\};$$

$$\overline{\mathcal{C}}(X) = \bigcup\{K : K \in \mathcal{C} \bigwedge K \bigcap X \neq \emptyset\};$$

$$B(X) = X - \underline{\mathcal{C}}(X).$$

(1) $(U; \mathcal{C})$ is called a covering approximation space.
(2) $\underline{C} : 2^U \longrightarrow 2^U$ is called lower covering approximation operator.
(3) $\overline{C} : 2^U \longrightarrow 2^U$ is called upper covering approximation operator.
(4) $\underline{\mathcal{C}}(X)$ is called lower covering approximation of X.
(5) $\overline{\mathcal{C}}(X)$ is called upper covering approximation of X.
(6) $B(X)$ is called boundary of X.
(7) X is called definable in $(U; \mathcal{C})$ if $\overline{\mathcal{C}}(X) = \underline{\mathcal{C}}(X)$.

However, in many applications of Pawlak rough-set theory, we need to consider the case that a cover $\mathcal{C}$ of a universe of discourse U is restricted on some subset U' of U (see [19], for example). More precisely, we are also interested in subspace $(U'; \mathcal{C}')$ of covering approximation space $(U; \mathcal{C})$.

Definition 1.2. *Let $(U; \mathcal{C})$ be a covering approximation space. $(U'; \mathcal{C}')$ is called a subspace of $(U; \mathcal{C})$ if $U' \subset U$ and $\mathcal{C}' = \{K \bigcap U' : K \in \mathcal{C}\}$.*

Remark 1.3. *For a subspace $(U'; \mathcal{C}')$ of a covering approximation space $(U; \mathcal{C})$ and a subset X of U', it is the same as Definition 1.1 to define lower covering approximation operator $\underline{C'}$, upper covering approximation operator $\overline{C'}$, lower covering approximation $\underline{\mathcal{C}'}(X)$ of X, upper covering approximation $\overline{\mathcal{C}'}(X)$ of X, boundary $B'(X)$ of X and definable subsets in $(U'; \mathcal{C}')$. We omit these definitions.*

Let $(U'; \mathcal{C}')$ be a subspace of a covering approximation space $(U; \mathcal{C})$ and $X \subset U'$. It is worthy to give some relations between covering approximations of subsets in $(U; \mathcal{C})$ and covering approximations of subsets in $(U'; \mathcal{C}')$ and to establish some connections between definable subsets in $(U; \mathcal{C})$ and definable subsets in $(U'; \mathcal{C}')$. It is well-known that if $(U;\mathcal{C})$ is a Pawlak approximation space, i.e., $\mathcal{C}$ is a partition of U, then $(U;\mathcal{C})$ is a topological space with a base $\mathcal{C}$. $\underline{\mathcal{C}}(X)$, $\overline{\mathcal{C}}(X)$ and $B(X)$ are exactly interior of X, closure of X and boundary of X in $(U;\mathcal{C})$, respectively (see [7, 15, 26], for example). Thus, $(U';\mathcal{C}')$ is a topological subspace of $(U;\mathcal{C})$ with a base $\mathcal{C}'$. $\underline{\mathcal{C}'}(X)$, $\overline{\mathcal{C}'}(X)$ and $B'(X)$ are exactly interior of X, closure of X and boundary of X in $(U';\mathcal{C}')$, respectively. So the following results are obtained naturally.

Proposition 1.4. *Let $(U';\mathcal{C}')$ be a subspace of a Pawlak approximation space $(U;\mathcal{C})$, and $X \subset U'$. Then the following hold.*

(1) $\overline{\mathcal{C}'}(X) = \overline{\mathcal{C}}(X) \bigcap U'$.

(2) $\underline{\mathcal{C}}(X) = \underline{\mathcal{C}'}(X) \bigcap \underline{\mathcal{C}}(U')$.

(3) $B'(X) \subset B(X) \bigcap U'$.

(4) If U' is definable in $(U;\mathcal{C})$, then X is definable in $(U;\mathcal{C})$ iff X is definable in $(U';\mathcal{C}')$.

By viewing Proposition 1.4, J. Li raised the following question in [9].

Question 1.5. *If $(U;\mathcal{C})$ is a covering approximation space, does Proposition 1.4 hold?*

In this paper, we investigate and answer Question 1.5. For a subspace $(U';\mathcal{C}')$ of a covering approximation space $(U;\mathcal{C})$ and a subset X of U', we show that $\overline{\mathcal{C}'}(X) = \overline{\mathcal{C}}(X) \bigcap U'$ and $B'(X) \subset B(X) \bigcap U'$. Also, $\underline{\mathcal{C}}(X) = \underline{\mathcal{C}'}(X) \bigcap \underline{\mathcal{C}}(U')$ iff $(U;\mathcal{C})$ has Property Multiplication. Furthermore, we establish some connections between outer (resp. inner) definable subsets in $(U;\mathcal{C})$ and outer (resp. inner) definable subsets in $(U';\mathcal{C}')$. These results are helpful to obtain further applications of Pawlak rough set theory in pattern recognition and artificial intelligence.

2 On Covering Approximations of subsets

The following lemma is known (see [18, 29], for example).

Lemma 2.1. *Let $(U;\mathcal{C})$ be a covering approximation space. Then the following hold.*

(1) If $X \subset U$, then $\underline{\mathcal{C}}(X) \subset X \subset \overline{\mathcal{C}}(X)$.
(2) If $X \subset Y \subset U$, then $\overline{\mathcal{C}}(X) \subset \overline{\mathcal{C}}(Y)$ and $\underline{\mathcal{C}}(X) \subset \underline{\mathcal{C}}(Y)$.
(3) If $X, Y \subset U$, then $\underline{\mathcal{C}}(X \bigcap Y) \subset \underline{\mathcal{C}}(X) \bigcap \underline{\mathcal{C}}(Y)$.
(4) If X is a union of some elements of $\mathcal{C}$, then $\underline{\mathcal{C}}(X) = X$.
(5) $\underline{\mathcal{C}}(U) = \overline{\mathcal{C}}(U) = U$.

Theorem 2.2. *Let $(U';\mathcal{C}')$ be a subspace of a covering approximation space $(U;\mathcal{C})$ and $X \subset U'$. Then the following hold.*
(1) $\overline{\mathcal{C}'}(X) = \overline{\mathcal{C}}(X) \bigcap U'$.
(2) $B'(X) \subset B(X) \bigcap U'$.

Proof. (1) Let $x \in \overline{\mathcal{C}'}(X)$, then there exists $K \in \mathcal{C}$ such that $x \in K \bigcap U'$ and $(K \bigcap U') \bigcap X \neq \emptyset$, so $x \in K$ and $K \bigcap X \neq \emptyset$. Thus $x \in \overline{\mathcal{C}}(X)$ and $x \in U'$, i.e., $x \in \overline{\mathcal{C}}(X) \bigcap U'$. On the other hand, let $x \in \overline{\mathcal{C}}(X) \bigcap U'$, then there exists $K \in \mathcal{C}$ such that $x \in K$ and $K \bigcap X \neq \emptyset$. Since $X \subset U'$, $(K \bigcap U') \bigcap X = K \bigcap X \neq \emptyset$. Note that $x \in K \bigcap U'$ and $K \bigcap U' \in \mathcal{C}'$. So $x \in \overline{\mathcal{C}'}(X)$.

(2) Since $B'(x) = U' - \underline{\mathcal{C}'}(X)$ and $B(X) \bigcap U' = (U - \underline{\mathcal{C}}(X)) \bigcap U' = U' - (\underline{\mathcal{C}}(X) \bigcap U')$, it suffices to prove that $\underline{\mathcal{C}}(X) \bigcap U' \subset \underline{\mathcal{C}'}(X)$. Let $x \in \underline{\mathcal{C}}(X) \bigcap U'$, then $x \in U'$ and there exists $K \in \mathcal{C}$ such that $x \in K \subset X$, So $x \in K \bigcap U' \subset X$. Note that $K \bigcap U' \in \mathcal{C}'$, so $x \in \underline{\mathcal{C}'}(X)$. This proves that $\underline{\mathcal{C}}(X) \bigcap U' \subset \underline{\mathcal{C}'}(X)$. □

Remark 2.3. *The following example shows that "$\subset$" in Theorem 2.2(2) can not be replaced by "$=$".*

Example 2.4. *There exist a subspace $(U';\mathcal{C}')$ of a covering approximation space $(U;\mathcal{C})$ and a subset X of U' such that $B'(X) \neq B(X) \bigcap U'$.*

Proof. Let $U = \{a, b, c\}$, $\mathcal{C} = \{\{a, b\}, \{b, c\}\}$, $U' = \{a, b\}$ and $\mathcal{C}' = \{\{a, b\}, \{b\}\}$, then $(U';\mathcal{C}')$ is a subspace $(U;\mathcal{C})$. Put $X = \{b\}$, then $X \subset U'$.

(1) $\underline{\mathcal{C}}(X) = \emptyset$, so $B(X) = X - \underline{\mathcal{C}}(X) = X$.
(2) $\underline{\mathcal{C}'}(X) = X$, so $B'(X) = X - \underline{\mathcal{C}'}(X) = \emptyset$.
Consequently, $B'(X) \neq B(X) \bigcap U'$. □

In general, Proposition 1.4(2) does not hold for covering approximation spaces (see [18, 29], for example). We give a sufficient and necessary condition such that it holds.

Definition 2.5. *Let $(U;\mathcal{C})$ be a covering approximation space. $(U;\mathcal{C})$ is called to have Property Multiplication (Property (M), in brief), if $\underline{C}(X\bigcap Y) = \underline{C}(X)\bigcap\underline{C}(Y)$ for any $X, Y \subset U$.*

Remark 2.6. *Every Pawlak approximation space has Property (M). In general, covering approximation spaces have not Property (M) (see [26, Proposition 4].*

The following lemma comes from [26, Theorem 1].

Lemma 2.7. *Let $(U;\mathcal{C})$ be a covering approximation space. Then the following are equivalent.*

(1) $(U;\mathcal{C})$ has Property (M).

(2) If $K_1, K_2 \in \mathcal{C}$ and $x \in K_1\bigcap K_2$, then there exists $K \in \mathcal{C}$ such that $x \in K \subset K_1\bigcap K_2$.

Theorem 2.8. *Let $(U;\mathcal{C})$ be a covering approximation space. Then the following are equivalent.*

(1) $(U;\mathcal{C})$ has Property (M).

(2) If $(U';\mathcal{C}')$ is a subspace of $(U;\mathcal{C})$ and $X \subset U'$, then $\underline{\mathcal{C}}(X) = \underline{\mathcal{C}'}(X)\bigcap\underline{\mathcal{C}}(U')$.

Proof. (1) $\Longrightarrow$ (2): Let $(U';\mathcal{C}')$ be a subspace of $(U;\mathcal{C})$ and $X \subset U'$. If $x \in \underline{\mathcal{C}}(X)$, then there exists $K \in \mathcal{C}$ such that $x \in K \subset X \subset U'$, so $x \in \underline{\mathcal{C}}(U')$. Note that $K\bigcap U' = K$, so $K \in \mathcal{C}'$, thus $x \in \underline{\mathcal{C}'}(X)$. Consequently, $x \in \underline{\mathcal{C}'}(X)\bigcap\underline{\mathcal{C}}(U')$. On the other hand, if $x \in \underline{\mathcal{C}'}(X)\bigcap\underline{\mathcal{C}}(U')$, then there exist $K_1, K_2 \in \mathcal{C}$ such that $x \in K_1\bigcap U' \subset X$ and $x \in K_2 \subset U'$. Since $(U;\mathcal{C})$ has Property (M), by Lemma 2.7, there exists $K \in \mathcal{C}$ such that $x \in K \subset K_1\bigcap K_2$. So $x \in K \subset K_1\bigcap K_2 \subset K_1\bigcap U' \subset X$. Thus $x \in \underline{\mathcal{C}}(X)$. This proves that $\underline{\mathcal{C}}(X) = \underline{\mathcal{C}'}(X)\bigcap\underline{\mathcal{C}}(U')$.

