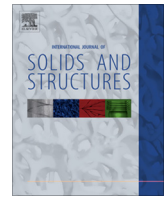




Contents lists available at ScienceDirect

International Journal of Solids and Structures

journal homepage: www.elsevier.com/locate/ijssolstr

Optimization algorithms for the solution of the frictionless normal contact between rough surfaces



A. Bemporad, M. Paggi*

IMT Institute for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca, Italy

ARTICLE INFO

Article history:

Received 1 November 2014
 Received in revised form 1 June 2015
 Available online 16 June 2015

Keywords:

Unilateral contact problem
 Frictionless normal contact
 Quadratic programming
 Optimization algorithms
 Boundary element method
 Roughness

ABSTRACT

This paper revisits the fundamental equations for the solution of the frictionless unilateral normal contact problem between a rough rigid surface and a linear elastic half-plane using the boundary element method (BEM). After recasting the resulting Linear Complementarity Problem (LCP) as a convex quadratic program (QP) with nonnegative constraints, different optimization algorithms are compared for its solution: (i) a Greedy method, based on different solvers for the unconstrained linear system (Conjugate Gradient CG, Gauss–Seidel, Cholesky factorization), (ii) a constrained CG algorithm, (iii) the Alternating Direction Method of Multipliers (ADMM), and (iv) the Non-Negative Least Squares (NNLS) algorithm, possibly warm-started by accelerated gradient projection steps or taking advantage of a loading history. The latter method is two orders of magnitude faster than the Greedy CG method and one order of magnitude faster than the constrained CG algorithm. Finally, we propose another type of warm start based on a refined criterion for the identification of the initial trial contact domain that can be used in conjunction with all the previous optimization algorithms. This method, called cascade multi-resolution (CMR), takes advantage of physical considerations regarding the scaling of the contact predictions by changing the surface resolution. The method is very efficient and accurate when applied to real or numerically generated rough surfaces, provided that their power spectral density function is of power-law type, as in case of self-affine fractal surfaces.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Contact mechanics between rough surfaces is a very active area of research for its paramount importance to address several practical applications in physics and engineering. Understanding the evolution of the contact domain and contact variables, such as load, real contact area, contact stiffness, and many others, that depend on the morphological properties of roughness, is still considered a challenging problem today. The reader is referred to (Barber, 2003; Nosonovsky and Bhushan, 2005; Ciavarella et al., 2006; Hyun and Robbins, 2007; Ciavarella et al., 2008a,b; Carbone and Bottiglione, 2008; Paggi and Ciavarella, 2010; Campana et al., 2001; Paggi and Barber, 2011; Paggi et al., 2014; Yastrebov et al., 2015) for an overview of research results developed during the last decade.

Semi-analytical contact theories that are able to provide synthetic predictions of the contact response is also a challenging topic. A comparison and validation on benchmark results is necessary to understand the limitations of existing approaches and

propose further improvements. Experimental investigations are difficult to make and involve approximations, for example very often the contact parameters can only be estimated by indirect measurements of thermal or electric resistances of compressed rough joints (McCool, 1986; Sridhar and Yovanovich, 1994) or are mostly limited to measurements of real contact area under special conditions (O'Callaghan and Probert, 1970; Hendriks and Visscher, 1995). Therefore, numerical methods are essential to acquire as much information as possible about the contact problem at hand and infer general conclusions.

In spite of its effectiveness and versatility, the finite element method (FEM) has been mainly applied in mechanics to solve contact problems between rough surfaces in which the constitutive behavior of the bulk is not linear elastic. For instance, the study of elasto-plastic contact problems with roughness (Hyun et al., 2004), where an explicit approach was used to reduce the high computational cost, and the study involving frictional dissipative phenomena in visco-elastic materials, where the energy dissipation in the bulk is essential and can be well predicted by FEM (Wriggers and Reinelt, 2009), are worth mentioning.

In the linear elastic regime, when the multi-scale character of roughness covering a wide spectrum of wavelengths is the main

* Corresponding author. Tel.: +39 0583 4326 604; fax: +39 0583 4326 565.
 E-mail address: marco.paggi@imtlucca.it (M. Paggi).

focus, the use of the boundary element method (BEM) is historically preferred over FEM (Andersson, 1981; Man, 1994). This is essentially due to the fact that only the surface must be discretized and not the bulk. Moreover, it is not necessary to adopt surface interpolation techniques, like Bezier curves, to discretize the interface (see, e.g., the rigorous studies in (Wriggers, 2006, Ch. 9) and in (Hyun et al., 2004)), which must be used with care to avoid smoothing out artificially the fine scale geometrical features of roughness.

In the application of BEM, the frictionless contact problem between two linear elastic rough surfaces is mathematically equivalent to the problem of the normal contact between a rigid rough surface and an elastic half-plane with equivalent elastic parameters, see (Barber, 2003) for a rigorous proof. The core of BEM is based on the so-called Green's functions, that relate the displacement of a generic point of the half-plane to the action of a concentrated force on the surface caused by contact interactions. An integral convolution of all the contact tractions provides the deformed contact configuration. After introducing a discretization of the half-plane consisting of a grid of boundary elements, the problem of point-force singularity is solved numerically by using the closed-form solution for a patch load acting on a finite-size boundary element (Johnson, 1985, Ch. 3,4). The contact problem is then set in terms of equalities and inequalities stemming from the unilateral contact constraints and can be solved by constrained optimization. In this regard, apart from the discretization error intrinsic in any numerical method, BEM provides the highest attainable accuracy for discrete problems (Polonsky and Keer, 1999). The basic version of BEM can be also extended to solve rough contact problems with friction (Li and Berger, 2003; Pohrt and Li, 2014) and between viscoelastic materials (Carbone and Putignano, 2013).

With the aim of investigating the effect of roughness at multiple scales, the availability of computational methods that can solve large contact problems in an efficient and fast way is of crucial importance. The size of the linear system of equations relating the contact pressures to the normal deflections can be in fact quite large, as it arises from high resolution profilometric surface samples of 512×512 heights and very large indentations. Hence, the computational challenges regard two main aspects: (i) efficiently solve the system of linear equations; (ii) impose the satisfaction of the unilateral contact constraints (contact inequalities). Regarding the first issue, iterative methods like the Conjugate Gradient algorithm or the Gauss–Seidel method (Francis, 1983; Borri-Brunetto et al., 1999; Borri-Brunetto et al., 2001) have been widely used. Alternatively, the capabilities of multigrid or multi-level methods have been exploited (Raous, 1999; Polonsky and Keer, 1999) to approximately solve the equation system on coarse grids and then project the results on finer grids. Finally, we mention the fast method and its variants based on the solution of the linear system of equations in the Fourier space (see, e.g., Nogi and Kato, 1998; Polonsky and Keer, 2000a,b; Batrouni et al., 2002; Scaraggi et al., 2013; Prodanov et al., 2014).

Regarding the imposition of the contact inequalities, Johnson (1985, p.149–150) suggested to apply a greedy approach: after solving the equation set for the unknown tractions, the boundary elements for which these are negative (tensile) are excluded in a following iteration from the assumed contact area and the corresponding pressures set equal to zero. Johnson (1985, p.149–150) stated that “experience confirms that repeated iterations converge to a set of values of pressures which are positive where contact takes place and zero otherwise”. To the best of the authors' knowledge, a rigorous proof of convergence of this method has not been provided in the literature. However, if valid, it allows to use any numerical method to solve the unconstrained set of linear equations and then impose a correction in a recursive way. Indeed, this

numerical approach has been successfully applied by many authors, such as Kubo et al. (1981) and Borri-Brunetto et al. (1999, 2001) who used this greedy approach in conjunction with a Gauss–Seidel iterative algorithm for the solution of the unconstrained set of linear equations, and Karpenko and Akay (2001) and Batrouni et al. (2002) who applied it together with a numerical scheme based on the Fast Fourier Transform (FFT).

In alternative to the greedy approach, Polonsky and Keer (1999) proposed a constrained Conjugate Gradient method based on the theory in (Hestenes et al., 1980, Ch. 2,3) to solve the linear system of equations and rigorously impose the satisfaction of the contact constraints. For the solution of the system of equations, a multi-grid solution scheme was proposed in (Polonsky and Keer, 1999) and then a FFT algorithm was considered in (Polonsky and Keer, 2000a,b).

In this paper, we first examine the validity of the greedy approach based on a monotonic elimination of tensile points. We show that this approach usually finds the exact solution but, as we prove by a counter-example, it may fail. Then, we show that other optimization algorithms such as Non-Negative Least Squares (NNLS) and the Alternating Direction Method of Multipliers (ADMM) can be used in alternative to the greedy approach, by exploiting the equivalence between the contact problem and quadratic programming with unilateral non-negativity constraints. Moreover, we propose warm starting techniques for the optimization algorithms that are especially useful in case of a solution of a sequence of increasing or decreasing displacements.

This paper provides a comprehensive comparison of the computational performance of the greedy approach (used in conjunction with different unconstrained solvers like the Conjugate Gradient, the Gauss–Seidel iterative scheme, or the MATLAB's `mldivide` solver¹), of the original constrained CG method by Polonsky and Keer (1999), and of novel optimization algorithms that are able to exploit warm starts for solving convex quadratic programs subject to non-negativity constraints. As a main conclusion, the proposed NNLS algorithm with warm start based on accelerated gradient projections (GPs) is found to be one order of magnitude faster than the algorithm by Polonsky and Keer (1999) and two orders of magnitude faster than the greedy approach.

Finally, by exploiting the morphological features of the contact domain of fractal surfaces, we propose a cascade multi-resolution algorithm that can further reduce computation time by at least a factor two with respect to the NNLS algorithm with accelerated GPs.

2. Mathematical formulation

In the framework of BEM, the normal displacements $u(\mathbf{x})$ at any point of the half-plane identified by the position vector \mathbf{x} are related to the contact pressures $p(\mathbf{y})$ at other points as follows (Johnson, 1985; Barber, 2010):

$$u(\mathbf{x}) = \int_S H(\mathbf{x}, \mathbf{y}) p(\mathbf{y}) d\mathbf{y}, \quad (1)$$

where $H(\mathbf{x}, \mathbf{y})$ represents the displacement at a point \mathbf{x} due to a surface contact pressure p acting at \mathbf{y} and S is the elastic half-plane. For homogeneous, isotropic, linear elastic materials, the influence coefficients are:

$$H(\mathbf{x}, \mathbf{y}) = \frac{1 - \nu^2}{\pi E} \frac{1}{\|\mathbf{x} - \mathbf{y}\|}, \quad (2)$$

¹ According to documentation, `mldivide` solves linear systems with symmetric positive definite matrices by computing a Cholesky factorization, see <http://www.mathworks.it/help/MATLAB/ref/mldivide.html>.

where E and ν denote, respectively, the composite Young's modulus and Poisson's ratio of half-plane, and $\|\cdot\|$ the standard Euclidean norm. The total contact force P is the integral of the contact traction field

$$P = \int_S p(\mathbf{x}) d\mathbf{x}. \quad (3)$$

By referring to Fig. 1, in the following we define for each surface point $\mathbf{x} \in S$ its elevation $\zeta(\mathbf{x})$, measured with respect to a reference frame, and set $\zeta_{\max} \triangleq \max_{\mathbf{x} \in S} \zeta(\mathbf{x})$ the maximum elevation. The indentation of the half plane at the points in contact is denoted by \bar{u} , whereas a generic displacement along the surface is u .

