INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING
Int. J. Numer. Meth. Engng. **42**, 145–173 (1998)

FORMULATION AND COMPARISON OF ALGORITHMS FOR FRICTIONAL CONTACT PROBLEMS

P. W. CHRISTENSEN[1], A. KLARBRING[1,*], J. S. PANG[2] AND N. STRÖMBERG[1]

[1] *Department of Mechanical Engineering, Linköping University, S-581 83 Linköping, Sweden*
[2] *Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218, U.S.A.*

ABSTRACT

This paper presents two algorithms for solving the discrete, quasi-static, small-displacement, linear elastic, contact problem with Coulomb friction. The algorithms are adoptions of a Newton method for solving B-differentiable equations and an interior point method for solving smooth, constrained equations. For the application of the former method, the contact problem is formulated as a system of B-differentiable equations involving the projection operator onto sets with simple structure; for the application of the latter method, the contact problem is formulated as a system of smooth equations involving complementarity conditions and with the non-negativity of variables treated as constraints. The two algorithms are numerically tested for two-dimensional problems containing up to 100 contact nodes and up to 100 time increments. Results show that at the present stage of development, the Newton method is superior both in robustness and speed. Additional comparison is made with a commercial finite element code. © 1998 John Wiley & Sons, Ltd.

KEY WORDS: contact; friction; complementarity; Newton's method; interior point method; finite elements

1. INTRODUCTION

Problems involving contact and friction are among the most difficult ones in mechanics and at the same time of crucial practical importance in many engineering branches. The chief mathematical difficulty lies in the severe contact non-linearities which are of such a nature that the natural first-order constitutive laws of contact and friction phenomena are expressed by non-smooth multivalued force–displacement or force–velocity relations. The present paper discusses the numerical solution of these constitutive laws when used in conjunction with a model of a discrete linear elastic structure. The problem may be seen as obtained by a formal finite element discretization of a continuous linear elastic body in contact with a frictional foundation. Such a formal finite element discretization is presented in the related paper,[1] where some new existence results for the semi-coercive version of the problem were established.

* Correspondence to A. Klarbring, Division of Mechanics, Department of Mechanical Engineering, S-581 83 Linköping, Sweden. E-mail: andkl@ikp.liu.se

Contract/grant sponsor: CENIIT
Contract/grant sponsor: ABB Atom AB
Contract/grant sponsor: TFR
Contract/grant sponsor: National Science Foundation; Contract/grant number: CCR-9213739
Contract/grant sponsor: Office of Naval Research; Contract/grant number: N00014-93-1-0228

CCC 0029–5981/98/010145–29$17.50

Received 27 November 1996
Revised 14 May 1997

Due to the practical importance of frictional contact problems, a large number of algorithms for the numerical solution of the related finite element equations and inequalities have been presented in the literature. Review papers that may be consulted for an extensive list of references are References 2–4. See also the monographs by Kikuchi and Oden[5] and Zhong.[6] However, very few of these algorithms are supported by any convergence theory and it should be realized that since quasistatic friction problems are known to exhibit 'catastrophic' features such as non-existence and non-uniqueness of solutions, see Reference 7, there is a substantial need for such results. Indeed, users of commercial software frequently report numerical difficulties when attempting to solve friction problems.

A large family of numerical methods that are based on sound mathematical principles and easily amenable to rigorous analysis are the so-called Mathematical Programming (MP) methods; see the review paper, Reference 3. These methods are based on reformulating the problem as a previously recognized mathematical problem such as a Linear Complementarity Problem or a finite-dimensional Variational Inequality. Contact problems with friction are then entered as an application of the field of MP and proven algorithms become available. The present paper is mainly a continuation of this idea and takes into account some of the latest developments in the MP field. In addition to being mathematically sound and rigorous, the MP methods are practically very effective for solving frictional contact problems of a realistic nature, as can be seen from the computational results reported in this and several accompanying papers.[8–10] Furthermore, when applied to problems which the 'trial-and-error' commercial codes can successfully solve (such as the frictionless problems), the MP methods are highly competitive (a fact confirmed by the algorithms described in this paper). Further discussion on our experience in comparing with the commercial code ABAQUS is given in the concluding remarks.

Two solution methods are presented, analysed, implemented and compared in this work. The first method is related to a so-called augmented Lagrangian formulation of the contact problem with friction, as was independently proposed by Alart and Curnier,[11] Simo and Laursen[12] and De Saxce and Feng.[13] See also References 9 and 14–16. This formulation has the same property as a penalty formulation in that it reformulates the contact and friction conditions as a system of equations in contrast to inequalities, but contrary to the penalty method it involves no approximation of the original problem and is not subject to penalty sensitivity. However, a drawback (actually shared by most penalty methods) is that the equation system is non-differentiable, and therefore, the classical Newton method for smooth equations fails to be applicable. The present development takes this fact as a point of departure and observes that the non-differentiable equation system in question is B-differentiable. It then follows, as proved by Pang,[17] that the systems can be solved by an extended Newton method which involves solving in each iteration a possibly non-smooth system and performing a line search which makes the method globally convergent. Furthermore, we suggest how to simplify the formulation of Alart and Curnier in such a way that the system is still B-differentiable. This approach is analysed and implemented in the present paper.