(2) $\Longrightarrow$ (1): Let $K_1, K_2 \in \mathcal{C}$ and $x \in K_1\bigcap K_2$. Put $U' = K_1$ and $\mathcal{C}' = \{K\bigcap U' : K \in \mathcal{C}\}$, then $(U';\mathcal{C}')$ is a subspace of $(U;\mathcal{C})$. Put $X = K_1\bigcap K_2 = K_2\bigcap U'$, then $x \in X \in \mathcal{C}'$. So $x \in X = \underline{\mathcal{C}'}(X)$ from

Lemma 2.1(4). On the other hand, $x \in K_1 = \underline{\mathcal{C}}(K_1) = \underline{\mathcal{C}}(U')$ from Lemma 2.1(4). Thus $x \in \underline{\mathcal{C}'}(X) \bigcap \underline{\mathcal{C}}(U')$. Since $\underline{\mathcal{C}}(X) = \underline{\mathcal{C}'}(X) \bigcap \underline{\mathcal{C}}(U')$, $x \in \underline{\mathcal{C}}(X)$, and so there exists $K \in \mathcal{C}$ such that $x \in K \subset X = K_1 \bigcap K_2$. By Lemma 2.7, $(U; \mathcal{C})$ has Property (M). □

Remark 2.9. *In the proof of Theorem 2.8(1) $\Longrightarrow$ (2), we can see that $\underline{\mathcal{C}}(X) \subset \underline{\mathcal{C}'}(X) \bigcap \underline{\mathcal{C}}(U')$ without requiring Property (M), and so $\underline{\mathcal{C}}(X) \subset \underline{\mathcal{C}'}(X)$ without requiring Property (M).*

3 On Outer and Inner Definable Subsets

As some applications of Theorem 2.2 and Theorem 2.8, we investigate definable subsets in covering approximation subspaces. The following definitions come from [15]

Definition 3.1. *Let $(U; \mathcal{C})$ be a covering approximation space and $X \subset U$.*

(1) X is called outer definable in $(U; \mathcal{C})$ if $\overline{\mathcal{C}}(X) = X$.

(2) X is called inner definable in $(U; \mathcal{C})$ if $\underline{\mathcal{C}}(X) = X$.

Remark 3.2. *It is easy to see that X is definable in $(U, \mathcal{C})$ iff it is both outer definable and inner definable in $(U; \mathcal{C})$.*

Lemma 3.3. *Let $(U; \mathcal{C})$ be a covering approximation space and $X \subset U$. Consider the following conditions.*

(1) X is definable in $(U, \mathcal{C})$.

(2) X is outer definable in $(U, \mathcal{C})$.

(3) X is inner definable in $(U, \mathcal{C})$.

Then (1) $\Longleftrightarrow$ (2) $\Longrightarrow$ (3).

Proof. By Remark 3.2, (1) $\Longrightarrow$ (2) and (1) $\Longrightarrow$ (3). It suffices to prove (2) $\Longrightarrow$ (1).

Let X be outer definable in $(U; \mathcal{C})$, i.e., $\overline{\mathcal{C}}(X) = X$. Let $x \in X$, then there is $K \in \mathcal{C}$ such that $x \in K$. So $K \bigcap X \neq \emptyset$, and hence $K \subset \overline{\mathcal{C}}(X) = X$. It follows that $x \in K \subset \underline{\mathcal{C}}(X)$. This proves that $X \subset \underline{\mathcal{C}}(X)$. By Lemma 2.1(1), $\underline{\mathcal{C}}(X) \subset X$, so $\underline{\mathcal{C}}(X) = X$. Consequently, X is inner definable in $(U, \mathcal{C})$. □

Remark 3.4. *(1) In Lemma 3.3, (3)* $\not\Longrightarrow$ *(2) (see Example 3.5).*

(2) If $(U;\mathcal{C})$ *is a Pawlak approximation space and* $X \subset U$*, then (1), (2) and (3) in Lemma 3.3 are equivalent ([15]).*

Example 3.5. *There exist a covering approximation space* $(U;\mathcal{C})$ *and a subset* X *of* U *such that* X *is inner definable in* $(U;\mathcal{C})$*, but* X *is not outer definable in* $(U;\mathcal{C})$*.*

Proof. Let $U = \{a,b,c\}$, $\mathcal{C} = \{\{a,b\},\{b,c\}\}$, $X = \{a,b\}$.

(1) Since $X \in \mathcal{C}$, $\underline{\mathcal{C}}(X) = X$ from Lemma 2.1(4), so X is inner definable in $(U;\mathcal{C})$.

(2) It is easy to see that, $\overline{\mathcal{C}}(X) = U \neq X$, so X is not outer definable in $(U;\mathcal{C})$. □

By Lemma 3.3, "outer definable" can be replaced by "definable" throughout the following.

Theorem 3.6. *Let* $(U';\mathcal{C}')$ *be a subspace of a covering approximation space* $(U;\mathcal{C})$ *and* $X \subset U$*. Then the following hold.*

(1) If X *is outer definable in* $(U;\mathcal{C})$*, then* $X \bigcap U'$ *is outer definable in* $(U';\mathcal{C}')$*.*

(2) If X *is inner definable in* $(U;\mathcal{C})$*, then* $X \bigcap U'$ *is inner definable in* $(U';\mathcal{C}')$*.*

Proof. (1) Let X is outer definable in $(U;\mathcal{C})$, i.e., $\overline{\mathcal{C}}(X) = X$. By Theorem 2.2(1), Lemma 2.1(3) and Lemma 2.1(1), $\overline{\mathcal{C}'}(X \bigcap U') = \overline{\mathcal{C}}(X \bigcap U') \bigcap U' \subset \overline{\mathcal{C}}(X) \bigcap \overline{\mathcal{C}}(U') \bigcap U' = X \bigcap U'$. On the other hand, $X \bigcap U' \subset \overline{\mathcal{C}'}(X \bigcap U')$ from Lemma 2.1(1). Thus $\overline{\mathcal{C}'}(X \bigcap U') = X \bigcap U'$, so $X \bigcap U'$ is outer definable in $(U';\mathcal{C}')$.

(2) Let X is inner definable in $(U;\mathcal{C})$, i.e., $\underline{\mathcal{C}}(X) = X$. Then $X \bigcap U' = \underline{\mathcal{C}}(X) \bigcap U' = (\bigcup\{K : K \in \mathcal{C} \bigwedge K \subset X\}) \bigcap U' = \bigcup\{K \bigcap U' : K \in \mathcal{C} \bigwedge K \subset X\} \subset \bigcup\{K \bigcap U' : K \in \mathcal{C} \bigwedge K \bigcap U' \subset X \bigcap U'\} = \underline{\mathcal{C}'}(X \bigcap U')$. On the other hand, $\underline{\mathcal{C}'}(X \bigcap U') \subset X \bigcap U'$ from Lemma 2.1(1). Thus $\underline{\mathcal{C}'}(X \bigcap U') = X \bigcap U'$, so $X \bigcap U'$ is inner definable in $(U';\mathcal{C}')$. □

Remark 3.7. *The following example shows that both (1) and (2) in Theorem 3.6 can not be reversed even if* $(U;\mathcal{C})$ *is a Pawlak approximation space.*

Example 3.8. *There exist a subspace* $(U';\mathcal{C}')$ *of a Pawlak approximation space* $(U;\mathcal{C})$ *and a subset* X *of* U*, where* U' *is outer definable in* $(U;\mathcal{C})$*, such that* $X\bigcap U'$ *is outer definable in* $(U';\mathcal{C}')$*, but* X *is not inner definable in* $(U;\mathcal{C})$*.*

Proof. Let $U=\{a,b,c,d\}$, $\mathcal{C}=\{\{a,b\},\{c,d\}\}$, then $(U;\mathcal{C})$ is a Pawlak approximation space. Put $U'=\{a,b\}$ and $\mathcal{C}'=\{\{a,b\}\}$, then $(U';\mathcal{C}')$ is a subspace $(U;\mathcal{C})$. Put $X=\{a,b,c\}$.

(1) It is clear that U' is outer definable in $(U;\mathcal{C})$.

(2) Since $X\bigcap U'=U'$, $X\bigcap U'$ is outer definable in $(U';\mathcal{C}')$.

(3) It is easy to see that $\underline{\mathcal{C}}(X)=U'\neq X$, so X is not inner definable in $(U;\mathcal{C})$. □

However, we have the following results.

Theorem 3.9. *Let* $(U';\mathcal{C}')$ *be a subspace of a covering approximation space* $(U;\mathcal{C})$ *and* $X\subset U'$*. If* U' *is outer definable in* $(U;\mathcal{C})$*, then the following are equivalent.*

(1) X *is outer definable in* $(U;\mathcal{C})$*.*

(2) X *is outer definable in* $(U';\mathcal{C}')$*.*

Proof. (1) $\Longrightarrow$ (2): It holds from Theorem 3.6(1).

(2) $\Longrightarrow$ (1): Let X be outer definable in $(U';\mathcal{C}')$, i.e., $\overline{\mathcal{C}'}(X)=X$. Since U' is outer definable in $(U;\mathcal{C})$, $\overline{\mathcal{C}}(U')=U'$. $\overline{\mathcal{C}}(X)\subset\overline{\mathcal{C}}(U')=U'$ from Lemma 2.1(2). By Theorem 2.2(1), $\overline{\mathcal{C}}(X)=\overline{\mathcal{C}}(X)\bigcap U'=\overline{\mathcal{C}'}(X)=X$. So X is outer definable in $(U;\mathcal{C})$. □

Remark 3.10. *(1) By Theorem 3.6(1), the condition "*U' *is outer definable in* $(U;\mathcal{C})$*" in Theorem 3.9(1)* $\Longrightarrow$ *(2) can be omitted.*

(2) The condition "U' *is outer definable in* $(U;\mathcal{C})$*" in Theorem 3.9(2)* $\Longrightarrow$ *(1) can not be relaxed to "*U' *is inner definable in* $(U;\mathcal{C})$*" (see Example 3.11).*

Example 3.11. *There exist a subspace $(U';\mathcal{C}')$ of a covering approximation space $(U;\mathcal{C})$ and a subset X of U' such that $(U;\mathcal{C})$ has Property (M), U' is inner definable in $(U;\mathcal{C})$, and X is outer definable in $(U';\mathcal{C}')$, but X is not outer definable in $(U;\mathcal{C})$.*

Proof. Let $U = \{a, b, c\}$, $\mathcal{C} = \{\{a, b\}, \{b, c\}, \{b\}\}$. Put $U' = X = \{a, b\}$ and $\mathcal{C}' = \{\{a, b\}, \{b\}\}$, then $(U';\mathcal{C}')$ is a subspace $(U;\mathcal{C})$.

(1) Using Lemma 2.7, it is easy to check that $(U;\mathcal{C})$ has Property (M).

(2) It is clear that U' is inner definable in $(U;\mathcal{C})$.

(3) Since $X \subset \overline{\mathcal{C}'}(X) \subset U' = X$, $\overline{\mathcal{C}'}(X) = X$, so X is outer definable in $(U';\mathcal{C}')$.

(4) $\overline{\mathcal{C}}(X) = U \neq X$, so X is not outer definable in $(U;\mathcal{C})$. □

Theorem 3.12. *Let $(U';\mathcal{C}')$ be a subspace of a covering approximation space $(U;\mathcal{C})$ and $X \subset U'$. If $(U;\mathcal{C})$ has Property (M) and U' is inner definable in $(U;\mathcal{C})$, then the following are equivalent.*

(1) X is inner definable in $(U;\mathcal{C})$.

(2) X is inner definable in $(U';\mathcal{C}')$.

Proof. (1) $\Longrightarrow$ (2): It holds from Theorem 3.6(2).