We consider the following problem:

Problem 1. For a given far-field displacement $\Delta \geq 0$ in the direction perpendicular to the undeformed half-plane, find the solution of the normal contact problem $u(\mathbf{x}), p(\mathbf{x})$ satisfying (1) and the unilateral contact (linear complementarity) conditions

$$u(\mathbf{x}) - \bar{u}(\mathbf{x}, \Delta) \geq 0, \quad (4a)$$

$$p(\mathbf{x}) \geq 0, \quad (4b)$$

$$(u(\mathbf{x}) - \bar{u}(\mathbf{x}, \Delta))p(\mathbf{x}) = 0, \quad (4c)$$

for all points $\mathbf{x} \in S$, where contact tractions are positive when compressive.

Introducing the quantity $w(\mathbf{x}, \Delta) = u(\mathbf{x}) - \bar{u}(\mathbf{x}, \Delta)$, Eq. (4) can be rewritten as:

$$w(\mathbf{x}, \Delta) \geq 0, \quad (5a)$$

$$p(\mathbf{x}) \geq 0, \quad (5b)$$

$$w(\mathbf{x}, \Delta)p(\mathbf{x}) = 0. \quad (5c)$$

Problem 1 is an infinite-dimensional linear complementarity problem. We find a finite-dimensional approximate solution by discretizing the surface as a square grid of spacing δ consisting of $N \times N$ average heights. Let S_{ij} be the cell of area δ^2 indexed by $i, j \in I_N$, with $I_N \triangleq \{1, \dots, N\} \times \{1, \dots, N\}$. Let $\mathbf{x}_{ij} \triangleq \int_{\mathbf{x} \in S_{ij}} \mathbf{x} d\mathbf{x}$, $\zeta_{ij} \triangleq \int_{\mathbf{x} \in S_{ij}} \zeta(\mathbf{x}) d\mathbf{x}$, $p_{ij} \triangleq \int_{\mathbf{x} \in S_{ij}} p(\mathbf{x}) d\mathbf{x}$, and $u_{ij} \triangleq \int_{\mathbf{x} \in S_{ij}} u(\mathbf{x}) d\mathbf{x}$ be, respectively, the barycentric coordinate, average height, resultant of the contact tractions, and the corresponding displacement on the surface element S_{ij} . Consider the following discretized version of (1)

$$u_{ij} = \sum_{k=1}^N \sum_{l=1}^N H_{i-k, j-l} p_{kl} \quad (6)$$

for all $(i, j) \in I_N$, where the term $H_{i-k, j-l}$ is the Green function used in (1) averaged over the elementary area δ^2 , which corresponds to the displacement induced by a uniformly loaded square:

$$H_{i-k, j-l} = \frac{1}{\delta^2} \int_{S_{ij}} \int_{S_{kl}} H(\mathbf{x}, \mathbf{y}) dy d\mathbf{x}, \quad (7)$$

and

$$p_{kl} \geq 0, \quad \forall (k, l) \in I_N. \quad (8)$$

For instance, Borri-Brunetto et al. (1999) used the following approximation related to a uniform pressure acting on a rounded patch of radius $\delta/2$:

$$H_{i-k, j-l} = \begin{cases} \frac{2}{E\pi\delta}, & \text{if } i = k \text{ and } j = l \\ \frac{2}{E\pi\delta} \arcsin \frac{\delta}{2\|\mathbf{x}_{ij} - \mathbf{x}_{kl}\|}, & \text{if } i \neq k, \quad j \neq l \end{cases} \quad (9)$$

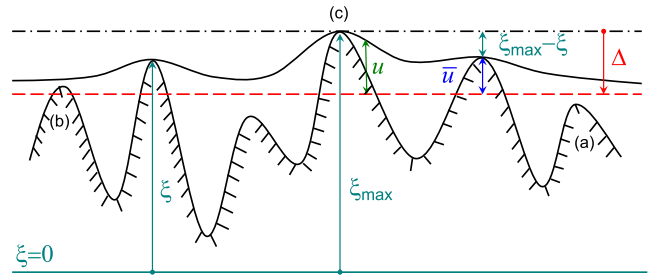


Fig. 1. Sketch of the contact problem between a rigid rough surface and an elastic half-plane. Its deformed configuration corresponding to the imposed far-field displacement Δ is depicted with a black solid line. The red dashed line corresponding to a rigid-body motion of the half-plane identifies the heights to be included in the initial trial contact domain. Once Problem 1 is solved we may have: (i) heights certainly not in contact from the beginning, type (a); (ii) heights losing contact due to elastic interactions, type (b); (iii) heights in contact, type (c). (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

but other formulae for a square patch can also be taken as in (Pohrt and Li, 2014).

Let $\bar{I}_C \triangleq \{(i, j) \in I_N : \zeta_{ij} < \zeta_{\max} - \Delta\}$ be the set of indices corresponding to elements S_{ij} that are certainly not in contact (cf. Fig. 1), and hence

$$p_{kl} = 0, \quad \forall (k, l) \in \bar{I}_C, \quad (10)$$

let $m = \#\bar{I}_C$ be the number of elements of \bar{I}_C and $n = \#I_C$ the number of elements belonging to the initial trial contact domain, $I_C \triangleq I_N \setminus \bar{I}_C$. The set I_C is only a superset of the set I_C^c of actual contact points, since the corrections to the displacements induced by elastic interactions may induce lack of contact in some elements (i, j) , i.e., $u_{ij} > \bar{u}_{ij}$, where $\bar{u}_{ij} \triangleq \Delta - \zeta_{\max} + \zeta_{ij}$ is the value of the compenetration of the height corresponding to the element (i, j) in the half-plane (see Fig. 1).

For a generic $(i, j) \in I_C$ corresponding to an element of the surface which is potentially in contact with the elastic half-plane, we denote by

$$w_{ij} \triangleq u_{ij} - \bar{u}_{ij} \geq 0 \quad (11)$$

the corresponding elastic correction to the displacement. Clearly, it must hold that

$$w_{ij} p_{ij} = 0, \quad \forall (i, j) \in I_C \quad (12)$$

since $w_{ij} > 0$ implies no contact between the surfaces and therefore no pressure, while $p_{ij} > 0$ implies contact, $u_{ij} = \bar{u}_{ij}$, or equivalently $w_{ij} = 0$.

By taking into account that $p_{kl} = 0$ for all $(k, l) \in \bar{I}_C$, Eq. (6) can be recast as the following condition

$$w_{ij} + \bar{u}_{ij} = \sum_{(k,l) \in I_C} H_{i-k, j-l} p_{kl}, \quad \forall (i, j) \in I_C, \quad (13)$$

which is limited to the nodes belonging to the initial trial contact domain I_C , whose number of elements is in general significantly smaller than those of I_N . The relations (8)–(13) can be recast in matrix form as the following Linear Complementarity Problem (LCP) (Cottle et al., 1992):

$$\mathbf{w} = \mathbf{H}\mathbf{p} - \bar{\mathbf{u}} \quad (14a)$$

$$\mathbf{w} \geq \mathbf{0}, \quad \mathbf{p} \geq \mathbf{0}, \quad \mathbf{w}'\mathbf{p} = 0, \quad (14b)$$

where $\mathbf{w} \in \mathbb{R}^n$ is the vector of unknown elastic corrections $w_{ij}, (i, j) \in I_C$, \mathbf{w}' denotes its transpose, $\mathbf{p} \in \mathbb{R}^n$ is the vector of unknown average contact forces $p_{ij}, (i, j) \in I_C$, $\bar{\mathbf{u}} \in \mathbb{R}^n$ is the vector of compenetrations $\bar{u}_{ij}, (i, j) \in I_C$, and $\mathbf{H} = \mathbf{H}'$ is the matrix obtained

by collecting the compliance coefficients H_{i-kj-l} , for $(i,j), (k,l) \in I_C$. Due to the properties of linear elasticity (Johnson, 1985, p.144) we have that

$$\mathbf{H} = \mathbf{H}' \succ 0, \quad (15)$$

that is \mathbf{H} is a symmetric positive definite matrix (with the additional property deriving from (9) of having all its entries positive). After solving (14), the vector $\mathbf{u} \in \mathbb{R}^n$ of normal displacements $u_{ij}, (i,j) \in I_C$ is then simply retrieved as $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{w}$.

By the positive definiteness property (15) of \mathbf{H} , we inherit immediately the following important property (Cottle et al., 1992, Th. 3.3.7):

Property 1. The discretized version (8), (10)–(13) of Problem 1 admits a unique solution \mathbf{p}, \mathbf{u} , for all $\Delta \geq 0$.

The LCP problem (14) corresponds to the Karush–Kuhn–Tucker (KKT) conditions for optimality of the following convex quadratic program (QP)

$$\min_{\mathbf{p}} \frac{1}{2} \mathbf{p}' \mathbf{H} \mathbf{p} - \bar{\mathbf{u}}' \mathbf{p} \quad (16a)$$

$$\text{s.t. } \mathbf{p} \geq \mathbf{0} \quad (16b)$$

in that the solution \mathbf{p} of (16) and its corresponding optimal dual solution \mathbf{w} solve (14), and vice versa.

Problem (16) is consistent with former pioneering considerations by Kalker and van Randen (1972) and also summarized in Johnson (1985, p.151–152). In fact, the contact pressures solving the unilateral contact problem can be obtained by minimizing the total complementary energy W of the linear elastic system, subject to the constraint $p(\mathbf{x}) \geq 0, \forall \mathbf{x} \in S$. For a continuous system, the total complementary energy is

$$W = U - \int_S p(\mathbf{x}) \bar{u}(\mathbf{x}, \Delta) d\mathbf{x}, \quad (17)$$

where U is the internal complementary energy of the deformed half-plane in contact. For linear elastic materials, we have:

$$U = \frac{1}{2} \int_S p(\mathbf{x}) u(\mathbf{x}) d\mathbf{x}. \quad (18)$$

Although such an energy-based approach can be used to derive FEM formulations, it is also possible to remain within BEM and introduce a surface discretization as before. By invoking the averaged Green's functions in (7), the discretized version of \tilde{W} leads to

$$\tilde{W} = \frac{1}{2} \sum_{(ij) \in I_C} \sum_{(kl) \in I_C} H_{i-kj-l} p_{kl} p_{ij} - \sum_{(ij) \in I_C} p_{ij} \bar{u}_{ij} \quad (19)$$

which represents a quadratic function of \mathbf{p} to be minimized, under the constraints $p_{ij} \geq 0, \forall (i,j) \in I_C$, as in (16). Since it is unlikely that the contact area is known a priori, the active set of nodes in contact results only after solving problem (14) or equivalently (16).

A large variety of solvers for LCP and QP problems were developed in the last 60 years (Beale, 1955; Fletcher, 1971; Goldfarb and Idnani, 1983; Cottle et al., 1992; Schmid and Biegler, 1994; Patrinos and Bemporad, 2014), and is still an active area of research in the optimization and control communities. Historically, in the mechanics community, Kalker and van Randen (1972) proposed the simplex method, although it was found to be practical only for relatively small N . More recent contributions adopt algorithms to solve the unconstrained linear system of equations and then correct the solution by eliminating the boundary elements bearing tensile tractions (Francis, 1983; Borri-Brunetto et al., 1999; Borri-Brunetto et al., 2001), or use a constrained version of the Conjugate Gradient (CG) algorithm (Polonsky and Keer, 1999). These methods are simply initialized by considering arbitrary non-

negative entries in \mathbf{p} , without taking advantage of the monotonic increase (or decrease) of pressures by increasing (or decreasing) the far-field displacement, an important property guaranteed by rigorous elasticity theorems (Barber, 1974). The history of pressures is saved during a contact simulation and it is easy to access and use and it can be beneficial to save computation time.