The second method treated in the paper is an interior point method proposed for dealing with complementarity systems in Reference 18. Interior point methods have a well-documented effectiveness for solving linear programs and monotone complementarity problems. Since frictional contact problems may be seen as a particular type of complementarity problem, we have chosen to investigate the application of a method of this type. It appears that this is the first time such a method is applied to deformable-body contact problems. In Reference 19 the method in Reference 18 was applied to a rigid-body contact problem. As in Reference 19, the method given here is based on a reformulation of the friction problem as a system of smooth equations involving

complementarity conditions and with the non-negativity of variables treated as constraints of the equations. The formulation is dealt with in full detail for the two-dimensional case. Although a similar treatment can be given for the three-dimensional case, our preliminary experience with the application of the interior point method to the resulting formulation suggests that this approach is not promising. Thus, a different formulation is needed for solving the three-dimensional problems. The latter subject has been independently investigated and our findings will be published separately.[8]

The rest of the paper is organized as follows. In the next section, the quasi-static, linear elastic, small-displacement, contact problem with Coulomb friction is formulated. This is done in a spatially discrete setting. A backward Euler time discretization is introduced to produce the time incremental problems, which are solved by the two methods described in subsequent sections. In the numerical examples, we solve a sequence of these incremental problems to produce physically interesting solutions. Section 3 presents the B-differentiable Newton method and the interior point method and summarize some theoretical properties of these methods. In Section 4, we reformulate the frictional contact problem in a way such that the B-differentiable Newton method becomes applicable. In Section 5, we reformulate this problem such that the interior point method becomes applicable. Section 6 presents the numerical examples. In Section 7, we give conclusions concerning the performance of the two methods.

2. SETTING OF THE PROBLEM

We consider a discrete or discretized frictional contact problem associated with a mechanical structure such as a finite element discretized elastic body or a naturally discrete structure such as a truss. In the numerical examples in Section 6, a finite element discretized plane linear elastic body is considered. We refer to, e.g. Reference 5 for the mathematical relation between the continuous problem and the discrete problem.

The basic problem is time evolutionary. In what follows, the given inputs, the unknown variables and the defining equations of the problem are all stated with reference to a fixed but arbitrary time instant.

The configurations of the mechanical structure will be represented by a vector $u \in \Re^{n_u}$ of displacements and the forces acting on it by a vector $f \in \Re^{n_u}$. A positive semi-definite stiffness matrix $K \in \Re^{n_u \times n_u}$ exists such that

$$f = Ku \tag{1}$$

The structure is situated in a three-dimensional physical space where the components of u are associated with the displacements of particular points of the structure, denoted as nodes. Assuming that there are no partially constrained nodes, there will be a total of $n_u/3$ such nodal points. Among these nodal points is a subset of dimension n_c that contains the nodes where contact can occur (called contact nodes). At each contact node, we visualize a rigid obstacle and associate with this obstacle a normal direction and two tangential directions. A vector of normal-direction-displacements and two vectors of tangential-direction-displacements are then obtained from u by means of kinematic transformation matrices C_n, C_t and C_o each belonging to $\Re^{n_c \times n_u}$, i.e.

$$w_n = C_n u \in \Re^{n_c}, \quad w_t = C_t u \in \Re^{n_c}, \quad w_o = C_o u \in \Re^{n_c} \tag{2}$$

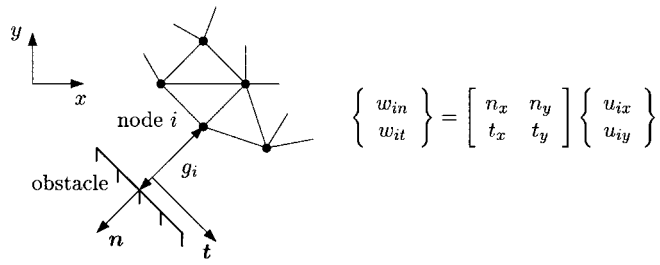

Figure 1. Illustration of the kinematic transformation matrices given in (2) for one contact node in the two-dimensional case

The subscripts n, t and o stand for normal, tangential and orthogonal, respectively. The rows of the matrices C_n, C_t and C_o are denoted $C_{i\mathrm{n}}$, $C_{i\mathrm{t}}$ and $C_{i\mathrm{o}}$, respectively, for $i = 1, \ldots, n_\mathrm{c}$. The transformations given in (2) are illustrated in Figure 1.