(2) $\Longrightarrow$ (1): Let X be inner definable in $(U';\mathcal{C}')$, i.e., $\underline{\mathcal{C}'}(X) = X$. Since $(U;\mathcal{C})$ has Property (M), $\underline{\mathcal{C}}(X) = \underline{\mathcal{C}'}(X) \bigcap \underline{\mathcal{C}}(U')$ from Theorem 2.8. Note that $\underline{\mathcal{C}}(U') = U'$ because U' is inner definable in $(U;\mathcal{C})$. Thus $\underline{\mathcal{C}}(X) = \underline{\mathcal{C}'}(X) \bigcap \underline{\mathcal{C}}(U') = X \bigcap \underline{U'} = X$. So X is inner definable in $(U;\mathcal{C})$. □

Remark 3.13. *(1) By Theorem 3.6(2), both condition "$(U;\mathcal{C})$ has Property (M)" and condition "U' is inner definable in $(U;\mathcal{C})$" in Theorem 3.12(1) $\Longrightarrow$ (2) can be omitted.*

(2) The condition "$(U;\mathcal{C})$ has Property (M)" in Theorem 3.12(2) $\Longrightarrow$ (1) can not be omitted (see Example 3.14).

(3) The condition "U' is inner definable in $(U;\mathcal{C})$" in Theorem 3.12(2) $\Longrightarrow$ (1) can not be omitted (see Example 3.15).

Example 3.14. *There exist a subspace $(U';\mathcal{C}')$ of a covering approximation space $(U;\mathcal{C})$ and a subset X of U', where U' is inner definable* *in* $(U;\mathcal{C})$*, such that* X *is inner definable in* $(U';\mathcal{C}')$*, but* X *is not inner definable in* $(U;\mathcal{C})$*.*

Proof. Let $U = \{a,b,c\}$, $\mathcal{C} = \{\{a,b\},\{b,c\}\}$. Put $U' = \{a,b\}$ and $\mathcal{C}' = \{\{a,b\},\{b\}\}$, then $(U';\mathcal{C}')$ is a subspace $(U;\mathcal{C})$. Put $X = \{b\}$, then $X \subset U'$.

(1) Since $U' \in \mathcal{C}$, $\underline{\mathcal{C}}(U') = U'$ from Lemma 2.1(4), so U' is inner definable in $(U;\mathcal{C})$.

(2) Since $X \in \mathcal{C}'$, $\underline{\mathcal{C}'}(X) = X$ from Lemma 2.1(4), so X is inner definable in $(U';\mathcal{C}')$.

(3) $\underline{\mathcal{C}}(X) = \emptyset \neq X$, so X is not inner definable in $(U;\mathcal{C})$. □

Example 3.15. *There exist a subspace* $(U';\mathcal{C}')$ *of a covering approximation space* $(U;\mathcal{C})$ *and a subset* X *of* U'*, where* $(U;\mathcal{C})$ *has Property (M), such that* X *is inner definable in* $(U';\mathcal{C}')$*, but* X *is not inner definable in* $(U;\mathcal{C})$*.*

Proof. Let $U = \{a,b,c\}$, $\mathcal{C} = \{\{a,b\},\{b,c\},\{b\}\}$. Put $U' = X = \{a,c\}$ and $\mathcal{C}' = \{\{a\},\{c\}\}$, then $(U';\mathcal{C}')$ is a subspace $(U;\mathcal{C})$.

(1) $(U;\mathcal{C})$ has Property (M) from Example 3.11.

(2) Since $\underline{\mathcal{C}'}(U') = U'$ from Lemma 2.1(5), $\underline{\mathcal{C}'}(X) = \underline{\mathcal{C}'}(U') = U' = X$, so X is inner definable in $(U';\mathcal{C}')$.

(3) $\underline{\mathcal{C}}(X) = \emptyset \neq X$, so X is not inner definable in $(U;\mathcal{C})$. □

4 Postscript

In this paper, our investigations on covering approximation subspaces are based on lower covering approximation operator $\underline{\mathcal{C}}$ and upper covering approximation operator $\overline{\mathcal{C}}$, which are endowed covering approximation spaces. Because there are also other covering approximation operators (see the following Definition 4.1), It is an interesting work to give some answers of Question 1.5 for these covering approximation operators.

Definition 4.1. *Let* $(U;\mathcal{C})$ *be a covering approximation space. For each* $x \in U$*, put*

$$Md(x) = \{K : (x \in K \in \mathcal{C}) \bigwedge (x \in S \in \mathcal{C} \bigwedge S \subset K \Longrightarrow S = K)\};$$

$$N(x) = \bigcap\{K : x \in K \in \mathcal{C}\}.$$

For each $i \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $\underline{C_i}$ and $\overline{C_i}$ are defined as follows and are called i-th lower covering approximation operator and i-th upper covering approximation operator on $(U; \mathcal{C})$, respectively.

(1) $\underline{\mathcal{C}_1}(X) = \bigcup\{K : K \in \mathcal{C} \bigwedge K \subset X\}$;
$\overline{\mathcal{C}_1}(X) = \underline{\mathcal{C}_1}(X) \bigcup(\bigcup\{\bigcup Md(x) : x \in X - \underline{\mathcal{C}_1}(X)\})$.

(2) $\underline{\mathcal{C}_2}(X) = \{x \in U : \forall K \in \mathcal{C}(x \in K \Longrightarrow K \subset X)\}$;
$\overline{\mathcal{C}_2}(X) = \bigcup\{K : K \in \mathcal{C} \bigwedge K \bigcap X \neq \emptyset\}$.

(3) $\underline{\mathcal{C}_3}(X) = \bigcup\{K : K \in \mathcal{C} \bigwedge K \subset X\}$;
$\overline{\mathcal{C}_3}(X) = \bigcup\{\bigcup Md(x) : x \in X\}$.

(4) $\underline{\mathcal{C}_4}(X) = \bigcup\{K : K \in \mathcal{C} \bigwedge K \subset X\}$;
$\overline{\mathcal{C}_4}(X) = \underline{\mathcal{C}_4}(X) \bigcup(\bigcup\{K : K \in \mathcal{C} \bigwedge K \bigcap(X - \underline{\mathcal{C}_4}(X)) \neq \emptyset\})$.

(5) $\underline{\mathcal{C}_5}(X) = \bigcup\{K : K \in \mathcal{C} \bigwedge K \subset X\}$;
$\overline{\mathcal{C}_5}(X) = \underline{\mathcal{C}_5}(X) \bigcup(\bigcup\{N(x) : x \in X - \underline{\mathcal{C}_5}(X)\})$.

(6) $\underline{\mathcal{C}_6}(X) = \{x \in U : N(x) \subset X\}$;
$\overline{\mathcal{C}_6}(X) = \{x \in U : N(x) \bigcap X \neq \emptyset\}$.

(7) $\underline{\mathcal{C}_7}(X) = \bigcup\{K : K \in \mathcal{C} \bigwedge K \subset X\}$;
$\overline{\mathcal{C}_7}(X) = U - \underline{\mathcal{C}_7}(U - X)$.

(8) $\underline{\mathcal{C}_8}(X) = \{x \in U : \exists u(u \in N(x) \bigwedge N(u) \subset X)\}$;
$\overline{\mathcal{C}_8}(X) = \{x \in U : \forall u(u \in N(x) \Longrightarrow N(u) \bigcap X \neq \emptyset)\}$.

(9) $\underline{\mathcal{C}_9}(X) = \{x \in U : \forall u(x \in N(u) \Longrightarrow N(u) \subset X)\}$;
$\overline{\mathcal{C}_9}(X) = \bigcup\{N(x) : x \in U \bigwedge N(x) \bigcap X \neq \emptyset\}$.

(10) $\underline{\mathcal{C}_{10}}(X) = \{x \in U : \forall u(x \in N(u) \Longrightarrow u \in X)\}$;
$\overline{\mathcal{C}_{10}}(X) = \bigcup\{N(x) : x \in X\}$.

Remark 4.2. *$\underline{C_i}$ and $\overline{C_i}$ (i=1,3) come from [29]; $\underline{C_2}$ and $\overline{C_2}$ come from [17]; $\underline{C_4}$ and $\overline{C_4}$ come from [26]; $\underline{C_5}$ and $\overline{C_5}$ come from [27]; $\underline{C_6}$ and $\overline{C_6}$ come from [18, 28]; $\underline{C_i}$ and $\overline{C_i}$ (i=7,8,9,10) come from [18].*

Thus, we have the following question, which is still worthy to be considered in subsequent research.

Question 4.3. *Let* $(U';\mathcal{C}')$ *be a subspace of a covering approximation space* $(U;\mathcal{C})$*,* $X \subset U'$ *and* $i = 1,2,3,4,5,6,7,8,9,10$*. Do following hold?*

(1) $\overline{\mathcal{C}'_i}(X) = \overline{\mathcal{C}_i}(X) \bigcap U'$*.*

(2) $\underline{\mathcal{C}_i}(X) = \underline{\mathcal{C}'_i}(X) \bigcap \underline{\mathcal{C}_i}(U')$*.*

(3) $B'_i(X) \subset B_i(X) \bigcap U'$*.*

(4) If U' *is definable in* $(U;\mathcal{C})$*, then* X *is definable in* $(U;\mathcal{C})$ *iff* X *is definable in* $(U';\mathcal{C}')$*.*

Acknowledgments. The author would like to thank the referee for his/her valuable amendments and corrections.

References

[1] Z.Bonikowski, E.Bryniarski and U.Wybraniec, *Extensions and intentions in the rough set theory*, Information Sciences, 107(1998), 149–167.

[2] X.Ge, *Granularity-wise separations in covering approximation spaces with some approximation operators*, 2008 International Symposium on Computer Science and Computational Technology, 2008, 303–306.

[3] Y.Ge, *Granularity-wise separations in covering approximation spaces*, 2008 IEEE International Conference on Granular Computing, 2008, 238–243.

[4] A.G.Jackson, Z.Pawlak and S.LeClair, *Rough sets applied to the discovery of materials knowledge*, Journal of Alloys and Compounds, 279(1998), 14–21.

[5] M.Kryszkiewicz, *Rough set approach to incomplete information systems*, Information Sciences, 112(1998), 39–49.

[6] M.Kryszkiewicz, *Rule in incomplete information systems*, Information Sciences, 113(1998), 271–292.

[7] E.Lashin, A.Kozae, A.Khadra and T.Medhat, *Rough set theory for topological spaces*, International Journal of Approximate Reasoning, 40(2005), 35–43.

[8] Y.Leung, W.Wu and W.Zhang, *Knowledge acquisition in incomplete information systems: A rough set approach*, European Journal of Operational Research, 168(2006), 164–180.

[9] J.Li, A discemibility matrix for the topological reduction, *2007 International General Topology Symposinm, in Nanning, P.R.China*, 2007.

[10] M.Liu, M.Shao, W.Zhang and C.Wu, *Reduction method for concept lattices based on rough set theory and its application*, Computers and Mathematics with Applications, 53(2007), 1390–1410.

[11] Z.Pawlak, *Rough sets*, International Journal of Computer and Information Sciences, 11(1982), 341-356.

[12] Z.Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*, Kluwer Academic Publishers, 1991.

[13] Z.Pawlak, *Rough classification*, Int. J. Human-computer Studies, 51(1999), 369–383.

[14] Z.Pawlak, *Theorize with data using rough sets*, Proceedings of the 26-th Annual International Computer Software and Applications Conference, 2002.

[15] D.Pei, On definable concepts of rough set models, *Information Sciences*, 177(2007), 4230–4239.

[16] J.F.Peters1, Z.Pawlak and A.Skowron, *A rough set approach to measuring information granules*, Proceedings of the 26-th Annual International Computer Software and Applications Conference, 2002.

[17] J.A.Pomykala, *Approximation operations in approximation spaces*, Bull. Pol. Acad. Sci, 35(9-10)(1987), 653–662.

[18] K.Qin, Y.Gao and Z.Pei, *On covering rough sets*, RSKT 2007, LNAI, Vol. 4481(2007), 34–41.

[19] S.Yang, J.Qian and Y.Ge, *Uncertain decision analysis in stocks prices fluctuations*, J. of Suzhou University, 24(2008), 6–10.

[20] Y.Yao, *Views of the theory of rough sets in finite universes*, International Journal of Approximate Reasoning, 15(1996), 291–317.

[21] Y.Yao, Constructive and algebraic methods of theory of rough sets, Information Sciences, 109(1998), 21–47.

[22] Y.Yao, *Relational interpretations of neighborhood operators and rough set approximation operators*, Information Sciences, 111(1998) 239C259.