Next section presents effective optimization algorithms for solving the QP problem (16) and compares their performance with respect to the Greedy CG method. Contrary to the latter, not only the considered QP have the guaranteed property of always converging to the unique solution \mathbf{p}, \mathbf{u} for any given $\Delta \geq 0$, but also the history of loading can be more efficiently taken into account as a warm-start, with a significant saving of computation time.

3. Optimization algorithms

Since now on, we use the subscript i to denote the i -th component of a vector or the i -th row of a matrix, the subscript \mathcal{I} to denote the subvector obtained by collecting all the components $i \in \mathcal{I}$ of a vector (or all the rows i of a matrix), and the double subscript $\mathcal{I}, \mathcal{I}_1$ to denote the submatrix obtained by collecting the i -th row and j -th column, for all $i \in \mathcal{I}, j \in \mathcal{I}_1$.

3.1. Greedy methods

A greedy method corresponds to solve problem (16) by iteratively solving the unconstrained linear system of equations $\mathbf{w} = \mathbf{H} \mathbf{p} - \bar{\mathbf{u}} = \mathbf{0}$ with respect to \mathbf{p} and increasingly zeroing negative elements of \mathbf{p} until the condition $\mathbf{p} \geq \mathbf{0}$ is satisfied. By construction we obtain $\mathbf{w}' \mathbf{p} = 0$. The method is described in Algorithm 1, in which a standard Conjugate Gradient employed to solve the unconstrained linear system of equations. Steps 2.1–2.4 can be replaced by any other algorithm for solving the linear system of equations, like the Gauss–Seidel iterative scheme as in (Borri-Brunetto et al., 1999, 2001), the MATLAB's `mldivide` solver, or even the FFT algorithm as in (Karpenko and Akay, 2001; Batrouni et al., 2002).

Algorithm 1. Greedy method with Conjugate Gradient (greedy CG)

Input: Matrix $\mathbf{H} = \mathbf{H}' \succ 0$, vector $\bar{\mathbf{u}}$; initial guess \mathbf{p} and initial active set $\mathcal{I} \subseteq \{1, \dots, n\}$ such that $\mathbf{p}_{\{1, \dots, n\} \setminus \mathcal{I}} = \mathbf{0}$; maximum number K_{\max} of iterations, tolerance $\epsilon > 0$.

(1) $i \leftarrow 0; \bar{\mathcal{I}} \leftarrow \{1, \dots, n\} \setminus \mathcal{I}$;

(2) **while** ($i \leq K_{\max}$ **and** $\min(\mathbf{p}) < -\epsilon$) **or** $i = 0$ **do**:

(2.1) $\mathbf{w}_{\mathcal{I}} \leftarrow \mathbf{H}_{\mathcal{I}, \mathcal{I}} \mathbf{p}_{\mathcal{I}} - \bar{\mathbf{u}}_{\mathcal{I}}$;

(2.2) $n_w \leftarrow \|\mathbf{w}_{\mathcal{I}}\|_2$;

(2.3) $\mathbf{b}_{\mathcal{I}} \leftarrow -\mathbf{w}_{\mathcal{I}}$

(2.4) **while** $n_w > \epsilon$ **and** $i \leq K_{\max}$ **do**:

(2.4.1) $\mathbf{s}_{\mathcal{I}} \leftarrow \mathbf{H}_{\mathcal{I}, \mathcal{I}} \mathbf{b}_{\mathcal{I}}$;

(2.4.2) $\mathbf{p}_{\mathcal{I}} \leftarrow \mathbf{p}_{\mathcal{I}} - \frac{\mathbf{w}_{\mathcal{I}}' \mathbf{b}_{\mathcal{I}}}{\mathbf{b}_{\mathcal{I}}' \mathbf{s}_{\mathcal{I}}} \mathbf{b}_{\mathcal{I}}$;

(2.4.3) $\bar{\mathbf{w}}_{\mathcal{I}} \leftarrow \mathbf{H}_{\mathcal{I}, \mathcal{I}} \mathbf{p}_{\mathcal{I}} - \bar{\mathbf{u}}_{\mathcal{I}}$;

(2.4.4) $\mathbf{b}_{\mathcal{I}} \leftarrow -\bar{\mathbf{w}}_{\mathcal{I}} + \frac{\mathbf{w}_{\mathcal{I}}' \mathbf{s}_{\mathcal{I}}}{\mathbf{b}_{\mathcal{I}}' \mathbf{s}_{\mathcal{I}}} \mathbf{b}_{\mathcal{I}}$;

(2.4.5) $\mathbf{w}_{\mathcal{I}} \leftarrow \bar{\mathbf{w}}_{\mathcal{I}}$;

(2.4.6) $n_w \leftarrow \|\mathbf{w}_{\mathcal{I}}\|_2$;

(2.4.7) $i \leftarrow i + 1$;

(2.5) **for** $j \in \bar{\mathcal{I}}$ **do**:

(2.5.1) **if** $\mathbf{p}_j < -\epsilon$ **then** $\mathbf{p}_j \leftarrow 0$;

$\bar{\mathcal{I}} \leftarrow \bar{\mathcal{I}} \setminus \{j\}; \bar{\mathcal{I}} \leftarrow \bar{\mathcal{I}} \cup \{j\}$;

(3) $\mathbf{p}^* \leftarrow \mathbf{p}$;

(4) $\mathbf{u}_{\mathcal{I}}^* \leftarrow \bar{\mathbf{u}}_{\mathcal{I}} + \mathbf{H}_{\mathcal{I}, \mathcal{I}}^{-1} \mathbf{p}_{\mathcal{I}}^*$;

(5) **end**.

Output: Contact force vector \mathbf{p}^* and normal displacement vector \mathbf{u}^* .

Assuming that the prescribed initial \mathbf{p} and \mathcal{I} are such that $p_j = 0$ for all $j \in \{1, \dots, n\} \setminus \mathcal{I}$, and K_{\max} is sufficiently large, the output of the greedy algorithm leads to a contact pressure vector \mathbf{p}^* and a normal displacement vector \mathbf{u}^* satisfying $\mathbf{u}^* = \mathbf{H}\mathbf{p}^*$, $\mathbf{p}^* \geq \mathbf{0}$, $(\mathbf{u}^* - \bar{\mathbf{u}})' \mathbf{p}^* = 0$. In fact, condition $\mathbf{p}^* \geq \mathbf{0}$ is guaranteed by the condition in Step 2 up to ϵ precision. By letting $\mathbf{w}^* \triangleq \mathbf{u}^* - \bar{\mathbf{u}}$, at termination of the algorithm we have $\mathbf{w}^*_{\mathcal{I}} = \mathbf{H}_{\mathcal{I},\mathcal{I}}\mathbf{p}^*_{\mathcal{I}} - \bar{\mathbf{u}}_{\mathcal{I}} = \mathbf{0}$ because of the solution of the CG method (Step 2.4), or equivalently $\mathbf{u}^*_{\mathcal{I}} = \bar{\mathbf{u}}_{\mathcal{I}}$ (cf. Step 5). By setting $\mathbf{u}^*_{\bar{\mathcal{I}}} \triangleq \mathbf{H}_{\bar{\mathcal{I}},\bar{\mathcal{I}}}\mathbf{p}^*_{\bar{\mathcal{I}}}$ in Step 5, and recalling that $\mathbf{p}^*_{\bar{\mathcal{I}}} = \mathbf{0}$, we have

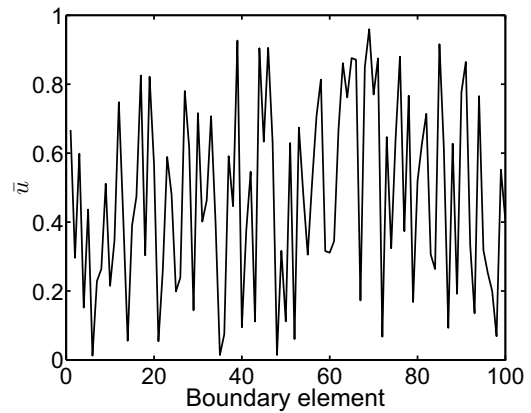
$$\begin{bmatrix} \mathbf{w}^*_{\mathcal{I}} \\ \mathbf{w}^*_{\bar{\mathcal{I}}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{H}_{\bar{\mathcal{I}},\mathcal{I}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{p}^*_{\mathcal{I}} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ -\bar{\mathbf{u}}_{\bar{\mathcal{I}}} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{\mathcal{I},\mathcal{I}} & \mathbf{H}_{\mathcal{I},\bar{\mathcal{I}}} \\ \mathbf{H}_{\bar{\mathcal{I}},\mathcal{I}} & \mathbf{H}_{\bar{\mathcal{I}},\bar{\mathcal{I}}} \end{bmatrix} \begin{bmatrix} \mathbf{p}^*_{\mathcal{I}} \\ \mathbf{p}^*_{\bar{\mathcal{I}}} \end{bmatrix} + \begin{bmatrix} -\bar{\mathbf{u}}_{\mathcal{I}} \\ -\bar{\mathbf{u}}_{\bar{\mathcal{I}}} \end{bmatrix}$$

and hence $\mathbf{u}^* = \mathbf{w}^* + \bar{\mathbf{u}} = \mathbf{H}\mathbf{p}^*$. The complementarity condition $(\mathbf{u}^* - \bar{\mathbf{u}})' \mathbf{p}^* = (\mathbf{w}^*)' \mathbf{p}^* = 0$ follows by construction, as Step 2.4 zeroes all the components of \mathbf{w}^*_j that correspond to nonnegative \mathbf{p}^*_j , $\forall j \in \mathcal{I}$, and zeroes all the components \mathbf{p}^*_j that correspond to possible non-zero components \mathbf{w}^*_j , $\forall j \in \bar{\mathcal{I}}$.