[23] Y.Yao, *On generalizing rough set theory*, Lecture Notes in AI, 2639(2003), 44–51.

[24] Y.Yao and N.Zhong, *Potential applications of granular computing in knowledge discovery and data mining.* Proceedings of World Multiconference on Systemics, Cybernetics and Informatics, 1999, 573–580.

[25] W.Zakowski, *Approximations in the space* $(U; \prod)$, Demonstratio Mathematica, 16(1983), 761–769

[26] W.Zhu, *Properties of the fourth type of covering based rough sets*, Proceedings of the Sixth International Conference on Hybrid Intelligent Systems, 2006.

[27] W.Zhu, *Topological approaches to covering rough sets*, Information Sciences, 177(2007), 1499–1508.

[28] W.Zhu, *Relationship between generalized rough sets based on binary relation and covering*, Information Sciences, 179(2009), 210C225.

[29] W.Zhu and F.Y.Wang, *On Three Types of Covering Rough Sets*, IEEE Transactions on Knowledge and Data Engineering, 19(2007), 1131–1144.

Xun Ge,
Received March, 18, 2009

Zhangjiagang college,
Jiangsu University of Science and Technology,
Zhangjiagang 215600, P. R. China
E–mail: *zhugexun*@163.*com*

Computer Science Journal of Moldova, vol.17, no.1(49), 2009

An approach for testing the primeness of attributes in relational schemas

Cotelea Vitalie

Abstract

In this paper there is proposed a method of partition the attributes of relation scheme in equivalence classes and in nonredundant equivalence classes. Several properties of these equivalence classes are proved. Their properties serve as the basis for an algorithm with a polynomial complexity, which determines the prime attributes of a database schema.

Keywords: Relation scheme, functional dependencies, equivalence classes, prime attributes, polynomial complexity tasks.

1 Introduction

The scope of this paper is to propose a solution to the problem that arises during design and analysis of database, that is determination of prime attributes (attributes that are contained in schema's possible keys). This problem is known to be NP-complete, due to the fact that the solution to this problem was reached through keys searching. But a schema can have an exponential number of keys with respect to number of functional dependencies [1].

In the current paper a different approach is taken for the searching of prime attributes that avoids the necessity of keys determination. Namely, the notion of contribution graph (Definition 1) of a reduced set of functional dependencies is proposed. The strongly connected components are computed, where each component represents a vertex of condensed graph (Definition 2). Over vertices of condensed graph a strict partial order is defined. Then it's presented how the inferred dependencies are reflected in contribution graph (Lemma 1 and Corollary 1).

Obviously, the strongly connected components of contribution graph split the set of attributes of relation scheme into equivalence classes. The notion of nonredundant equivalence classes of attributes is given (Definition 3). In section 4 several lemmas and theorems (Lemmas 2-3, Theorems 1-4) are proved that reflect the properties of equivalence classes of attributes.

It should be mentioned that redundant attributes represent the set of nonprime attributes of scheme (Corollary 4), and nonredundant equivalence classes of attributes consist only of prime attributes (Corollary 3). Proved properties in section 4, allow the determination of prime and nonprime attributes without scheme's keys finding.

In section 5 it is shown that the determination of prime and nonprime attributes can be performed in a polynomial time. This approach can be a part of the database analysis and design toolset.

2 Some basic concepts

In order to facilitate exposure of this paper's material, some preliminary notions are presented [2].

Let $Sch(R, F)$ be a relation scheme, where F is a set of functional dependencies defined on set R of attributes. Given a set F of functional dependencies on R, the *closure* of F, written as F^+, consists of all functional dependencies that are logically implied by F, that is $F^+ =$ $= \{V \to W | F| = V \to W\}$.

Given a set F of functional dependencies on set R of attributes and a subset X of R, the *closure* of the set X under the set F, written as X^+, contains all attributes, each of which is functionally dependent on Xunder F, that is $X^+ = \{A | X \to A \in F^+\}$.

Let X and Y be two sets of attributes, where $X, Y \subseteq R$. The set X is a determinant for Y, under the set F of functional dependencies, if $X \to Y \in F^+$ and for every proper subset X' of the set X, the expression $X' \to Y \notin F^+$ takes place.

A subset K of R is a key for a relation scheme $Sch(R,F)$, if K is a determinant of the set R under the set F of dependencies. A relation scheme can have more than one key, but it always has at least one.

An attribute A in R is *prime* if A belongs to some key, and *nonprime* otherwise.

In this paper, it is considered that the set F of functional dependencies is reduced. Let $Sch(R,F)$ be a relation scheme. The set of functional dependencies F is reduced [2], if there is no attribute A in R and no dependency $X \to Y$ in F, so that they satisfy the following conditions:

1. $A \in X$ and $F \equiv F - \{X \to Y\} \bigcup \{(X - \{A\}) \to Y\}$,
2. $A \in Y$ and $F \equiv F - \{X \to Y\} \bigcup \{X \to (Y - \{A\})\}$.

For functional dependencies an inference tool, named maximal derivation [3], will be used. Maximal derivation of the set X of attributes under the set F of dependencies, is a sequence of sets $H =< X_0, X_1, ..., X_n >$ of attributes, where

1. $X_0 = X$;
2. $X_i = X_{i-1} \bigcup Z$, where $Z = \bigcup_j W_j$ for $\forall V_j \to W_j \in F$ that satisfy $V_j \subseteq X_{i-1}$ and $W_j \not\subset X_{i-1}$;
3. Nothing else is in X_i.

The last term of maximal derivation X_n is, in fact, the closure of the set X of attributes under the set F of dependencies, that is $X_n = X^+$.

Claim 1. [3]. *$X \to Y \in F^+$, if and only if there exists a derivation $H =< X_0, X_1, ..., X_k >$ for $X \to Y$ under F, where X_k is the first term that contains the set of attributes Y.*

Claim 2. [3]. *If $X \to Y \in F^+$ and X is a determinant for Y under F, then for every attribute A in $X - Y$ there exists in F a dependency $V \to W$ used in derivation $H =< X_0, X_1, ..., X_k >$ for $X \to Y$ under F, such that $A \in V$.*

3 Graphical representation of functional dependencies

Given a set F of functional dependencies on the set R of attributes, that are part of the relation scheme $Sch(R, F)$, a contribution graph is drawn, in order to represent F.

Definition 1. *Contribution graph $G = (S, E)$ of set F is a graph that:*

- *$\forall A \in R$ there exists in S a vertex labeled with attribute A;*
- *$\forall X \to Y \in F$ and $\forall A \in X$ and $\forall B \in Y$ there exists in E an edge $a = (A, B)$, that is directed from vertex A to vertex B.*

Example 1. *If $F = \{C \to B, AD \to B, AB \to DC, B \to E\}$ and $R = \{A, B, C, D, E\}$ then the contribution graph of set F of dependencies is presented in Figure 1.*

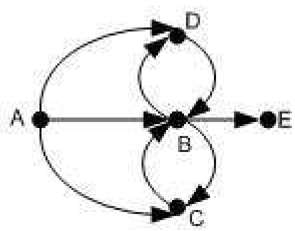

Figure 1. A contribution graph for set F

Two vertices $A, B \in S$ are strongly connected, if and only if there exists in graph G a path from A to B and backwards, from B to A. It is obvious that the relation of strong connectivity is an equivalence relation. So, there is a partition of set of vertices S into pairwise disjoint subsets. That is, $S = \bigcup_{i=1}^{n} S_i$ and all vertices in S_i, $i = \overline{1, n}$, are strongly connected, and every two vertices from different subsets are not strongly connected.

In accordance with this partition, subgraphs $G_i = (S_i, E_i)$, $i = \overline{1,n}$ are called strongly connected components [4] of the graph G, where E_i represents the set of edges that connect pairs of vertices in S_i.

Example 2. *The set of vertices of the graph represented in Figure 1 are split into three equivalence classes $S_1 = \{A\}$, $S_2 = \{B, C, D\}$ and $S_3 = \{E\}$.*

Definition 2. *Let G^* be the condensed graph of the graph G. Set of vertices of graph G^* represents set $\{G_1, ..., G_n\}$ of all strongly connected components of graph G and there is an edge from vertex G_i to vertex G_j of graph G^*, if there exists in G at least one edge that connects one vertex from component G_i to one vertex from component G_j.*

Obviously the graph G^* is an acyclic one.

Example 3. *The condensed graph of graph from Figure 1 has three vertices and two edges, as shown in Figure 2.*

Figure 2. Condensed graph of the graph from Figure 1

Over the set of vertices of graph G^* a strict partial order is defined. Vertex G_i precedes vertex G_j, if G_j is accessible from G_i. Now, the equivalence classes $S_1, ..., S_n$ will be sorted based on the corresponding order graph's G^* vertices.

Lemma 1. *If $X \to Y \in F^+$ and X is a determinant of set Y under F, then for every attribute $A \in (X - Y)$ there is an attribute $B \in Y$ so that in the contribution graph G there exists a path from vertex A to vertex B and for every attribute $B \in (Y - X)$ there exists in X an attribute A, from which the vertex B can be reached.*

Proof. Let attribute $B \in (Y - X)$ and let the subset X' of set X be determinant for B under F. Because $X' \to B \in F^+$, according to

Claim 1, there is a derivation $H = < X_0', X_1', ..., X_m' >$ for dependency $X' \to B$ under F. Then, based on Claim 2, there exists a sequence of dependencies $V_1 \to W_1, ..., V_q \to W_q$ in F, where $A \in V_1$, $B \in W_q$ and $W_i \bigcap V_{i-1} \neq \emptyset$, for $i = \overline{1, q-1}$.

Contribution graph has a structure, such that for every dependency $V_j \to W_j$ in F, from each vertex labeled with an attribute in V_j an edge leaves to every vertex labeled with an attribute in W_j. So, there exists a path from every vertex $A \in X'$ to vertex B.

It must be mentioned that, if $\overline{X}$ is considered the union of all determinants of attributes in $Y - X$, then $\overline{X} \bigcup X \bigcap Y = X$. Indeed, if we suppose that the set $\overline{X} \bigcup X \bigcap Y$ is a proper subset of set X, this will contradict the supposition that X is a determinant for Y under F.

Corollary 1. *If reduced dependency $V \to W$ is used nonredundantly in building the derivation H for dependency $X \to Y$ under F, then in contribution graph G there exists a path from every vertex labeled with an attribute in V to every vertex labeled with an attribute in Y.*

4 Properties of equivalence classes of attributes

Theorem 1. *If X is a determinant under F of set $S_1 \bigcup ... \bigcup S_j$, where $j = \overline{1, n}$, then $X \subseteq S_1 \bigcup ... \bigcup S_j$.*

Proof. Let $X \not\subset S_1 \bigcup ... \bigcup S_j$. Then there exists an equivalence class S_t, where $t = \overline{j, n}$, such that $X \bigcap S_t \neq \emptyset$. By Lemma 1, in the contribution graph G, from every attribute $A \in X \bigcap S_t$ there is a path towards B, where $B \in S_1 \bigcup ... \bigcup S_j$. But this fact contradicts the supposition that the sets $S_1, ..., S_j$ precede the set S_t.

Corollary 2. *If X is a determinant of set $S_1 \bigcup ... \bigcup S_n$ under F, then $X \bigcap S_1 \neq \emptyset$.*

Proof. Indeed, for every attribute B in S_1 or $B \in X$, or, according to Lemma 1, there is in X an attribute A from which vertex B is accessible in contribution graph G. But then A is also a member of equivalence class S_1.

Definition 3. *Equivalence class S_j is called nonredundant, if and only if for every attribute A in S_j, the expression $(\bigcup_{i=1}^{n} S_i - S_j) \rightarrow A \notin F^+$ holds.*

Considering Lemma 1, it can be concluded that set S_j is nonredundant, if and only if for every attribute A in S_j, the expression $(\bigcup_{i=1}^{j-1} S_i) \rightarrow A \notin F^+$ holds.

From the ordered sequence of sets $S_1, ..., S_n$ a sequence of ordered nonredundant sets can be built $T_1, ..., T_n$, where $T_1 = S_1$ and $T_j = S_j - (\bigcup_{i=1}^{j-1} T_i)_F^+$ for $j = \overline{2, n}$. As a result of this process, some sets T_j can become empty. These empty sets can be excluded from the sequence and a sequence of nonempty sets $T_1, ..., T_m$ will be obtained, keeping the precedence of prior sets.

Proposition 1. $T_1 = S_1$.

Proposition 2. $(T_1 \bigcup ... \bigcup T_m) \rightarrow (S_1 \bigcup ... \bigcup S_n) \in F^+$.

Example 4. *Sequence of equivalence classes of attributes $S_1 = \{A\}$, $S_2 = \{B, C, D\}$ and $S_3 = \{E\}$ turns into the following sequence of non redundant equivalence classes of attributes: $T_1 = \{A\}$, $T_2 = \{B, C, D\}$.*

Theorem 2. *Set X is a determinant of set $S_1 \bigcup ... \bigcup S_n$ under F, if and only if X is determinant of set $T_1 \bigcup ... \bigcup T_m$ under F.*

Proof. Necessity. Because X is a determinant of set $S_1 \bigcup ... \bigcup S_n$ and $T_1 \bigcup ... \bigcup T_m \subseteq S_1 \bigcup ... \bigcup S_n$, then $X \rightarrow (T_1 \bigcup ... \bigcup T_m) \in F^+$. Supposing X is not a determinant of set $T_1 \bigcup ... \bigcup T_m$ under F, thus there exists at least one attribute A in X for which the expression $(X - \{A\}) \rightarrow (T_1 \bigcup ... \bigcup T_m) \in F^+$ holds. Then, according to Proposition 2, the expression $(X - \{A\}) \rightarrow (S_1 \bigcup ... \bigcup S_n) \in F^+$ holds, fact that contradicts the hypothesis that X is a determinant of set $S_1 \bigcup ... \bigcup S_n$ under F.

Sufficiency. Let X be a determinant of set $T_1 \bigcup ... \bigcup T_m$ under F. Since $(T_1 \bigcup ... \bigcup T_m) \rightarrow (S_1 \bigcup ... \bigcup S_n) \in F^+$ and $T_1 \bigcup ... \bigcup T_m \subseteq S_1 \bigcup ... \bigcup S_n$, then X is a determinant for $S_1 \bigcup ... \bigcup S_n$ under F.

Lemma 2. *If X is a determinant under F of set $S_1 \bigcup ... \bigcup S_n$, then Z, where $Z = X \bigcap (S_1 \bigcup ... \bigcup S_j)$ and $j = \overline{1, n}$, is a determinant for $S_1 \bigcup ... \bigcup S_j$ under F.*

Proof. According to Theorem 1, the expression $X \subseteq S_1 \bigcup ... \bigcup S_n$ takes place. First it will be shown that $Z \to (S_1 \bigcup ... \bigcup S_j) \in F^+$. Lets suppose the contrary: $Z \to (S_1 \bigcup ... \bigcup S_j) \notin F^+$. Then there exists a set Z', where $Z' \subseteq X$, which is a determinant of set $S_1 \bigcup ... \bigcup S_j$ and $Z' \bigcap (\bigcup_{i=j+1}^{n} S_i) \neq \emptyset$. Considering Lemma 1, there is a path from every vertex labeled with A in $Z' \bigcap (\bigcup_{i=j+1}^{n} S_i)$ that leads to a vertex B in $\bigcup_{i=1}^{j} S_i$. A contradiction has been encountered. Therefore, $Z \to (S_1 \bigcup ... \bigcup S_j) \in F^+$.

To complete the proof of this lemma, it will be shown that Z is a determinant under F of set $S_1 \bigcup ... \bigcup S_j$. Indeed, if it is considered that Z is not a determinant of F under F, then there must exist in Z an attribute A, such that $(Z - \{A\}) \to (S_1 \bigcup ... \bigcup S_j) \in F^+$. But then $(Z - \{A\}) \to Z \in F^+$ takes place, fact that implies $(X - \{A\}) \to X \in F^+$. So, a contradiction has been encountered, that X is a determinant of set $S_1 \bigcup ... \bigcup S_n$ under X.

Theorem 3. *If set $Z = X \bigcap (T_1 \bigcup ... \bigcup T_j)$of attributes is a determinant of set $S_1 \bigcup ... \bigcup S_n$, then $X \subseteq T_1 \bigcup ... \bigcup T_m$.*

Proof. Let S_j be the first set of attributes that doesn't coincide with T_j and assume that there is an attribute A in X, such that $A \in S_j$ and $A \notin T_j$. Lemma 2 implies that $(X \bigcap (S_1 \bigcup ... \bigcup S_j)) \to (S_1 \bigcup ... \bigcup S_j) \in F^+$. Since $A \notin T_j$, then $(X \bigcap (S_1 \bigcup ... \bigcup S_j)) \to A \in F^+$. So $(X - \{A\}) \to X \in F^+$, thus X is not a determinant of set $S_1 \bigcup ... \bigcup S_n$ under F.

Corollary 3. *If an attribute A in R is prime in scheme M, then $A \in \bigcup_{i=1}^{m} T_i$.*

Corollary 4. *If an attribute A in $O(||F||)$ is nonprime in scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$, then $A \in (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$.*

Example 5. *Considering Corollaries 3 and 4, and Example 4, for the scheme $Sch(R, F)$, where $F = \{C \to B, AD \to B, AB \to DC, B \to E\}$*

and $R = \{A, B, C, D, E\}$*,* $\{A, B, C, D\}$ *is set of prime attributes and* E *is nonprime attribute.*

Theorem 3 and Lemma 2 can be paraphrased for nonredundant equivalence classes of attributes.

Lemma 3. *If* X *is a determinant under* F *of set* $T_1 \bigcup ... \bigcup T_m$*, then* Z*, where* $Z = X \bigcap (T_1 \bigcup ... \bigcup T_j)$ *and* $j = \overline{1, m}$*, is a determinant for* $T_1 \bigcup ... \bigcup T_j$ *under* F*.*

Proposition 3. *If set of attributes* X *is a determinant of set* $T_1 \bigcup ... \bigcup \bigcup T_j$*, then* $X \subseteq T_1 \bigcup ... \bigcup T_j$*, where* $j = \overline{1, m}$*.*

The soundness of this affirmation follows from theorems 1, 2 and 3.

Theorem 4. *If set of attributes* X *is a determinant of set* $T_1 \bigcup ... \bigcup T_m$*, then* $X \bigcap T_i \neq \emptyset$*, where* $i = \overline{1, m}$*.*

Proof. Let for a set T_j, where $j = \overline{1, m}$, the equality $X \bigcap T_j = \emptyset$ holds. From Corollary 2 and Proposition 1, follows that $X \bigcap T_1 \neq \emptyset$. According to Lemma 3 set Z, where $Z = X \bigcap (T_1 \bigcup ... \bigcup T_j)$ and $j = \overline{1, m}$, is a determinant for $T_1 \bigcup ... \bigcup T_j$ under F. From the fact that $X \bigcap T_j = \emptyset$ it follows that $Z \subseteq T_1 \bigcup ... \bigcup T_{j-1}$ and then $(T_1 \bigcup ... \bigcup T_{j-1}) \rightarrow T_j \in F^+$. But this contradicts the assumption that set $T_1 \bigcup ... \bigcup T_m$ is nonredundant.

5 Algorithmic aspects

From the algorithmic point of view, the problem of testing the primeness of attributes consists of two parts, construction of equivalence classes of scheme's attributes and elimination of the redundancy in these classes. In other words, being given a relation scheme $Sch(R, F)$, the sets $S_1 \bigcup ... \bigcup S_n = R$ and $T_1 \cup ... \cup T_m$ are to be build, respectively.

The method for determination of equivalence classes of attributes consists in the fact that for every attribute A in R, the list of attributes that label accessible vertices from A on the contribution graph is computed. So, accessibility matrix M will be computed, that will consist of 0 and 1, with a dimension $|R| \times |R|$, where $|R|$ is cardinality of set R. The element $M(i, j) = 1$ if and only if there exists a path from vertex i to vertex j. Based on matrix M the set of equivalence classes of attributes R is constructed.

In the speciality literature (for example, in [5]) it is described an algorithm of finding the strongly connected components of a directed graph with a complexity $O(\max(|S|, |E|))$, where $|S|$- number of vertices, and $|E|$- number of edges. But, it is easy to observe that, using this algorithm is non suitable, because the computing of the contribution graph (for example its representation in form of adjacency lists) for a set F of functional dependencies requires $O(|R| \cdot ||F||)$ operations and the graph will have a number of edges proportionally to $|R|^2$. Where $||F||$ is the number of attributes involved in F, when duplicates are also considered. As $||F|| > |R|$, algorithm of computing the equivalence classes of attributes needs $O(|R| \cdot ||F||)$ operations.

Because the closure of a set of attributes under a set of functional dependencies is computed in a time $O(||F||)$ [2], then for equivalence classes of attributes the elimination of redundancies requires $O(|EquivClasses| \cdot ||F||)$, where $|EquivClasses|$ represents the number of equivalence classes of attributes. Since $|EquivClasses| \leq |R|$, this algorithm requires a time proportionally to $|R| \cdot ||F||$.

References

[1] Yu C.T., Johnson D.T. *On the complexity of finding the set of candidate keys for a given set of functional dependencies.* Information Processing Letters, 1978, V.5, N.4, p.100-101.

[2] Maier, D. *The theory of relational database.* Computer Science Press, 1983, 637 p.

[3] Cotelea, Vitalie. *Relational databases: logical design.* ASEM, Chisinau, 1997, 290 p. (in Romanian)

[4] Even, Shimon. *Graph Algorithms.* Computer Science press, 1979, 250 p.

[5] Aho, Alfred V. Hopcroft, John E. and Ullman, Jeffrey D. *The Design and Analysis of Computer Algorithms.* Addison-Wesley, placeCityReading, StateMA, 1974, 470 p.

Vitalie Cotelea Received April 28, 2009

Vitalie Cotelea
Academy of Economic Studies of Moldova
Phone: (+373 22) 40 28 87
E–mail: *vitalie.cotelea@gmail.com*

Ramanujan-like formulas for $\frac{1}{\pi^2}$ á la Guillera and Zudilin and Calabi-Yau differential equations

Gert Almkvist

Abstract

Using the PSLQ-algorithm J.Guillera found some formulas for $\frac{1}{\pi^2}$. He proved three of them using WZ-pairs. Then W. Zudilin showed how to produce formulas for $\frac{1}{\pi^2}$ by squaring formulas for $\frac{1}{\pi}$. The success of this depends on facts related to Calabi-Yau differential equations of string theory. Here some examples of this is worked out. Also some formulas containing harmonic numbers are found by differentiating formulas for $\frac{1}{\pi^2}$.

1 Introduction

Ramanujan [10] found several formulas for $\frac{1}{\pi}$ of the following form

$$\sum_{n=0}^{\infty} a_n x_0^n(\alpha + \beta n) = \frac{1}{\pi}$$

where

$$v(x) = \sum_{n=0}^{\infty} a_n x^n$$

satisfies a third order differential equation with polynomial coefficients. J.Guillera [9] found eight (and proved three of them) formulas for $\frac{1}{\pi^2}$ of the form

$$\sum_{n=0}^{\infty} A_n x_0^n(c_0 + c_1 n + c_2 n^2) = \frac{1}{\pi^2}$$

where

$$w(x)=\sum_{n=0}^{\infty}A_nx^n$$

satisfies a differential equation of order five. It is quite remarkable that

$$w(x)=x(y_0(x)y_1'(x)-y_0'(x)y_1(x))$$

where y_0, y_1 are solutions of a fourth order differential equation (the pullback) of Calabi-Yau type (see [1] for definitions). With the notation of [2] the w are $\widehat{3},\widehat{6},\widehat{7},\widehat{8},\widehat{11},\widehat{12},A*\beta=\#40,C*\vartheta$. Guillera used the PSLQ-algorithm to find and WZ-pairs to prove his formulas. Also this paper uses modern computer algebra to find the formulas.