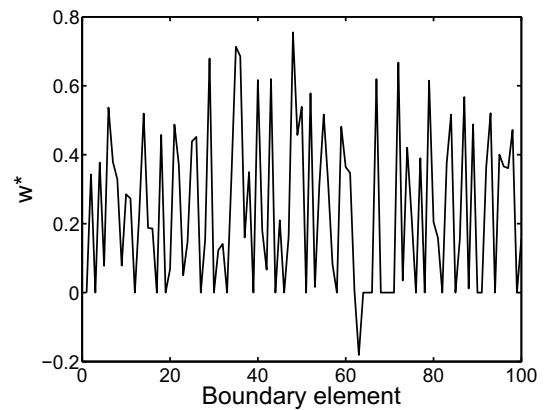
However, to the best of the authors' knowledge, no formal proof exists that the condition $\mathbf{w}^*_{\mathcal{I}} \geq \mathbf{0}$ is satisfied after the algorithm terminates, i.e., that $\mathbf{u}^* \geq \bar{\mathbf{u}}$. If the algorithm is applied to randomly generated $\bar{\mathbf{u}}$ vectors and \mathbf{H} positive definite matrices with positive coefficients, in many cases the LCP is not solved exactly. In contact mechanics, the only evidence that this condition is satisfied has been shown in simulations (see, e.g., [Batrouni et al., 2002](#)). Indeed, we obtained the following counterexample in which the greedy method failed in getting the solution also for \mathbf{H} whose coefficients are given by Eq. (9).²

Example 1. Consider a square mesh with grid spacing δ consisting of $N \times N$ boundary elements indexed by $(i,j) \in I_N, I_N = \{1, \dots, N\} \times \{1, \dots, N\}$. Suppose that all the boundary elements are included in the initial trial contact domain I_C ($n = N \times N$) by assigning a positive value of $\bar{\mathbf{u}}_{i,j}$ to all elements. This may represent a situation where a cluster of densely packed heights comes into contact. Since $\bar{\mathbf{u}}_{i,j}$ depends on the height field $\xi_{i,j}$, which is a random variable, for the sake of generality we extract the values $\bar{\mathbf{u}}_{i,j}$ randomly from a uniform distribution in the interval $(0, 1)$. The matrix \mathbf{H} is assembled according to (9). By running a sequence of 100 random simulations, we find that in approximately 40% of the simulations the greedy method provides a solution which violates the condition $\mathbf{w}^*_{i,j} \geq 0$ in at least one element. This lack of convergence to the right solution is observed for any size n of the problem. One of the wrong results obtained for $n = 100$ is shown in Fig. 2. The assigned random values of $\bar{\mathbf{u}}$ are plotted in Fig. 2(a) for the sequence of boundary elements (from 1 to 100) composing the mesh. The solution \mathbf{w}^* presents a negative entry in one single element (element 62 in Fig. 2(b)). The computed contact forces are compared in Fig. 2(c) with the values corresponding to the exact solution of the problem (green dots) obtained by using the NNLS algorithm presented in Section 3.3, that is proven to satisfy the LCP conditions (14) exactly. Although just one value of \mathbf{w}^* is negative, the overall solution is affected by this violation. We observe in fact a false contact detection for the element number 62 violating the condition $\mathbf{w}^*_{i,j} > 0$, a contact not detected (element 81) and 7 contact forces significantly underestimated with respect to the exact ones. ■

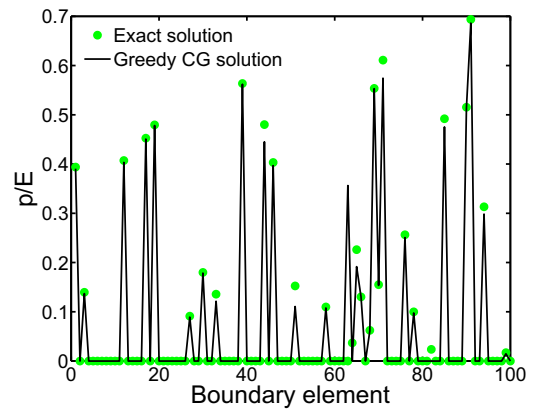
For less densely packed boundary elements belonging to I_C , for instance with a minimum distance of 2δ between them instead of δ as in Example 1, the algorithm was found to always provide a solution satisfying the condition $\mathbf{w}^* \geq \mathbf{0}$. Other benchmark tests



(a) Random $\bar{\mathbf{u}}$ as input



(b) Computed $\mathbf{w}^*_{\bar{\mathcal{I}}}$



(c) Computed \mathbf{P}/E

Fig. 2. Counterexample showing that the Greedy CG method fails in getting the correct solution ($\delta = 1$ a.u. of L , as \mathbf{u} and \mathbf{w}^* : $E = 0.01 F/L^2$). Green dots correspond to the correct contact forces satisfying the LCP and are obtained by using the NNLS method, Section 3.2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

considering a deterministic smooth variation of $\bar{\mathbf{u}}$, as in case of an indentation by a smooth sphere or by a flat punch, did not show any convergence problem to the solution as well, although the boundary elements in contact are densely packed as in the

² The MATLAB routine of the counterexample is available for download at <http://musam.imtluccha.it/counterexample.m>.

counterexample shown before. In conclusion, although it is likely that the diagonally dominant property of the matrix \mathbf{H} plays a role in the robustness of the algorithm, it remains an open problem to find exact mathematical requirements for \mathbf{H} and $\bar{\mathbf{u}}$ that guarantee the greedy method to provide a solution satisfying $\mathbf{w}^* \geq \mathbf{0}$, so that all the LCP conditions (14) are met.

Therefore, as a word of caution, the reliability of the greedy method should be carefully checked in case of applications of BEM to contact problems governed by other forms of \mathbf{H} , as in the case of contact with an anisotropic or an inhomogeneous half-space, or in the presence of multiple fields.

Another drawback of the algorithm is the difficulty to warm start the method with a proper choice of the initial active set \mathcal{I} . Since at Step 2.5.1 the number of elements in the sequence \mathcal{I} is decreased by removing negative enough components \mathbf{p}_j of the current solution vector, i.e., eliminating the points bearing tensile (negative) forces, in a monotonic way (no index j that has been removed from \mathcal{I} can be added back), a safe cold start is to set $\mathcal{I} = \{1, \dots, n\}$ and pick up a vector $\mathbf{p} \geq \mathbf{0}$, usually a vector with arbitrary non-negative numbers. The history of contact forces obtained during the solution of a sequence of imposed displacements is not taken into account by the method to accelerate its convergence, although we know that contact forces are monotonically increasing functions of the far-field displacement. In any case, for a complex sequence of loading with an increased or decreased far-field displacement, any warm starting on forces cannot be implemented in the method, since the elimination of contact points is irreversible.

3.2. Constrained Conjugate Gradient

A constrained CG algorithm was proposed by Polonsky and Keer (1999) based on the theory by Hestenes et al. (1980, Ch. 2.3) to solve the linear system of equations and rigorously impose the satisfaction of the contact constraints. Algorithm 2 has been applied by Polonsky and Keer (1999) to simulations under load control. However, it can be used also for displacement control. The condition for convergence set by Polonsky and Keer (1999) in terms of relative variation in the local contact forces from an iteration to the next has been recast in terms of the error in the local contact displacements. The two criteria are completely equivalent.

This constrained CG algorithm does not remove the points bearing tensile forces from the active set. Therefore, the size of the linear system of equations is not reduced during the iterations, increasing the computation time for its solution. On the other hand, the method assures the satisfaction of the LCP conditions (14) and it is found to converge with a reduced number of iterations as compared to the Greedy CG algorithm. Although not investigated in (Polonsky and Keer, 1999), it can be warm started in case of a sequence of loading steps by considering both an initial trial contact domain and a set of contact pressures derived from the previous converged solution. The FFT method can be used to accelerate step (3.8) as in (Polonsky and Keer, 2000a).

Algorithm 2. Constrained Conjugate Gradient

Input: Matrix $\mathbf{H} = \mathbf{H}^T \succ \mathbf{0}$, vector $\bar{\mathbf{u}}$, initial guess $\mathbf{p} \geq \mathbf{0}$, initial active set $\mathcal{I} = \{1, \dots, n\}$; maximum number K_{\max} of iterations, tolerance $\epsilon > 0$.

- (1) $i \leftarrow 0, n_{w,\text{old}} = 1, d = 0, \text{err} = +\infty$;
- (2) $\mathbf{w} \leftarrow \mathbf{H}\mathbf{p} - \bar{\mathbf{u}}$;

- (3) **while** ($i \leq K_{\max}$ **and** $\text{err} > \epsilon$):
 - (3.1) **if** $i = 0$ **then** $\mathbf{t} \leftarrow \mathbf{w}$ **else:** $\mathbf{t} \leftarrow \mathbf{w} + d \frac{n_w}{n_{w,\text{old}}} \mathbf{t}_{\text{old}}$;
 - (3.2) $\tau = \frac{\mathbf{w}^T \mathbf{t}}{\mathbf{t}^T \mathbf{t}}$;
 - (3.3) $\mathbf{p} \leftarrow \mathbf{p} - \tau \mathbf{t}$;
 - (3.4) $\forall j \in \mathcal{I} : p_j \leftarrow \max\{p_j, 0\}$;
 - (3.5) Find $I_{ol} = \{j \in \mathcal{I} : p_j = 0, w_j < 0\}$; **if** $I_{ol} = \emptyset$ **then** $d = 1$ **else** $d = 0; p_j \leftarrow p_j - \tau w_j, \forall j \in I_{ol}$;
 - (3.6) $\mathcal{I} \leftarrow \{j : p_j > 0\} \cup I_{ol}$;
 - (3.7) $\mathbf{t}_{\text{old}} \leftarrow \mathbf{t}, n_{w,\text{old}} \leftarrow n_w$;
 - (3.8) $\mathbf{w} \leftarrow \mathbf{H}\mathbf{p} - \bar{\mathbf{u}}$;
 - (3.9) $n_w = \|\mathbf{w}\|_2$;
 - (3.10) $\text{err} \leftarrow |n_w - n_{w,\text{old}}|/n_{w,\text{old}}$;
 - (3.11) $i \leftarrow i + 1$;
- (4) $\mathbf{p}^* \leftarrow \mathbf{p}; \mathbf{u}^* = \mathbf{H}\mathbf{p}^*$;
- (5) **end.**

Output: Contact force vector \mathbf{p}^* and normal displacement vector \mathbf{u}^* .

3.3. Non-Negative Least Squares (NNLS)

In this section we show how a QP problem with positive definite Hessian matrix having the special form (16) can be effectively solved as a nonnegative least squares problem.

Thanks to property (15), matrix \mathbf{H} admits a Cholesky factorization $\mathbf{H} = \mathbf{C}^T \mathbf{C}$. Hence we can *theoretically* recast problem (16) as the Non-Negative Least Squares (NNLS) problem:

$$\min_{\mathbf{p}} \frac{1}{2} \|\mathbf{C}\mathbf{p} - \mathbf{C}^{-T}\bar{\mathbf{u}}\|_2^2 \quad (20a)$$

$$\text{s.t. } \mathbf{p} \geq \mathbf{0} \quad (20b)$$

A simple and effective active-set method for solving the NNLS problem (20) is the one in (Lawson and Hanson, 1974, p.161), that is extended here in Algorithm 3 to directly solve (16) without explicitly computing the Cholesky factor \mathbf{C} and its inverse \mathbf{C}^{-1} and to handle warm starts. After a finite number of steps, Algorithm 3 converges to the optimal contact force vector \mathbf{p}^* and returns the normal displacement vector \mathbf{u}^* whose components p_{ij}, u_{ij} satisfy $p_{ij} \geq 0, u_{ij} \geq \bar{u}_{ij}, (u_{ij} - \bar{u}_{ij})p_{ij} = 0$, and (13), $\forall (i, j) \in I_C$.

The method is easy to warm start in case of a loading scenario consisting of an alternating sequence of increasing or decreasing far-field displacements. The contact forces determined for a given imposed displacement are used to initialize vector \mathbf{p} . Due to the monotonicity of the contact solution, this initialization is certainly much closer to the optimal solution \mathbf{p}^* than a zero vector. This usually significantly reduces the iterations of the method to convergence. Such a warm start has a fast implementation requiring a projection of the forces of the points belonging to $I_C^e(\Delta_k)$ to the same points of the trial domain $I_C^e(\Delta_{k+1})$ for a new imposed far field displacement Δ_{k+1} . For an increasing far-field displacement, i.e., $\Delta_{k+1} > \Delta_k$ the forces in the elements belonging to $I_C^e(\Delta_{k+1}) - I_C^e(\Delta_k)$ are simply initialized equal to zero. In the numerical experiments of Section 4, the time required for this projection will be added to the global solution time for a consistent comparison with the greedy method with cold start and with the constrained CG algorithm.