In [14] Zudilin showed how to "square" a Ramanujan-like formula for $\frac{1}{\pi}$ to get a formula for $\frac{1}{\pi^2}$. The success of this depends on the fact that $v(x)=u(x)^2$ where $u(x)$ satisfies a second order differential equation. Hence $w(x)=u(x)^4$ which leads to that the Yukawa coupling of the pullback is trivial. This is proved in section 2. In section 1 we give some examples of Zudilin's square. Finally in section 3 we give some examples of formulas containing harmonic numbers obtained by differentiating the formulas for $\frac{1}{\pi^2}$.

2 The square of Ramanujan

In [14] Zudilin has given the recipe for how to obtain a formula for $\frac{1}{\pi^2}$ from a formula for $\frac{1}{\pi}$. The key fact is that for all known formulas

$$\sum_{n=0}^{\infty}B_nx_0^n(\alpha+\beta n)=\frac{1}{\pi}$$

then

$$v=\sum_{n=0}^{\infty}B_nx^n$$

satisfies a third order differential equation

$$v'''+s_2v''+s_1v'+s_0v=0$$

which is the symmetric square of a second order differential equation

$$u'' + p_1 u' + p_0 u = 0.$$

This means that $v = u^2$ and in [1] it is shown that it is equivalent to

$$\frac{s_1 s_2}{27} + \frac{s_1'}{2} - \frac{s_2''}{6} - \frac{s_2 s_2'}{3} - s_0 = 0$$

and

$$p_0 = \frac{s_1}{4} - \frac{s_2^2}{18} - \frac{s_2'}{12} \ ,$$

$$p_1 = \frac{s_2}{3} \ .$$

In Zudilin [14] it is shown that squaring the formula for $\frac{1}{\pi}$ one obtains the following formula for $\frac{1}{\pi^2}$:

$$\sum_{n=0}^{\infty} A_n x_0^n (c_0 + c_1 n + c_2 n^2) = \frac{1}{\pi^2}$$

where

$$v^2 = \sum_{n=0}^{\infty} A_n x^n$$

and

$$c_0 = \alpha^2 + \frac{4}{3}\beta^2 x_0^2 p_0(x_0) = \alpha^2 + \frac{1}{27}\beta^2 x_0^2 (9 s_1(x_0) - 2 s_2(x_0)^2 - 3 s_2'(x_0)) \ ,$$

$$c_1 = \alpha\beta + \frac{1}{3}\beta^2 (x_0 p_1(x_0) - 1) = \alpha\beta + \frac{1}{9}\beta^2 (x_0 s_2(x_0) - 3) \ ,$$

$$c_2 = \frac{1}{3}\beta^2 \ .$$

The hypergeometric case.
Assume that

$$v = \sum_{n=0}^{\infty} \frac{(\frac{1}{2})_n (a)_n (1-a)_n}{n!^3} x^n.$$

Then $v = u^2$ where

$$u = \sum_{n=0}^{\infty} \frac{(\frac{a}{2})_n(\frac{1-a}{2})_n}{n!^2} x^n$$

by Clausen's identity and u satisfies

$$u'' + \frac{2-3x}{2x(1-x)}u' - \frac{a(1-a)}{4x(1-x)}u = 0$$

which gives

$$c_0 = \alpha^2 - \frac{1}{3}\beta^2 a(1-a)\frac{x_0}{1-x_0} \ ,$$

$$c_1 = \alpha\beta - \frac{1}{6}\beta^2 \frac{x_0}{1-x_0} \ ,$$

$$c_2 = \frac{1}{3}\beta^2 \ .$$

Case a=1/2.
Here

$$v = \sum_{n=0}^{\infty} \frac{(\frac{1}{2})_n^3}{n!^3} x^n$$

and

$$A_n = \sum_{k=0}^{n} \frac{(\frac{1}{2})_k^3(\frac{1}{2})_{n-k}^3}{k!^3(n-k)!^3}$$

with

$$w = \sum_{n=0}^{\infty} A_n x^n$$

satisfying the 5-th order differential equation ($\theta = x\frac{d}{dx}$)

$$8\theta^5 - x(2\theta+1)(8\theta^4 + 16\theta^3 + 17\theta^2 + 9\theta + 2) + 8x^2(\theta+1)^5 \ .$$

Example (Ramanujan [10])

$$\sum_{n=0}^{\infty} \frac{(\frac{1}{2})_n^3}{n!^3}(5+42n)\frac{1}{64^n} = \frac{64}{\pi}$$

Here $x_0 = \frac{1}{64}$, $\alpha = \frac{5}{16}$, $\beta = \frac{42}{16}$ which gives $c_0 = \frac{17}{192}$, $c_1 = \frac{77}{96}$, $c_2 = \frac{147}{64}$ and we find

$$\sum_{n=0}^{\infty} A_n \frac{17 + 154n + 441n^2}{192} \frac{1}{64^n} = \frac{1}{\pi^2} \ .$$

Case a=1/3.

Here

$$v = \sum_{n=0}^{\infty} \frac{(\frac{1}{2})_n(\frac{1}{3})_n(\frac{2}{3})_n}{n!^3} x^n$$

and

$$A_n = 108^{-n} \sum_{k=0}^{n} \binom{2k}{k}^2 \binom{2n-2k}{n-k}^2 \binom{3k}{k} \binom{3n-3k}{n-k}$$

with

$$w = \sum_{n=0}^{\infty} A_n x^n$$

satisfying the 5-th order differential equation

$$324\theta^5 - 18x(2\theta+1)(18\theta^4 + 36\theta^3 + 37\theta^2 + 19\theta + 4) +$$
$$+ x^2(\theta+1)(3\theta+2)(3\theta+4)(6\theta+5)(6\theta+7) \ .$$

Example (Chan-Liaw-Tan [7])

$$\sum_{n=0}^{\infty} \frac{(\frac{1}{2})_n(\frac{1}{3})_n(\frac{2}{3})_n}{n!^3} \frac{(-1)^n}{500^{2n}} (827 + 14151n) = \frac{1500\sqrt{3}}{\pi}$$

Here $x_0 = -\frac{1}{500^2}$, $\alpha = \frac{827}{1500\sqrt{3}}$, $\beta = \frac{14151}{1500\sqrt{3}}$, which gives $c_0 = \frac{410393}{4050000}$, $c_1 = \frac{2600669}{1500000}$, $c_2 = \frac{22250089}{2250000}$ and we find

$$\sum_{n=0}^{\infty} A_n \frac{4103930 + 70218063n + 400501602n^2}{40500000} \frac{(-1)^n}{500^{2n}} = \frac{1}{\pi^2} \ .$$

Case a=1/4.

Here

$$v=\sum_{n=0}^{\infty}\frac{(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{n!^3}x^n$$

and

$$A_n=256^{-n}\sum_{k=0}^{n}\frac{(4k)!}{k!^4}\frac{(4n-4k)!}{(n-k)!^4}$$

with

$$w:=\sum_{n=0}^{\infty}A_nx^n$$

satisfying the differential equation

$$64\theta^5-2x(2\theta+1)(32\theta^4+64\theta^3+63\theta^2+31\theta+6)+$$
$$+x^2(\theta+1)(2\theta+1)(2\theta+3)(4\theta+3)(4\theta+5)\ .$$

Example. (J.Borwein-P.Borwein [4])

$$\sum_{n=0}^{\infty}\frac{(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{n!^3}\frac{(-1)^n}{882^{2n}}(1123+21460n)=\frac{3528}{\pi}$$

Here $x_0=-\frac{1}{882^2}$, $\alpha=\frac{1123}{3528}$, $\beta=\frac{21460}{3528}$ which gives $c_0=\frac{630583}{6223392}$, $c_1=\frac{18074759}{9335088}$, $c_2=\frac{28783225}{2333772}$ and we find

$$\sum_{n=0}^{\infty}A_n\frac{1891749+36149518n+230265800n^2}{18670176}\frac{(-1)^n}{882^{2n}}=\frac{1}{\pi^2}\ .$$

Example. (D.V.Chudnovsky-G.V.Chudnovsky [8])

$$\sum_{n=0}^{\infty}\frac{(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{n!^3}\frac{1}{7^{4n}}(3+40n)=\frac{49\sqrt{3}}{9\pi}$$

Here $x_0 = \frac{1}{7^4}$, $\alpha = \frac{27}{49\sqrt{3}}$, $\beta = \frac{360}{49\sqrt{3}}$, which gives $c_0 = \frac{1935}{19208}$, $c_1 = \frac{3237}{2401}$, $c_2 = \frac{14400}{2401}$ and we find

$$\sum_{n=0}^{\infty} A_n \frac{1935 + 25896n + 115200n^2}{19208} \frac{1}{7^{4n}} = \frac{1}{\pi^2} \ .$$

Case a=1/6.
Here

$$v = \sum_{n=0}^{\infty} \frac{(\frac{1}{2})_n (\frac{1}{6})_n (\frac{5}{6})_n}{n!^3} x^n$$

and

$$A_n = 1728^{-n} \binom{2n}{n} \binom{3n}{n} \sum_{n=0}^{\infty} 16^{-k} \binom{2k}{k}^3 \binom{2n-2k}{n-k}$$

with

$$w = \sum_{n=0}^{\infty} A_n x^n$$

satisfies

$$648\theta^5 - 9x(2\theta+1)(3\theta+1)(3\theta+2)(8\theta^2+8\theta+5) +$$
$$+8x^2(\theta+1)(3\theta+1)(3\theta+2)(3\theta+4)(3\theta+5)$$

(this is the Hadamard product $B * \vartheta$ which is deleted from the big table since its fourth order pullback has trivial Yukawa coupling)

Example (J.Borwein-P.Borwein [4])

$$\frac{1}{2E\sqrt{3}} \sum_{n=0}^{\infty} \frac{(\frac{1}{2})_n (\frac{1}{6})_n (\frac{5}{6})_n}{n!^3} (A + Bn) \frac{1}{E^{2n}} = \frac{1}{\pi}$$

where

$$A = 1657145277365 + 212175710912\sqrt{61} \ ,$$
$$B = 107578229802750 + 13773980892672\sqrt{61} \ ,$$
$$E = 4752926464 + 608549875\sqrt{61} \ .$$

We obtain

$$\sum_{n=0}^{\infty} A_n(c_0 + c_1 n + c_2 n^2)\frac{1}{E^{2n}} = \frac{1}{\pi^2}$$

where

$$c_0 = \frac{1}{1320^3 E^2 F^3}(111664689387605862532927043117398929778009833 4$$
$$+142971984278650150521031407984764718461880160\sqrt{61}) \ ,$$
$$c_1 = \frac{1}{1320^3 E^2 F^3}(7249026250027431080646010302794456456357868150 3$$
$$+928142703605173341663106184983443065374812048 0\sqrt{61}) \ ,$$
$$c_2 = \frac{1}{1320^3 E^2 F^3}\times$$
$$\times(15686361809859452157973642156623168538259819499 03$$
$$+20084328236329394569722857360962957949842982400 0\sqrt{61})$$

where

$$F = 236674 + 30303\sqrt{61} \ .$$

Sporadic formulas

Example. (H.H.Chan-S.H.Chan-Z.G.Liu [6])

$$\sum_{n=0}^{\infty}\sum_{k=0}^{n}\binom{n}{k}^2\binom{2k}{k}\binom{2n-2k}{n-k}(1+5n)\frac{(-1)^n}{64^n} = \frac{8}{\pi\sqrt{3}} \ .$$

Here (case (α))

$$v = \sum_{n=0}^{\infty}\sum_{k=0}^{n}\binom{n}{k}^2\binom{2k}{k}\binom{2n-2k}{n-k}x^n$$

satisfies

$$v''' + \frac{3(128x^2-30x+1)}{x(64x^2-20x+1)}v'' + \frac{448x^2-68x+1}{x^2(64x^2-20x+1)}v' + \frac{4}{x^2(4x-1)}v = 0 \ .$$

We get

$$A_n = \sum_{k=0}^{n} \sum_i \sum_j \binom{k}{i}^2 \binom{2i}{i} \binom{2k-2i}{k-i} \binom{n-k}{j}^2 \binom{2j}{j} \binom{2n-2k-2j}{n-k-j}$$

and

$$w = \sum_{n=0}^{\infty} A_n x^n$$

satisfies

$$\theta^5 - 2x(2\theta+1)(10\theta^4 + 20\theta^3 + 25\theta^2 + 15\theta + 4)$$
$$+2^2 x^2(\theta+1)(132\theta^4 + 528\theta^3 + 947\theta^2 + 838\theta + 312)$$
$$-2^7 x^3(2\theta+3)(10\theta^4 + 60\theta^3 + 145\theta^2 + 165\theta + 74)$$
$$+2^{12} x^4 (\theta+2)^5 \ .$$

Then we have $x_0 = -\frac{1}{64}$, $\alpha = \frac{\sqrt{3}}{8}$, $\beta = \frac{5\sqrt{3}}{8}$ which gives $c_0 = -\frac{1}{72}$, $c_1 = \frac{5}{32}$, $c_2 = \frac{25}{64}$ and

$$\sum_{n=0}^{\infty} A_n \frac{-8 + 90n + 225n^2}{576} \frac{(-1)^n}{64^n} = \frac{1}{\pi^2} \ .$$

Example.

$$\sum_{n=0}^{\infty} \sum_{k=0}^{n} \binom{2k}{k}^2 \binom{2n-2k}{n-k}^2 \frac{1 + (3+2\sqrt{3})n}{4} \left(\frac{2-\sqrt{3}}{64} \right)^n = \frac{1}{\pi}$$

Here (case (β))

$$v = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \binom{2k}{k}^2 \binom{2n-2k}{n-k}^2 x^n$$

satisfies

$$v''' + \frac{3(32x-1)}{x(16x-1)} v'' + \frac{1792x^2 - 112x + 1}{x^2(16x-1)^2} v' + \frac{8(32x-1)}{x^2(16x-1)^2} v = 0 \ .$$

We get

$$A_n = \sum_{k=0}^{n} \sum_i \sum_j \binom{2i}{i}^2 \binom{2k-2i}{k-i}^2 \binom{2j}{j}^2 \binom{2n-2k-2j}{n-k-j}$$

where

$$w = \sum_{n=0}^{\infty} A_n x^n$$

satisfies

$$\begin{gathered}\theta^5 - 2^3 x(2\theta+1)(4\theta^4 + 8\theta^3 + 11\theta^2 + 7\theta + 2) \\ +2^9 x^2(\theta+1)(3\theta^4 + 12\theta^3 + 23\theta^2 + 22\theta + 9) \\ -2^{11} x^3(2\theta+3)(4\theta^4 + 24\theta^3 + 59\theta^2 + 69\theta + 32) \\ +2^{16} x^4(\theta+2)^5 \ .\end{gathered}$$

Then we have $x_0 = \frac{2-\sqrt{3}}{64}$, $\alpha = \frac{1}{4}$, $\beta = \frac{3+2\sqrt{3}}{4}$ which gives $c_0 = 0$, $c_1 = \frac{1+\sqrt{3}}{8}$, $c_2 = \frac{7+4\sqrt{3}}{16}$ and we get

$$\sum_{n=0}^{\infty} A_n \frac{(2+2\sqrt{3})n + (7+4\sqrt{3})n^2}{16} \left(\frac{2-\sqrt{3}}{64}\right)^n = \frac{1}{\pi^2} \ .$$

Example (T.Sato [12])

$$\sum_{n=0}^{\infty} \sum_k \binom{n}{k}^2 \binom{n+k}{n}^2 (10-3\sqrt{5}+20n) \left(\frac{\sqrt{5}-1}{2}\right)^{12n} = \frac{20\sqrt{3}+9\sqrt{15}}{6\pi}$$

Here (case (γ))

$$v = \sum_{n=0}^{\infty} \sum_k \binom{n}{k}^2 \binom{n+k}{n}^2 x^n$$

satisfies

$$v''' + \frac{3(2x^2-51x+1)}{x(x^2-34x+1)} v'' + \frac{7x^2-112x+1}{x^2(x^2-34x+1)} v' + \frac{x-5}{x^2(x^2-34x+1)} v = 0 \, .$$

We get

$$A_n = \sum_{k=0}^{n} \sum_i \sum_j \binom{k}{i}^2 \binom{k+i}{k}^2 \binom{n-k}{j}^2 \binom{n-k+j}{j}^2$$

and

$$w = \sum_{n=0}^{\infty} A_n x^n$$

satisfies

$$\theta^5 - 2x(2\theta+1)(17\theta^4 + 34\theta^3 + 38\theta^2 + 21\theta + 5)$$

$$+2x^2(\theta+1)(579\theta^4 + 2316\theta^3 + 3604\theta^2 + 2576\theta + 714)$$

$$-2x^3(2\theta+3)(17\theta^4 + 102\theta^3 + 242\theta^2 + 267\theta + 115)$$

$$+x^4(\theta+2)^5 \ .$$

We have $x_0 = \left(\frac{\sqrt{5}-1}{2}\right)^{12}$, $\alpha = \frac{6(10-3\sqrt{5})}{20\sqrt{3}+9\sqrt{15}}$, $\beta = \frac{120}{20\sqrt{3}+9\sqrt{15}}$ which gives

$$c_0 = \frac{1473122}{9} - 73200\sqrt{5} \ ,$$

$$c_1 = 183680 - 82144\sqrt{5} \ ,$$

$$c_2 = 51520 - 23040\sqrt{5}$$

and

$$\sum_{n=0}^{\infty} A_n(c_0 + c_1 n + c_2 n^2) \left(\frac{\sqrt{5}-1}{2}\right)^{12n} = \frac{1}{\pi^2}.$$

Example (H.H.Chan-H.Verrill [5])

$$\sum_{n=0}^{\infty} \sum_k (-1)^{n+k} 3^{n-3k} \binom{n}{3k} \binom{n+k}{n} \frac{(3k)!}{k!^3} (1+4n) \frac{1}{81^n} = \frac{3\sqrt{3}}{2\pi}$$

Here (case (δ))

$$v=\sum_{n=0}^{\infty}\sum_{k}(-1)^{n+k}3^{n-3k}\binom{n}{3k}\binom{n+k}{n}\frac{(3k)!}{k!^3}x^n$$

satisfies

$$v'''+\frac{3(162x^2+21x+1)}{x(81x^2+14x+1)}v''+\frac{567x^2+48x+1}{x^2(81x^2+14x+1)}v'+$$

$$+\frac{3(27x+1)}{x^2(81x^2+14x+1)}v=0\ .$$

We get

$$A_n=\sum_{k=0}^{n}\sum_{i}\sum_{j}(-1)^{n+i+j}3^{n-3i-3j}\binom{k}{3i}\binom{n-k}{3j}\binom{k+i}{k}\times$$

$$\times\binom{n-k+j}{j}\frac{(3i)!}{i!^3}\frac{(3j)!}{j!^3}$$

and

$$w:=\sum_{n=0}^{\infty}A_nx^n$$

satisfies

$$\begin{gathered}\theta^5+2x(2\theta+1)(7\theta^4+14\theta^3+18\theta^2+11\theta+3)\\+2x^2(\theta+1)(179\theta^4+716\theta^3+1364\theta^2+1296\theta+522)\\+2\cdot3^4x^3(2\theta+3)(7\theta^4+42\theta^3+102\theta^2+117\theta+53)\\+3^8x^4(\theta+2)^5\ .\end{gathered}$$

We have $x_0=\frac{1}{81}$, $\alpha=\frac{2}{3\sqrt{3}}$, $\beta=\frac{8}{3\sqrt{3}}$, which gives $c_0=\frac{50}{243}$, $c_1=\frac{160}{243}$, $c_2=\frac{64}{81}$ and we get

$$\sum_{n=0}^{\infty}A_n\frac{50+160n+192n^2}{243}\frac{1}{81^n}=\frac{1}{\pi^2}\ .$$

Example (Yifan Yang [13])

$$\sum_{n=0}^{\infty}\sum_{k}\binom{n}{k}^4(1+4n)\frac{1}{36^n}=\frac{18}{\pi\sqrt{15}}$$

We have

$$v=\sum_{n=0}^{\infty}\sum_{k}\binom{n}{k}^4 x^n$$

which satisfies

$$v'''+\frac{3(128x^2+18x-1)}{x(64x^2+12x-1)}v''+\frac{444x^2+40x-1}{x^2(64x^2+12x-1)}v'+$$

$$+\frac{2(30x+1)}{x^2(64x^2+12x-1)}v=0\ .$$

We have

$$A_n=\sum_{k=0}^{n}\sum_{i}\sum_{j}\binom{k}{i}^4\binom{n-k}{j}^4$$

and

$$w:=\sum_{n=0}^{\infty}A_n x^n$$

satisfies

$$\theta^5-4x(2\theta+1)(\theta^2+\theta+1)(3\theta^2+3\theta+1)$$

$$+16x^2(\theta+1)(\theta^4+4\theta^3-9\theta^2-26\theta-17)$$

$$+8x^3(2\theta+3)(96\theta^4+576\theta^3+1361\theta^2+1491\theta+634)$$

$$+64x^4(\theta+2)(2\theta+3)(2\theta+5)(4\theta+7)(4\theta+9)\ .$$

We have $x_0=\frac{1}{36}$, $\alpha=\frac{\sqrt{15}}{18}$, $\beta=\frac{2\sqrt{15}}{9}$ which gives $c_0=-\frac{1}{60}$, $c_1=\frac{8}{81}$, $c_2=\frac{20}{81}$ and

$$\sum_{n=0}^{\infty}A_n\frac{-27+160n+400n^2}{1620}\frac{1}{36^n}=\frac{1}{\pi^2}$$

Example (Rogers [11])

$$\sum_{n=0}^{\infty}(-1)^n\binom{2n}{n}\sum_k\binom{n}{k}^2\binom{2k}{k}(159-48\sqrt{3}+520n)(\frac{8-5\sqrt{3}}{22})^{2n}=$$

$$=\frac{2(64+29\sqrt{3})}{\pi}$$

We have

$$v=\sum_{n=0}^{\infty}\binom{2n}{n}\sum_k\binom{n}{k}^2\binom{2k}{k}x^n$$

which satisfies

$$v'''+\frac{3(1-60x+288x^2)}{x(1-40x+144x^2)}v''+\frac{11-132x+972x^2}{x^2(1-40x+144x^2)}v'+$$

$$+\frac{6(-1+18x)}{x^2(1-40x+144x^2)}v=0\ .$$

We have

$$A_n=\sum_{k=0}^{\infty}\binom{2k}{k}\binom{2n-2k}{n-k}\sum_{i,j}\binom{k}{i}^2\binom{2i}{i}\binom{n-k}{j}^2\binom{2j}{j}\ .$$

We have $x_0=-(\frac{8-5\sqrt{3}}{22})^2$, $\alpha=\frac{159-48\sqrt{3}}{2(64+29\sqrt{3})}$, $\beta=\frac{260}{64+29\sqrt{3}}$ which gives $c_0=\frac{1084121-624832\sqrt{3}}{11^4}$, $c_1=\frac{10(581611-333800\sqrt{3}}{3\cdot 11^4}$ $c_2=\frac{100}{3\cdot 11^4}(26476-14848\sqrt{3})$ and

$$\sum_{n=0}^{\infty}A_n(c_0+c_1n+c_2n^2)(-1)^n(\frac{8-5\sqrt{3}}{22})^{2n}=\frac{1}{\pi^2}\ .$$

3 Symmetric squares of third order differential equations

For a fourth order differential equation there are six 2×2-wronskians of the four solutions. In general they satisfy a differential equation of order 6. But there is an interesting exception, the Calabi-Yau equations, for which the wronskians satisfy a fifth order equation. Dually, there are six symmetric squares of the three solutions to a third order differential equation. When do these satisfy a fifth order equation? The answer is:

Theorem:
Consider the differential equation

$$v''' + s_2 v'' + s_1 v' + s_0 v = 0 \ .$$

Then $w = v^2$ satisfies a fifth order equation if and only if

$$\frac{1}{3}s_1 s_2 - \frac{2}{27}s_2^3 + \frac{1}{2}s_1' - \frac{1}{6}s_2'' - \frac{1}{3}s_2 s_2' - s_0 = 0 \ .$$

This means that already v is a square and w is a fourth power of a solution to a second order differential equation.