Algorithm 3. Non-Negative Least Squares (NNLS)

Input: Matrix $\mathbf{H} = \mathbf{H}' > 0$, vector $\bar{\mathbf{u}}$, initial guess \mathbf{p} ; maximum number K_{\max} of iterations, tolerance $\epsilon > 0$.

- (1) $\mathcal{I} \leftarrow \{i \in \{1, \dots, n\} : \mathbf{p}_i > \mathbf{0}\}$; $init \leftarrow FALSE$; $k \leftarrow 0$;
- (2) **if** $\mathcal{I} = \emptyset$ **then** $init \leftarrow TRUE$;
- (3) $\mathbf{w} \leftarrow \mathbf{H}\mathbf{p} - \bar{\mathbf{u}}$;
- (4) **if** $((\mathbf{w} \geq -\epsilon$ **or** $\mathcal{I} = \{1, \dots, n\})$ **and** $init = TRUE)$ **or** $k \geq K_{\max}$ **then go to** Step 13;
- (5) **if** $init = TRUE$ **then** $i \leftarrow \arg \min_{i \in \{1, \dots, n\} \setminus \mathcal{I}} \mathbf{w}_i$; $\mathcal{I} \leftarrow \mathcal{I} \cup \{i\}$; **else** $init \leftarrow TRUE$;
- (6) $\mathbf{s}_i \leftarrow$ solution of the linear system $\mathbf{H}_i \mathbf{s}_i = \bar{\mathbf{u}}_i$;
- (7) **if** $\mathbf{s}_i \geq -\epsilon$ **then** $\mathbf{p} \leftarrow \mathbf{s}$ **and go to** Step 3;
- (8) $j \leftarrow \arg \min_{h \in \mathcal{I} : \mathbf{s}_h \leq \mathbf{0}} \left\{ \frac{\mathbf{p}_h}{\mathbf{p}_h - \mathbf{s}_h} \right\}$;
- (9) $\mathbf{p} \leftarrow \mathbf{p} + \frac{\mathbf{p}_j}{\mathbf{p}_j - \mathbf{s}_j} (\mathbf{s} - \mathbf{p})$;
- (10) $\mathcal{I}_0 \leftarrow \{h \in \mathcal{I} : \mathbf{p}_h = \mathbf{0}\}$;
- (11) $\mathcal{I} \leftarrow \mathcal{I} \setminus \mathcal{I}_0$; $k \leftarrow k + 1$;
- (12) **go to** Step 6;
- (13) $\mathbf{p}^* \leftarrow \mathbf{p}$;
- (14) $\mathbf{u}^* \leftarrow \mathbf{w} + \bar{\mathbf{u}}$;
- (15) **end.**

Output: Contact force vector \mathbf{p}^* and normal displacement vector \mathbf{u}^* satisfying $\mathbf{u}^* = \mathbf{H}\mathbf{p}^*, \mathbf{u}^* \geq \bar{\mathbf{u}}, \mathbf{p}^* \geq \mathbf{0}, (\mathbf{u}^* - \bar{\mathbf{u}})' \mathbf{p} = 0$.

Note that Step 6 of Algorithm 3 is equivalent to Step 2.4 of Algorithm 1 and has been performed by using the MATLAB's `mldivide` solver. This step can be accelerated by the use of an approach based on the FFT (for its implementation, see e.g. Batrouni et al., 2002). Alternatively, since the set \mathcal{I}_0 changes incrementally during the iterations of the algorithm, more efficient iterative QR (Lawson and Hanson, 1974, Chap. 24) or LDL^T (Bemporad, in press) factorization methods can be employed.

3.3.1. Warm-started NNLS via accelerated Gradient Projection (NNLS + GP)

An alternative method to solve Problem (16) is to use an accelerated gradient projection (GP) method for QP (Nesterov, 1983; Patrinos and Bemporad, 2014). Because of the simple nonnegative constraints in (16), rather than going to the dual QP formulation as in (Patrinos and Bemporad, 2014), we formulate the GP problem directly for the primal QP problem (16). Numerical experiments have shown slow convergence of a pure accelerated GP method to solve (16). However, we can use the method to *warm start* Algorithm 3, as described in Algorithm 4. If Algorithm 4 is executed ($K > 0$), it returns a vector \mathbf{p} that is immediately used as an input to Algorithm 3, otherwise one can simply set $\mathbf{p} = \mathbf{0}$ (cold start). As shown in Section 4, GP iterations provide large benefits in warm starting the NNLS solver, therefore allowing taking the best advantages of the two methods: quickly getting in the neighborhood of the optimal solution (GP iterations of Algorithm 4) and getting solutions up to machine precision after a finite number of iterations (the active-set NNLS Algorithm 3).

Algorithm 4. Accelerated Gradient Projection (GP)

Input: Matrix $\mathbf{H} = \mathbf{H}' > 0$ and its Frobenius norm L , vector $\bar{\mathbf{u}}$, initial guess \mathbf{p} , number K of iterations.

- (1) $\bar{\mathbf{p}} \leftarrow \mathbf{p}$;
- (2) **for** $i = 0, \dots, K - 1$ **do**:
 - (2.1) $\beta = \max \left\{ \frac{i-1}{i+2}, 0 \right\}$;
 - (2.2) $\mathbf{s} = \mathbf{p} + \beta(\mathbf{p} - \bar{\mathbf{p}})$;
 - (2.3) $\mathbf{w} = \mathbf{H}\mathbf{s} - \bar{\mathbf{u}}$;
 - (2.4) $\bar{\mathbf{p}} \leftarrow \mathbf{p}$;
 - (2.5) $\mathbf{p} \leftarrow \max \left\{ \mathbf{s} - \frac{1}{L} \mathbf{w}, \mathbf{0} \right\}$;
- (3) **end.**

Output: Warm start for contact force vector \mathbf{p} and elastic correction vector \mathbf{w} .

3.4. Alternating Direction Method of Multipliers (ADMM)

The QP problem (16) can also be solved by the Alternating Direction Method of Multipliers (ADMM), which belongs to the class of augmented Lagrangian methods. The reader is referred to (Boyd et al., 2011) for mathematical details. The method treats the QP (16) as the following problem

$$\min_{\mathbf{p}, \mathbf{s}} \frac{1}{2} \mathbf{p}' \mathbf{H} \mathbf{p} - \bar{\mathbf{u}}' \mathbf{p} + g(\mathbf{s}) \quad (21)$$

$$\text{s.t. } \mathbf{p} = \mathbf{s}$$

where

$$g(\mathbf{s}) = \begin{cases} 0 & \text{if } \mathbf{s} \geq \mathbf{0} \\ +\infty & \text{if } \mathbf{s} < \mathbf{0} \end{cases}$$

Then, the augmented Lagrangian function

$$L_\rho(\mathbf{p}, \mathbf{s}, \mathbf{w}) = \frac{1}{2} \mathbf{p}' \mathbf{H} \mathbf{p} - \bar{\mathbf{u}}' \mathbf{p} + g(\mathbf{s}) + \mathbf{w}'(\mathbf{p} - \mathbf{s}) + \frac{\rho}{2} \|\mathbf{p} - \mathbf{s}\|_2^2$$

is considered, where $\rho > 0$ is a parameter of the algorithm. The basic ADMM algorithm consists of the following iterations:

$$\begin{aligned} \mathbf{p}^{k+1} &= \arg \min_{\mathbf{p}} L_\rho(\mathbf{p}, \mathbf{s}^k, \mathbf{w}^k) \\ \mathbf{s}^{k+1} &= \arg \min_{\mathbf{s}} L_\rho(\mathbf{p}^{k+1}, \mathbf{s}, \mathbf{w}^k) \\ \mathbf{w}^{k+1} &= \mathbf{w}^k + \rho(\mathbf{p}^{k+1} - \mathbf{s}^{k+1}) \end{aligned} \quad (22)$$

A scaled form with over-relaxation of the ADMM iterations (22) is summarized in Algorithm 5. The algorithm is guaranteed to converge asymptotically to the solution $\mathbf{p}^*, \mathbf{u}^*$ of the problem. The over-relaxation parameter $\alpha > 1$ is introduced to improve convergence, typical values for α suggested in (Boyd et al., 2011) are $\alpha \in [1.5, 1.8]$.

A warm start of the algorithm that takes into account the loading history is possible in a way analogous to that described for the NNLS approach of Section 3.3. However, as an additional complexity, also an initialization for the dual variable vector \mathbf{w} must be provided, possibly obtained by projecting the solution obtained for a certain Δ_k to that for Δ_{k+1} .

Algorithm 5. Alternative Direction Method of Multipliers (ADMM)

Input: Matrix $\mathbf{H} = \mathbf{H}' > 0$, vector $\bar{\mathbf{u}}$, initial guesses \mathbf{p}, \mathbf{w} , parameter $\rho > 0$, over-relaxation parameter $\alpha > 1$, maximum number K_{\max} of iterations, tolerance $\epsilon > 0$.

- (1) $\mathbf{M} \leftarrow \left(\frac{1}{\rho} \mathbf{H} + \mathbf{I} \right)$;
- (2) $\mathbf{w}_\rho \leftarrow -\frac{1}{\rho} \mathbf{w}$;
- (3) $\mathbf{s} \leftarrow \mathbf{p}$;
- (4) $i \leftarrow 0$;
- (5) **while** $(i \leq K_{\max}$ **and** $\|\mathbf{p} - \mathbf{s}\|_\infty > \epsilon)$ **or** $i = 0$ **do**:
 - (5.1) $\mathbf{s} \leftarrow \mathbf{M}(\mathbf{p} - \mathbf{w}_\rho - \frac{1}{\rho} \bar{\mathbf{u}})$;
 - (5.2) $\bar{\mathbf{s}} \leftarrow \alpha \mathbf{s} + (1 - \alpha) \mathbf{p}$;
 - (5.3) $\mathbf{p} \leftarrow \max \{ \bar{\mathbf{s}} + \mathbf{w}_\rho, \mathbf{0} \}$;
 - (5.4) $\mathbf{w}_\rho \leftarrow \mathbf{w}_\rho + \bar{\mathbf{s}} - \mathbf{p}$;
 - (5.5) $i \leftarrow i + 1$;
- (6) $\mathbf{p}^* \leftarrow \mathbf{p}$;
- (7) $\mathbf{u}^* \leftarrow \bar{\mathbf{u}} - \rho \mathbf{w}_\rho$;
- (8) **end.**

Output: Contact force vector \mathbf{p}^* and normal displacement vector \mathbf{u}^* satisfying $\mathbf{u}^* = \mathbf{H}\mathbf{p}^*, \mathbf{u}^* \geq \bar{\mathbf{u}}, \mathbf{p}^* \geq \mathbf{0}, (\mathbf{u}^* - \bar{\mathbf{u}})' \mathbf{p} = 0$.

4. Performance comparison of the algorithms

The optimization algorithms presented in the previous section are herein applied to the frictionless normal contact problem between a numerically generated pre-fractal rough surface and a

half-plane, in order to compare their performance in terms of number of iterations required to achieve convergence and computation time.