Proof:
Differentiating $w = v^2$ five times and eliminating $vv', vv'', v'v''$ we get

$$w^{(5)} + \frac{10}{3}s_2 w^{(4)} + 5(s_1 + \frac{5}{9}s_2^2 + \frac{1}{3}s_2')w''' +$$
$$+(11s_0 + 2s_1' + s_2'' + \frac{19}{3}s_1 s_2 + \frac{4}{9}s_1 s_2^2 + 2s_2 s_2')w''$$
$$+(7s_0' + s_1'' + 4s_1^2 + \frac{32}{3}s_0 s_2 + \frac{4}{9}s_1 s_2^2 - \frac{1}{3}s_1 s_2' + \frac{7}{3}s_1' s_2)w' +$$
$$+(2s_0'' + 8s_0 s_1 + \frac{8}{9}s_0 s_2^2 - \frac{2}{3}s_0 s_2' + \frac{14}{3}s_0' s_2)w$$
$$= -12v'^2(\frac{1}{3}s_1 s_2 - \frac{2}{27}s_2^3 + \frac{1}{2}s_1' - \frac{1}{6}s_2'' - \frac{1}{3}s_2 s_2' - s_0)$$

The right hand side is zero if and only if v is the square of a solution to a second order equation.

Corollary.

The fifth order equation in the Theorem is Calabi-Yau (but its fourth order pullback has trivial Yukawa coupling)

Proof: The C-Y2 condition for

$$w^{(5)} + b_4 w^{(4)} + b_3 w''' + b_2 w'' + b_1 w' + b_0 w = 0$$

is (see [3])

$$-b_2 + \frac{3}{2}b_3' + \frac{3}{5}b_3 b_4 - b_4'' - \frac{6}{5}b_4 b_4' - \frac{4}{25}b_4^3 = 0$$

and we compute the left hand side

$$-b_2 + \frac{3}{2}b_3' + \frac{3}{5}b_3 b_4 - b_4'' - \frac{6}{5}b_4 b_4' - \frac{4}{25}b_4^3 =$$

$$= 11(\frac{1}{3}s_1 s_2 - \frac{2}{27}s_2^3 + \frac{1}{2}s_1' - \frac{1}{6}s_2'' - \frac{1}{3}s_2 s_2' - s_0) = 0 \ .$$

We have $w_0 = v_0^2$, $w_1 = v_0 v_1$, $w_2 = v_1^2$. To show that the fourth order pullback has trivial Yukawa coupling we use the identities in [3]

$$x^2 f y_0^2 = \begin{vmatrix} w_0 & w_1 \\ w_0' & w_1' \end{vmatrix} = v_0^2 \begin{vmatrix} v_0 & v_1 \\ v_0' & v_1' \end{vmatrix}$$

$$x^2 f y_0 y_1 = \begin{vmatrix} w_0 & w_1 \\ w_0' & w_1' \end{vmatrix} = 2 v_0 v_1 \begin{vmatrix} v_0 & v_1 \\ v_0' & v_1' \end{vmatrix}$$

$$x^2 f y_0 y_2 = v_1^2 \begin{vmatrix} v_0 & v_1 \\ v_0' & v_1' \end{vmatrix}$$

which implies

$$\frac{y_2}{y_0} = \frac{v_1^2}{v_0^2} = \frac{1}{4}\left(\frac{y_1}{y_0}\right)^2 \ .$$

4 Harmonic sums

The expansions on pp.58-59 of [9] lead to, after differentiation, formulas containing harmonic numbers H_n defined by

$$H_n = \sum_{k=1}^{n} \frac{1}{k}$$

and $H_0 = 0$. As a curiosity I mention the asymptotic expansion

$$H_n = \log(n) + \gamma - \sum_{k=1}^{\infty} \frac{B_k}{kn^k}$$

which could be a strange definition of the Bernoulli numbers

$$\sum_{n=0}^{\infty} \{120 + 4(9 + 120n)(H_{4n} - H_n)\} \frac{(4n)!}{n!^4} \frac{1}{2^{8n}7^{4n}} =$$

$$= \frac{49(8\log(2) + 4\log(7))}{\pi\sqrt{3}}$$

$$\sum_{n=0}^{\infty} \{52780 + 4(2206 + 52780n)(H_{4n} - H_n)\} \frac{(4n)!}{n!^4} \frac{1}{2^{8n}99^{4n}} =$$

$$= \frac{99^2(8\log(2) + 4\log(99))}{\pi\sqrt{2}}$$

$$\sum_{n=0}^{\infty} (-1)^n \{51 + (7 + 51n)(3H_{3n} + 2H_{2n} - 5H_n)\} \binom{2n}{n}^2 \binom{3n}{n} \frac{1}{2^{4n}108^n} =$$

$$= \frac{36(6\log(2) + 3\log(3))}{\pi\sqrt{3}}$$

$$\sum_{n=0}^{\infty} (-1)^n \{545140134 + (13591409 + 545140134n)(6H_n - 3H_{3n} - 3H_n)\} \times$$

$$\times \frac{(6n)!}{(3n)!n!^3} \frac{1}{12^{3n}53360^{3n}} = \frac{9 \cdot 53360^2(\log(12) + \log(53360))}{2\pi\sqrt{10005}}$$

$$\sum_{n=0}^{\infty}(-1)^n\left\{45+410n+10(\frac{13}{4}+45n+205n^2)(H_{2n}-H_n)\right\}\times$$
$$\times\binom{2n}{n}^5\frac{1}{2^{20n}}=\frac{640\log(2)}{\pi^2}$$
$$\sum_{n=0}^{\infty}(-1)^n\left\{1+5n+10(\frac{1}{8}+n+\frac{5}{2}n^2)(H_{2n}-H_n)\right\}\binom{2n}{n}^5\frac{1}{2^{12n}}=$$
$$=\frac{12\log(2)}{\pi^2}$$
$$\sum_{n=0}^{\infty}\left\{38+480n+(\frac{15}{8}+38n+240n^2)(8H_{8n}-4H_{4n}+2H_{2n}-6H_n)\right\}\times$$
$$\times\binom{2n}{n}\frac{(8n)!}{(4n)!n!^4}\frac{1}{2^{18n}7^{4n}}=\frac{49(18\log(2)+4\log(7)}{\pi^2\sqrt{7}}$$

$$\sum_{n=0}^{\infty}(-1)^n\left\{693+10836n+6(29+693n+5418n^2)(H_{6n}-H_n)\right\}\times$$
$$\times\frac{(6n)!}{n!^6}\frac{1}{2880^{3n}}=\frac{384\sqrt{5}\log(2880)}{\pi^2}$$

$$\sum_{n=0}^{\infty}\sum_k\sum_i\sum_j\{160+800n+4(-27+160n+400n^2)(H_{n-k}-H_{n-k-j})\}\times$$
$$\times\binom{k}{i}^4\binom{n-k}{j}^4\frac{1}{36^n}=\frac{1620\log(36)}{\pi^2}$$

I have continued the expansions one term longer for a few of Guillera's expansions on p.43 and 46 in [9]

$$\sum_{n=0}^{\infty}\frac{(-1)^n}{1024^{n+x}}\frac{(\frac{1}{2})_{n+x}^5}{(1)_{n+x}^5}(13+180(n+x)+820(n+x)^2)=$$

$$= \frac{128}{\pi^2} - 320x^2 + \frac{4880}{3}\pi^2 x^4 - 114688\zeta(3)x^5 + O(x^6)$$

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{1024^{n+x}} \frac{(1)_{n+x}^5}{(\frac{3}{2})_{n+x}^5}(13 + 180(n + \frac{1}{2} + x) + 820(n + \frac{1}{2} + x)^2) =$$

$$= 256\zeta(3) + \frac{64}{3}\pi^4 x + O(x^2)$$

$$\sum_{n=0}^{\infty} \frac{1}{64^{n+x}} \frac{(\frac{1}{2})_{n+x}^7}{(1)_{n+x}^7}(1 + 14(n + x) + 76(n + x)^2 + 168(n + x)^3) =$$

$$= \frac{32}{\pi^2}(1 - \pi^2 x^2 + \frac{4}{3}\pi^4 x^4 - \frac{257}{45}\pi^6 x^6 + O(x^7))$$

$$\sum_{n=0}^{\infty} \frac{1}{64^{n+x}} \frac{(1)_{n+x}^7}{(\frac{3}{2})_{n+x}^7}(1+14(n+\frac{1}{2}+x)+76(n+\frac{1}{2}+x)^2+168(n+\frac{1}{2}+x)^3) =$$

$$= \frac{1}{2}\pi^4 - 186\zeta(5)x + O(x^2)$$

Errata: In the thesis on p.58 in [9] $\frac{59\sqrt{3}}{49}$ should be $\frac{9\sqrt{3}}{49}$ (on p.33 formula (1.9) is correct)

References

[1] G.Almkvist and W.Zudilin, *Differential equations, mirror maps and zeta values*, Mirror Symmetry V, N.Yui, S.-T. Yau and J.D. Lewis (eds), Proc. of BIRS workshop on Calabi-Yau Varities and Mirror Symmetry (December 6-11, 2003), AMS/IP Stud. Adv. Math. 38, Amer. Math. Soc. & International Press, Providence, RI (2007), pp. 481–515, math.NT/0402386

[2] G.Almkvist, C. van Enckevort, D. van Straten and W. Zudilin, *Tables of Calabi-Yau equations*, math.AG/ 0507430

[3] G. Almkvist, *Calabi-Yau differential equations of degree 2 and 3 and Yifan Yang's pullback*, math.AG/ 0612215

[4] J.M.Borwein and P.B.Borwein, *Pi and the AGM.* New York, Wiley Interscience, 1987.

[5] H.H.Chan and H.Verrill, *The Apéry numbers, the Almkvist-Zudilin numbers and a new series for* $\frac{1}{\pi}$. Preprint 2007

[6] H.H.Chan, S.H.Chan and Z.Liu, *Domb's numbers and Ramanujan-Sato type series for* $\frac{1}{\pi}$, Adv. Math. 186 (2004), pp. 396–410.

[7] H.H.Chan, W.C.Liaw and V.Tan, *Ramanujan's class invariant* λ_n *and a new class of series for* $\frac{1}{\pi}$. J.London Math. Soc. 64 (2001), pp. 93–106.

[8] D.V.Chudnovsky and G.V.Chudnovsky, *Approximations and complex multiplication according to Ramanujan*, in Ramanujan Revisited: Proceedings of the centennial conference, 1987, edited by G.E.Andrews, B.C.Berndt and R.A.Rankin, pp. 375–472. Academic Press, Boston 1987.

[9] J.Guillera, *Series de Ramanujan: Generalizaciones y conjecturas*, Thesis, Zaragoza 2007, available on http://personal.auna.com/jguillera/

[10] S.Ramanujan, *Modular equations and approximations to* π, Quartily J. of Math. 45 (1914), 350-372.

[11] M.D.Rogers, *New* $_5F_4$ *hypergeometric transformations, three-variable Mahler measures, and formulas for* $\frac{1}{\pi}$, math/NT0704-2438

[12] T.Sato, *Apéry numbers and Ramanujan's series for* $\frac{1}{\pi}$, Abstrat of a talk presented at The Annual Meeting of the Mathematical Society of Japan, March 28-31, 2002.

[13] Y.Yang, *On differential equations satisfied by modular forms*, Math. Z. 246 (2004), pp. 1–19.

[14] W.Zudilin, *Quadratic transformations and Guillera's formulae for* $\frac{1}{\pi^2}$, Preprint 2005, http://wain.mi.ras.ru /publications.html

[15] W.Zudilin, *Ramanujan-type formulae for* $\frac{1}{\pi}$ *: A second wind? in Modular forms and duality and string duality*, Fields Institute Communications, AMS, Providence 2008.

Gert Almkvist

Received May 27, 2009

Lund University
Department of Mathematics
P.O. Box 118
S-221 00 Lund, Sweden
E–mail: *gert.almkvist@yahoo.se*