The random midpoint displacement algorithm (Peitgen and Saupe, 1988) is used to generate the synthetic height field of surfaces with multiscale fractal roughness, i.e., with a power spectral density (PSD) function of the height field of power-law type. The surface with a given resolution (pre-fractal) is realized by a successive refinement of an initial coarse representation by adding a sequence of intermediate heights whose elevation is extracted from a Gaussian distribution with a suitable rescaled variance, see a qualitative sketch in Fig. 3. Several applications of the method to model rough surfaces for contact mechanics simulations are available in (Zavarise et al., 2004, 2007; Paggi and Ciavarella, 2010).

In particular, we consider a test problem consisting of a surface with Hurst exponent $H = 0.7$, lateral size $L = 100 \mu\text{m}$ and 512 heights per side, which corresponds to the highest discretization used to sample real surfaces with a confocal profilometer, like the Leica DCM3D available at the Multi-scale Analysis of Materials (MUSAM) Laboratory of IMT Lucca, Italy. Similar discretizations are obtained in case of AFM. The surface is brought into contact with an elastic half-plane under displacement control. Ten displacement steps are imposed to reach a maximum far-field displacement which is set equal to $(\xi_{\text{max}} - \xi_{\text{ave}})/2$, where ξ_{max} and ξ_{ave} are the maximum and the average elevations of the rough surface, respectively. All the simulations are carried out with the server 653745-421 Proliant DL585R07 from Hewlett Packard with 128 GB Ram, 4 processors AMD Opteron 6282 SE 2.60 GHz with 16 cores running MATLAB R2014b.

The parameters for the Greedy CG method are the maximum number of iterations $K_{\text{max}} = 1 \times 10^5$ and the convergence tolerance $\epsilon = 1 \times 10^{-8}$. The contact forces are initialized at zero (cold start). The constrained CG method also considers $K_{\text{max}} = 1 \times 10^5$ and the same tolerance $\epsilon = 1 \times 10^{-8}$. Both the original version by Polonsky and Keer (1999) (labeled P&K1999 in Fig. 4) and its warm-started variant (labeled P&K1999 + warm start in Fig. 4) are considered.

For the NNLS algorithm (Algorithm 3) we adopt the warm start strategy based on the projection of contact forces from the solution corresponding to a previous displacement step. Alternatively, for NNLS + GP, 100 gradient projections are used to initialize vector \mathbf{p} . For the ADMM method we use $\alpha = 1.5$, $\rho = 1$, $K_{\text{max}} = 3 \times 10^3$ and $\epsilon = 10^{-8}$. The total number n of optimization variables is varying with Δ and therefore with the force level. For the highest indentation we have $n = 35555$. Warm starting the algorithm is achieved by projecting primal variables as for the NNLS and dual variables \mathbf{w} as well. The projection simply consists of assigning the values of p_{ij}^* and w^* of the boundary elements in contact for the step Δ_k to the same boundary elements belonging to the trial contact domain I_c corresponding to the higher indentation Δ_{k+1} .

Once convergence is achieved for each imposed far-field displacement, the optimization algorithms provide the same normal

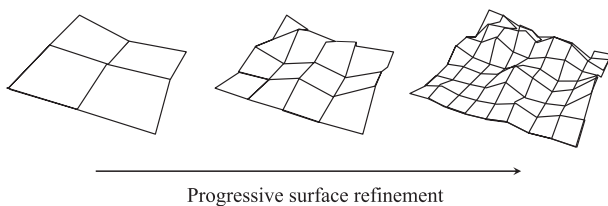


Fig. 3. Rough surfaces with multi-scale roughness and different resolution, numerically generated by the random midpoint displacement algorithm.

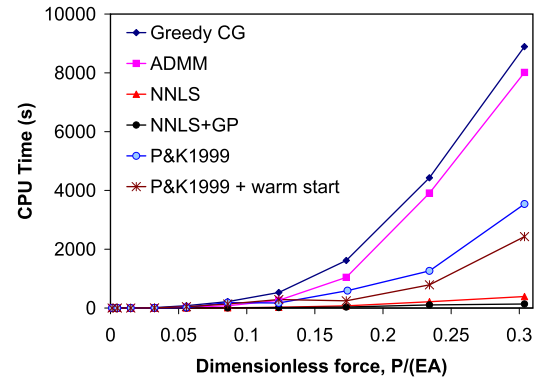


Fig. 4. Comparison between the optimization algorithms in terms of computation time.

force P and contact domains, with small roundoff errors due to finite machine precision. The CPU time required by each method to achieve convergence are shown in Fig. 4 vs. the dimensionless normal force $P/(EA)$, where E is the Young's modulus and $A = L^2$ is the nominal contact area. The best performance is achieved by the application of the NNLS method with 100 gradient projections (GP), which is 26 times faster than the original constrained CG method by Polonsky and Keer (1999) and about two orders of magnitude faster than the ADMM and the Greedy CG algorithms.

As outlined in the introduction, the Greedy method can be used in conjunction with other algorithms for solving the unconstrained linear system of equations (Step 2.4) than the CG algorithm. Although an extensive comparison of different solvers of linear systems of equations with positive definite matrices is outside the scope of this paper, we tested the Greedy algorithm after replacing the CG Step 2.4 with the optimized built-in `mldivide` function of MATLAB, or with the Gauss–Seidel algorithm, as proposed in (Borri-Brunetto et al., 1999, 2001).

The MATLAB's `mldivide` solver (which employs the Cholesky factorization) leads to a reduction of computation time of 30–40%, almost regardless of the size of the system n , see Fig. 5. Even with this gain in computation speed, the overall performance is still quite far from that of the NNLS Algorithm 3 on the platform used for the tests. Moreover, the MATLAB solver leads to an error of lack of memory for $n > 20000$, a serious problem for large systems that is not suffered by the CG solver described in Step 2.4 of Algorithm 1. The Gauss–Seidel algorithm does not suffer for the lack of memory but it is about 3 times slower than the CG method.

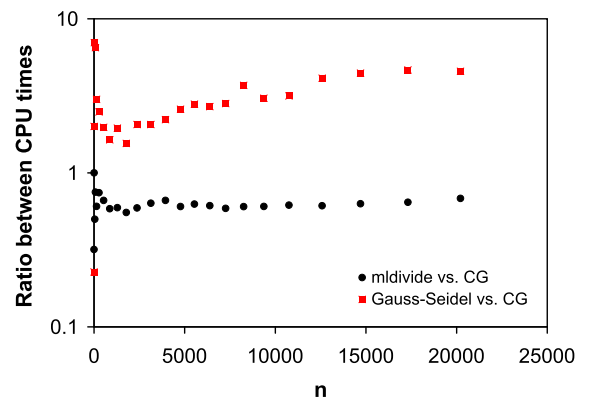


Fig. 5. Computation times for the Greedy method for different sizes n of the contact superset I_c : CG vs. MATLAB's `mldivide` solver and CG vs. Gauss–Seidel algorithm.

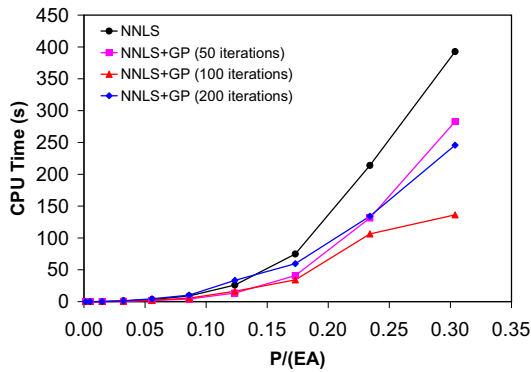


Fig. 6. Computation times of the NNLS algorithm depending on the number K of gradient projection (GP) iterations.

The effect of the number K of GP iterations applied before the NNLS algorithm is investigated in Fig. 6 for the same test problem whose results were shown in Fig. 4. By increasing K from 0 to 100 we observe a reduction in the total computation time due to a decrease in the number of iterations requested by the NNLS algorithm to achieve convergence thanks to a better initial guess of \mathbf{p} . However, a further increase in K (see, e.g., the blue curve in Fig. 6 corresponding to $K = 200$ iterations) does not correspond to further savings of CPU time. This is due to the fact that the number of NNLS iterations was already reduced to its minimum for $K = 100$ GP iterations, so that the application of further gradient projections are just leading to additional CPU time without further benefit.

5. Cascade multi-resolution (CMR) method

5.1. Algorithm

A further speed-up of computation time, as compared to the NNLS method, can be achieved by improving the criterion for the guess of the initial set I_c of points in contact. The standard criterion based on checking the interpenetration of the surface heights into the half-plane in case of a rigid body motion is the most conservative.

However, at convergence, only a small subset I_c^* of that initial set is actually in contact. Therefore, a better choice of the initial trial contact domain would reduce the size of the system of linear equations with an expected gain in terms of computation time.

As shown in (Borri-Brunetto et al., 1999) via numerical simulations on pre-fractal surfaces with Hurst exponent $H > 0.5$ and different resolution, by refining the surface height field via a recursive application of the random midpoint displacement algorithm the real contact area of each surface representation decreases by reducing δ , as illustrated in the sketch in Fig. 7. In the fractal limit of $\delta \rightarrow 0$, the real contact area vanishes. Therefore, this property of lacunarity implies that the heights that are not in contact for a coarser surface representation are not expected to come into contact by a successive refining of the height field, for the same imposed far-field displacement.

Therefore, as a better criterion, the initial trial contact domain can be selected by retaining, among all the heights selected by the rigid body interpenetration check, only those located within the areas of influence of the nodes belonging to the contact domain of a coarser representation of the rough surface for the same imposed displacement Δ .

As graphically shown in Fig. 7, an area of influence of a given node in contact can be defined by the radius $\sqrt{2}\delta$, where δ is the grid size of the coarser surface representation. Since the criterion is not exact, it is convenient to consider a multiplicative factor h larger than unity for the radius defining the nodal area of influence. It is remarkable to note that this numerical scheme can be applied recursively to a cascade of coarser representations of the same rough surface. As a general trend, computation time is expected to drastically diminish by increasing the number of cascade projections. However, the propagation of errors due to the wrong exclusion of heights that would actually make contact cannot be controlled by the algorithm and it is expected to increase with the number of projections as well. The advantage of the method is represented by the fact that, in addition to saving computation time with respect to that required by the NNLS algorithm to solve just the contact problem for the finest surface, all the contact predictions for the coarser scale representations of the same surface will be available for free, which is a useful result for the multi-scale characterization of contact problems. Moreover, the

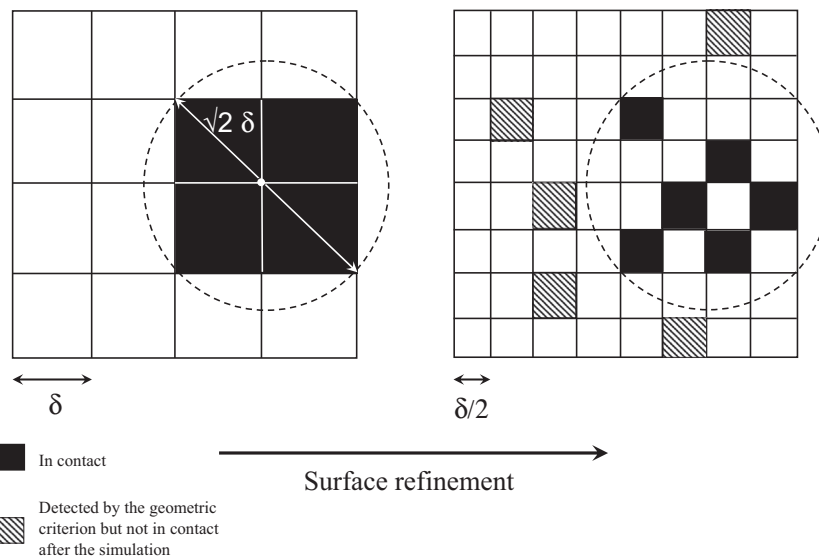


Fig. 7. A sketch illustrating the property of lacunarity of the contact domain: the real contact area progressively diminishes by refining the surface, until vanishes in the fractal limit of $\delta \rightarrow 0$. This implies that some boundary elements detected by the rigid-body interpenetration criterion (dashed gray elements) can be neglected a priori since they are outside the real contact area corresponding to the coarse scale contact solution.

CMR method can be used in conjunction with any of the optimization algorithms presented in the previous sections.

The algorithm is illustrated in Algorithm (6).

Algorithm 6. Cascade multi-resolution (CMR) algorithm

Input: $s = 1, \dots, l$ surface representations with different resolution or grid spacing $\delta(s)$; area of influence parameter $h \geq 1$.

(1) **for** $s = 1, \dots, l$ **do:**
 (1.1) **Determine**
 $I_C(s) = \{(i, j) \in I_N(s) : \zeta_{ij} \geq \zeta_{\max}(s) - \Delta\}$;
 (1.2) **if** $s = 1$ **then** $I_{C,p}(s) = I_C(s)$
else $I_{C,p}(s) = \{(i, j) \in I_C(s) : r_{i-kj-l} = \|\mathbf{x}_{ij} - \mathbf{x}_{k,l}\| \leq h\delta(s-1)\}$,
 $\forall (k, l) \in I_C^*(s-1)$
end
 (2) Construct \mathbf{H} based on the projected trial contact domain $I_{C,p}(s)$;
 (3) Apply optimization algorithms (e.g., NNLS) and determine \mathbf{p}^* , \mathbf{u}^* , $I_C^*(s)$;
 (4) **end.**

5.2. Validation in case of numerically generated and real rough surfaces

To assess the computational performance of the approach described in Section 5.1, the CMD method is applied in conjunction with the NNLS algorithm to pre-fractal surfaces with different H numerically generated by the RMD method. As an example, the lateral size is $100 \mu\text{m}$ for all the surfaces and the finest resolution whose contact response has to be sought corresponds to 256 heights per side. The method requires the storage of the coarser representations of such surfaces that are in any case available by the RMD algorithm during its various steps of random addition.

We apply the cascade of projections starting by a coarser representation of the surfaces with only 16 heights per side and then

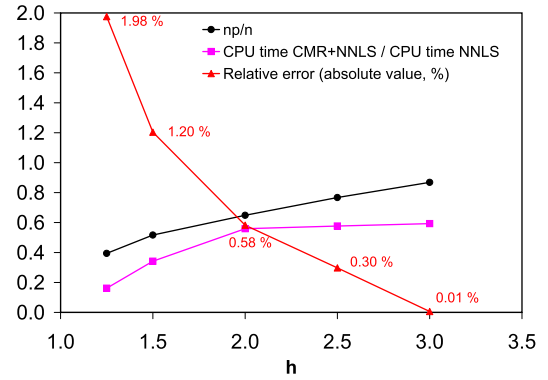


Fig. 9. Performance of the CMR+NNLS method with respect to NNLS for a numerically generated fractal surface with $H = 0.7$, depending on the parameter h .

considering 32, 64, 128 and finally 256 heights per side. A parameter $h = 2$ has been used for the definition of the area of influence. The solution of the contact problem for the surface with 16 heights per side is obtained in an exact form since it is the starting point of the cascade, whereas the contact predictions for the finer surface representations can be affected by an error intrinsic in the criterion. The approximate predictions for the surface with 256 heights per side are compared with the reference solution corresponding to the application of the NNLS algorithm with warm start directly to the finest representation of the rough surface.

The computation time of the CMR + NNLS solution is the sum of the CPU time required to solve all the coarser surface representations and it is found to be much less than the CPU time required by the NNLS algorithm to solve just one single surface with the finest resolution, see Fig. 8, where we observe a reduction of 50% in CPU time almost regardless of H . The relative error in the computation of the maximum normal force between the predicted solution and the reference one is a rapid decreasing function of H , as shown in Fig. 8(d). Considering that real surfaces have often a Hurst exponent $H > 0.5$, this is very promising.

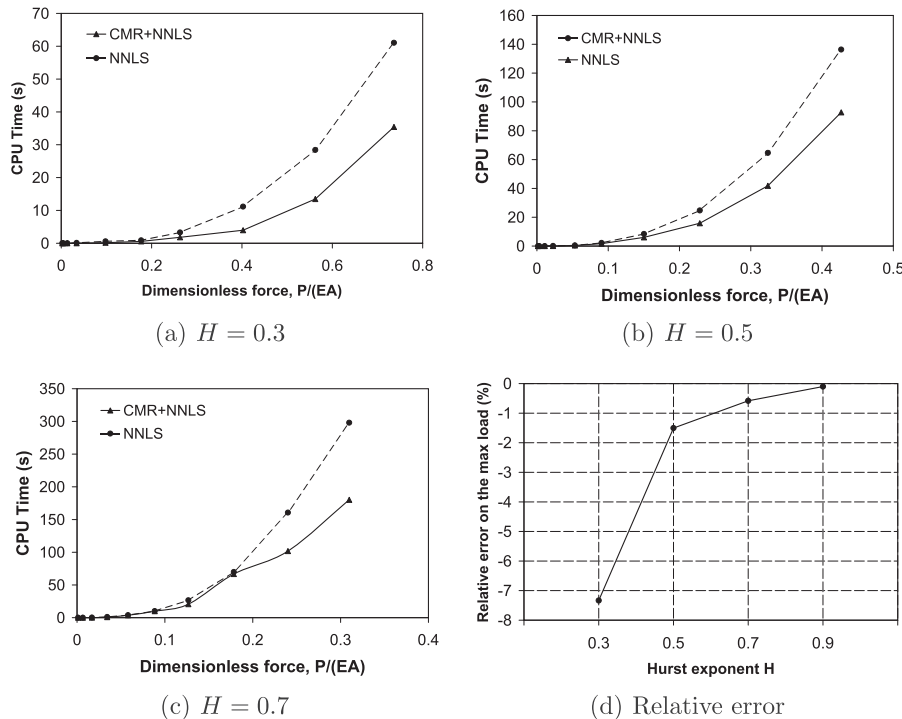


Fig. 8. Performance of the CMR + NNLS method applied to numerically generated fractal surfaces with different Hurst exponent H , $h = 2$.

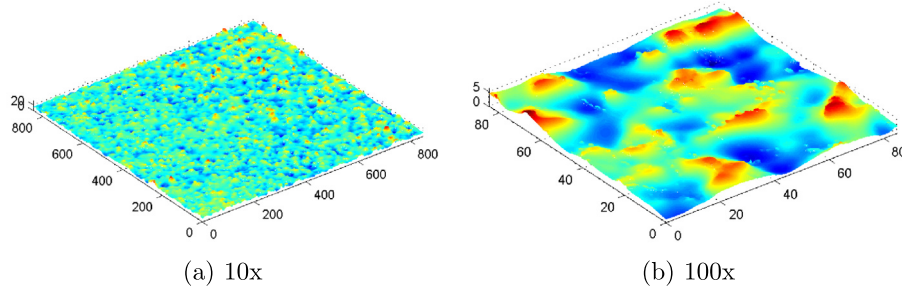


Fig. 10. Surface of textured silicon solar cells sampled with a confocal profilometer at two different magnifications (10 \times and 100 \times) obtained by using two different lenses.

A synthetic diagram illustrating the effect of the parameter h for the surface with $H = 0.7$ and for a single imposed displacement corresponding to the maximum load in Fig. 8(c) is shown in Fig. 9. The relative error is rapidly decreasing to values less than 1% by increasing h . The ratio between the number of points expected to be in contact after the application of the CMR projection criterion, n_p , and the number of points that would be included by using the classic rigid-body interpenetration check, n , is ranging from 0.4 to 0.8 by increasing h from 1.25 to 3.0. The ratio between CPU times, on the other hand, tends to an asymptotic value of 0.6, which implies a saving of 40% of computation time as compared to the exact solution, with less than 0.01% of relative error.

We also check the CMR method for warm starting on real surfaces not displaying the ideal fractal scaling at any length scale, to better assess possible limits of applicability. As a practical example we consider the surface of textured silicon solar cells sampled with two different lenses in order to achieve two different magnifications (10 \times and 100 \times) by using the confocal profilometer Leica DCM3D, see Fig. 10. The PSD function of such a surface sampled with 512 points per side presents a power-law trend for high frequencies (fine resolutions) and a cut-off to the power-law at low frequencies (coarse resolutions). In the power-law regime the surface is characterized by a Hurst exponent $H \cong 0.6$ that can be determined by the slope of the PSD function as customary.

As a main difference with respect to pre-fractal rough surfaces generated by the RMD algorithm, the application of the CMR method requires a filter to downsample the acquired surfaces and extract their coarser representations. The CMR method is applied to the two surfaces acquired with 10 \times and 100 \times magnifications using $h = 1.5$ and considering a cascade of projections involving coarser representations of the finest surfaces with 64 and 128 heights per side. A single contact step corresponding to

an imposed far-field normal displacement equal to $(\xi_{\max} - \xi_{\text{ave}})/5$ is examined.

The application of the CMR + NNLS method to the surface acquired at 100 \times leads to very good results in line with those observed for ideal fractal surfaces. The relative error in the prediction of the normal load is -0.4% , with a saving of CPU time of 18% as compared to the direct application of the NNLS algorithm. On the other hand, the method applied to the surface acquired at 10 \times leads to poor results in terms of accuracy with -98% of relative error and almost no saving in computation time. This bad performance is due to the fact that the property of lacunarity of the contact domain, strictly connected with the self-affine scaling of roughness due to fractality, does not hold anymore for the surface sampled at 10 \times due to the cut-off to its power-law PSD (see Fig. 11). As a consequence, the CMR method erroneously excludes many possible points from the initial contact domain suggested by the rigid body interpenetration check that are actually relevant for contact. Therefore, in conclusion, the CMR method is efficient for warm starting the NNLS algorithm, but it should be strictly applied to numerically generated or real rough surfaces provided that the self-affine properties of roughness are confirmed by a PSD function of power-law type.

6. Conclusion

This paper has shown how the problem of frictionless normal contact between rough surfaces within the BEM framework can be solved very efficiently by exploiting ideas from convex quadratic programming. A series of efficient optimization algorithms has been proposed and compared with the traditional Greedy method and constrained CG algorithm. As the lack of convergence of the Greedy method seems to be a rare phenomenon, it remains an open question to establish the conditions on \mathbf{H} and $\bar{\mathbf{u}}$ for which the algorithm is guaranteed to converge.

The NNLS algorithm warm started by accelerated gradient projections was shown at least two orders of magnitude faster than the Greedy method and 26 times faster than the original constrained CG algorithm.

Finally, we explored another method for warm starting the optimization algorithms, this time focusing on a selective reduction of the size of the initial trial contact domain based on the multi-resolution properties of roughness. The resulting cascade multi-resolution (CMR) method allows a further saving of about 50% of CPU time as compared to NNLS for contact simulations involving numerically generated fractal surfaces. Relative errors were found less than 2% for surfaces with $H > 0.5$, by using $h = 2$, that was found a good compromise between accuracy and computation time. Moreover, it has to be remarked that not only the solution of the finest contact problem is gained by the CMR + NNLS method with much less CPU time, but also the contact problems involving all the coarser representations of the finest surface. These results are particularly important for speeding up

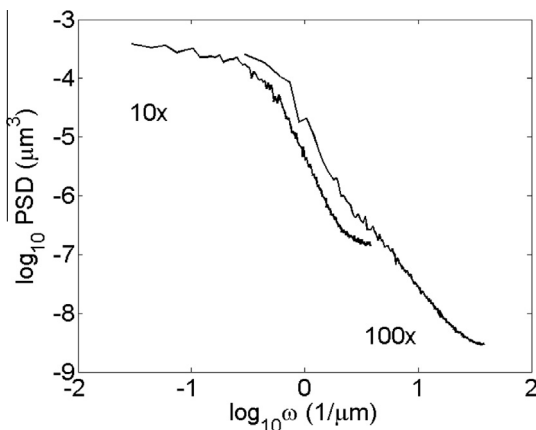


Fig. 11. Power spectral density function (PSD) of the two sampled rough surfaces shown in Fig. 10.

intensive Monte Carlo simulations involving a sequence of contact simulations for a population of fractal surfaces with different resolution. So far, to the best of the authors' knowledge, such extensive simulations, that are important to determine more reliable trends from the statistical point of view, have been limited to populations of 20 to 50 randomly generated surfaces.

In case of real surfaces, a very good performance (less than 2% of error with 3 cascades and at least 18% of CPU time saved for one single imposed displacement step) has been demonstrated in case of power-law PSDs, assuring the self-affine scaling of roughness which represents the main underlying assumption for the algorithm applicability. For surfaces with a cut-off to the power-law PSD, on the other hand, the CMR + NNLS method has given poor results in terms of accuracy and in any case almost no saving in CPU time as compared to the pure application of NNLS. Therefore, this warm start method should be used with care and only in a range where the PSD is of power-law type.

Finally, we point out that the proposed optimization methods can also be applied to frictional contact problems by using for instance the complete BEM formulation as in (Pohrt and Li, 2014). Although this issue is left for further investigation, we expect an even more significant gain in CPU time by applying the algorithms presented in this paper instead of other optimization methods, since the size of the problem is by far significantly increased as compared to the frictionless case.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement n. 306622 (ERC Starting Grant "Multi-field and multi-scale Computational Approach to Design and Durability of PhotoVoltaic Modules" – CA2PVM; PI: Prof. M. Paggi).

References

- Andersson, T., 1981. The boundary element method applied to two-dimensional contact problems with friction. *Boundary Elem. Methods* 3, 239–258.
- Barber, J.R., 1974. Determining the contact area in elastic-indentation problems. *J. Strain Anal.* 9, 230–232.
- Barber, J.R., 2003. Bounds on the electrical resistance between contacting elastic rough bodies. *Proc. R. Soc. London Ser. A* 459, 53–66.
- Barber, J.R., 2010. *Elasticity*, third ed. Springer, Dordrecht.
- Batrouni, G.G., Hansen, A., Schmittbuhl, J., 2002. Elastic response of rough surfaces in partial contact. *Europhys. Letters* 60, 724–730.
- Beale, E.M.L., 1955. On minimizing a convex function subject to linear inequalities. *J. R. Stat. Soc. Ser. B*, 173–184.
- Bemporad, A., A quadratic programming algorithm based on nonnegative least squares with applications to embedded model predictive control, *IEEE Trans. Autom. Control*. Conditionally accepted for publication.
- Borri-Brunetto, M., Carpinteri, A., Chiaia, B., 1999. Scaling phenomena due to fractal contact in concrete and rock fractures. *Int. J. Fract.* 95, 221–238.
- Borri-Brunetto, M., Chiaia, B., Ciavarella, M., 2001. Incipient sliding of rough surfaces in contact: a multiscale numerical analysis. *Comput. Methods Appl. Mech. Eng.* 190, 6053–6073.
- Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* 3, 1–122.
- Campana, C., Persson, B.N.J., Müser, M.H., 2001. Transverse and normal interfacial stiffness of solids with randomly rough surfaces. *J. Phys.: Condens. Matter* 23, 085001.
- Carbone, G., Bottiglione, F., 2008. Asperity contact theories: do they predict linearity between contact area and load? *J. Mech. Phys. Solids* 56, 2555–2572.
- Carbone, G., Putignano, C., 2013. A novel methodology to predict sliding and rolling friction of viscoelastic materials: theory and experiments. *J. Mech. Phys. Solids* 61, 1822–1834.
- Ciavarella, M., Delfino, V., Demelio, G., 2006. A re-vitalized Greenwood & Williamson model of elastic contact between fractal surfaces. *J. Mech. Phys. Solids* 54, 2569–2591.
- Ciavarella, M., Dibello, S., Demelio, G., 2008a. Conductance of rough random profiles. *Int. J. Solids Struct.* 45, 879–893.
- Ciavarella, M., Greenwood, J.A., Paggi, M., 2008b. Inclusion of interaction in the Greenwood and Williamson contact theory. *Wear* 265, 729–734.
- Cottle, R.W., Pang, J.-S., Stone, R.E., 1992. *The Linear Complementarity Problem*. Academic Press.
- Fletcher, R., 1971. A general quadratic programming algorithm. *IMA J. Appl. Math.* 7, 76–91.
- Francis, H.A., 1983. The accuracy of plane strain models for the elastic contact of three-dimensional rough surfaces. *Wear* 85, 239–256.
- Goldfarb, D., Idnani, A., 1983. A numerically stable dual method for solving strictly convex quadratic programs. *Math. Program.* 27 (1), 1–33.
- Hestenes, M.R., 1980. *Conjugate Direction Methods in Optimization*. Springer, New York, Ch. 2 and 3.
- Hendriks, C.P., Visscher, M., 1995. Accurate real area of contact measurements on polyurethane. *ASME J. Tribol.* 117, 607–611.
- Hyun, S., Pei, L., Molinari, J.-F., Robbins, M.O., 2004. Finite-element analysis of contact between elastic self-affine surfaces. *Phys. Rev. E* 70, 026117.
- Hyun, S., Robbins, M.O., 2007. Elastic contact between rough surfaces: effect of roughness at large and small wavelengths. *Tribol. Int.* 40, 1413.
- Johnson, K.L., 1985. *Contact Mechanics*. Cambridge University Press, Cambridge, UK.
- Kalker, J.J., van Randen, Y.A., 1972. A minimum principle for frictionless elastic contact with application to non Hertzian problems. *J. Eng. Math.* 6, 193–206.
- Karpenko, Y.A., Akay, A., 2001. A numerical model of friction between rough surfaces. *Tribol. Int.* 34, 531–545.
- Kubo, A., Okamoto, T., Kurokawa, N., 1981. Contact stress between rollers with surface irregularity. *J. Tribol.* 116, 492–498.
- Lawson, C., Hanson, R., 1974. Solving least squares problems. *SIAM* 161, Ch. 24.
- Li, J., Berger, E.J., 2003. A semi-analytical approach to three-dimensional normal contact problems with friction. *Comput. Mech.* 30, 310–322.
- Man, K.W., 1994. *Contact Mechanics Using Boundary Elements*, Topics in Engineering, vol. 22. Computational Mechanics Publications, Southampton, Boston.
- McCool, J.I., 1986. Comparison of models for the contact of rough surfaces. *Wear* 107, 37–60.
- Nesterov, Y., 1983. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Sov. Math. Dokl.* 27, 372–376.
- Nogi, T., Kato, T., 1998. Influence of a hard surface layer on the limit of elastic contact – Part I: analysis using a real surface model. *J. Tribol.* 110, 493–500.
- Nosonovsky, M., Bhushan, B., 2005. Roughness optimization for biomimetic superhydrophobic surfaces. *Microsyst. Technol.* 11, 535–549.
- O'Callaghan, P.W., Probert, S.D., 1970. Real area of contact between a rough surface and a softer optically flat surface. *J. Mech. Eng. Sci.* 11, 259–267.
- Paggi, M., Barber, J.R., 2011. Contact conductance of rough surfaces composed of modified RMD patches. *Int. J. Heat Mass Transfer* 54, 4664–4672.
- Paggi, M., Ciavarella, M., 2010. The coefficient of proportionality κ between real contact area and load, with new asperity models. *Wear* 268, 1020–1029.
- Paggi, M., Pohrt, R., Popov, V.L., 2014. Partial-slip frictional response of rough surfaces. *Sci. R.* 4, 5178.
- Patrinos, P., Bemporad, A., 2014. An accelerated dual gradient-projection algorithm for embedded linear model predictive control. *IEEE Trans. Autom. Control* 59, 18–33.
- Peitgen, H.O., Saupe, D., 1988. *The Science of Fractal Images*. Springer-Verlag, New York.
- Pohrt, R., Li, Q., 2014. Complete boundary element formulation for normal and tangential contact problems. *Phys. Mesomech.* 17, 334–340.
- Polonsky, I.A., Keer, L.M., 1999. A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques. *Wear* 231, 206–219.
- Polonsky, I.A., Keer, L.M., 2000a. A fast and accurate method for numerical analysis of elastic layered contacts. *J. Tribol.* 122, 30–35.
- Polonsky, I.A., Keer, L.M., 2000b. Fast methods for solving rough contact problems: a comparative study. *J. Tribol.* 122, 36–41.
- Prodanov, N., Dapp, W.B., Müser, M.H., 2014. On the contact area and mean gap of rough, elastic contacts: dimensional analysis, numerical corrections and reference data. *Tribol. Lett.* 53, 433–448.
- Raous, M., 1999. Quasistatic Signorini problem with Coulomb friction and coupling to adhesion, In: *New Developments in Contact Problems*, International Centre for Mechanical Sciences, 384, pp. 101–178.
- Scaraggi, M., Putignano, C., Carbone, G., 2013. Elastic contact of rough surfaces: a simple criterion to make 2D isotropic roughness equivalent to 1D one. *Wear* 297, 811–817.
- Schmid, C., Biegler, L.T., 1994. Quadratic programming methods for reduced hessian SQP. *Comput. Chem. Eng.* 18 (9), 817–832.
- Sridhar, M.R., Yovanovich, M.M., 1994. Review of elastic and plastic contact conductance models: comparison with experiments. *J. Thermophys. Heat Transfer* 8, 633–640.
- Wriggers, P., Reinelt, J., 2009. Multi-scale approach for frictional contact of elastomers on rough rigid surfaces. *Comput. Methods Appl. Mech. Eng.* 191, 1333–1348.
- Wriggers, P., 2006. *Computational Contact Mechanics*, second ed. Springer Verlag, Berlin, Heidelberg.
- Yastrebov, V.A., Anciaux, G., Molinari, J.-F., 2015. From infinitesimal to full contact between rough surfaces: evolution of the contact area. *Int. J. Solids Struct.* 52, 83–102.
- Zavarise, G., Borri-Brunetto, M., Paggi, M., 2004. On the reliability of microscopical contact models. *Wear* 257, 229–245.
- Zavarise, G., Borri-Brunetto, M., Paggi, M., 2007. On the resolution dependence of micromechanical contact models. *Wear* 262, 42–54.