Next, normal and tangential contact forces $p_\mathrm{n} \in \Re^{n_\mathrm{c}}$, $p_\mathrm{t} \in \Re^{n_\mathrm{c}}$ and $p_\mathrm{o} \in \Re^{n_\mathrm{c}}$, (negatively) work conjugate to w_n, w_t and w_o, respectively, are introduced together with a vector of prescribed forces $f^{\mathrm{ext}} \in \Re^{n_u}$. It then holds that

$$f = -C_\mathrm{n}^\mathrm{T} p_\mathrm{n} - C_\mathrm{t}^\mathrm{T} p_\mathrm{t} - C_\mathrm{o}^\mathrm{T} p_\mathrm{o} + f^{\mathrm{ext}} \tag{3}$$

where the superscript T denotes the transpose of a vector or matrix. As mentioned above, the basic problem considered here is time evolutionary; i.e. even though it is not notationally made explicit, the external force f^{ext} is time dependent, making the unknown vectors u, p_n, p_t and p_o also time dependent. Furthermore, the friction law introduced below explicitly involves a time derivative. The components of the vectors p_n, p_t, p_o, w_n, w_t and w_o are denoted $p_{i\mathrm{n}}$, $p_{i\mathrm{t}}$, $p_{i\mathrm{o}}$, $w_{i\mathrm{n}}$, $w_{i\mathrm{t}}$ and $w_{i\mathrm{o}}$, respectively, for $i = 1, \ldots, n_\mathrm{c}$.

We turn attention to the contact and friction conditions. Specifically, the normal contact conditions are stated in a complementarity form

$$w_\mathrm{n} \leqslant g, \quad p_\mathrm{n} \geqslant 0, \quad p_\mathrm{n}^\mathrm{T}(w_\mathrm{n} - g) = 0 \tag{4}$$

where $g \in \Re^{n_\mathrm{c}}$ represents the initial contact gaps between contact nodes and rigid obstacles. We note that (4) is equivalent to the variational inequality:

$$p_\mathrm{n} \in \Re_+^{n_\mathrm{c}}: \quad (w_\mathrm{n} - g)^\mathrm{T}(q_\mathrm{n} - p_\mathrm{n}) \leqslant 0, \ \ \forall q_\mathrm{n} \in \Re_+^{n_\mathrm{c}} \tag{5}$$

where $\Re_+^{n_\mathrm{c}}$ is the non-negative orthant in $\Re^{n_\mathrm{c}}$ and q_n is a 'trial' contact force.

As a condition relating to the tangential forces and displacements, we take Coulomb's law of friction which we state for each contact node. A convenient way of expressing this law for an arbitrary $i \in \{1, \ldots, n_\mathrm{c}\}$ is through the maximum dissipation principle in the form of a variational inequality:

$$(p_{i\mathrm{t}}, p_{i\mathrm{o}}) \in \mathscr{F}(\mu_i p_{i\mathrm{n}}): \quad \dot{w}_{i\mathrm{t}}(q_{i\mathrm{t}} - p_{i\mathrm{t}}) + \dot{w}_{i\mathrm{o}}(q_{i\mathrm{o}} - p_{i\mathrm{o}}) \leqslant 0, \ \ \forall (q_{i\mathrm{t}}, q_{i\mathrm{o}}) \in \mathscr{F}(\mu_i p_{i\mathrm{n}}) \tag{6}$$

where the superposed dot denotes time derivative, and

$$\mathscr{F}(\gamma) \equiv \{(q_{i\mathrm{t}}, q_{i\mathrm{o}}) \in \Re^2: \ q_{i\mathrm{t}}^2 + q_{i\mathrm{o}}^2 \leqslant \gamma^2\} \quad \text{for } \gamma \in \Re_+$$

is a set of admissible tangential contact forces (with $\gamma = \mu_i p_{i\mathrm{n}}$) and $\mu_i \geqslant 0$ is the friction coefficient at node i. Other forms of $\mathscr{F}(\gamma)$ could also be considered, see Reference 1.

What has been formulated so far is the quasi-static contact problem with friction, i.e. for a given load history $\theta \mapsto f^{\text{ext}}(\theta)$, $\theta \in [0, \mathscr{T}]$, where $[0, \mathscr{T}]$ is a given time interval, we want to find the functions $\theta \mapsto u(\theta)$, $\theta \mapsto p_{\text{n}}(\theta)$, $\theta \mapsto p_{\text{t}}(\theta)$ and $\theta \mapsto p_{\text{o}}(\theta)$ such that for each $\theta \in [0, \mathscr{T}]$, the tuple $(u(\theta), p_{\text{n}}(\theta), p_{\text{t}}(\theta), p_{\text{o}}(\theta))$ satisfies conditions (1)–(6). There are two distinct approaches to compute these functions. One possibility was developed in References 20 and 21: it involves introducing a piecewise linearization of the friction cone $\mathscr{F}(\gamma)$; the solution of such a problem is piecewise linear in time. By reformulating the problem as a type of Linear Complementarity Problem (LCP) involving time derivatives, it is possible to develop a principle pivoting algorithm to compute the piecewise linear solution functions. Another approach, which is perhaps more obvious and historically older, is to introduce a backward Euler time discretization of the time derivatives $\dot{w}_{it}$ and $\dot{w}_{io}$. This approach is more versatile; in particular, it is applicable when non-linearities other than a non-linear friction law (like the one used here) are present, see e.g. Reference 9. We adopt the time discretization approach in this work.

Let the time interval $[0, \mathscr{T}]$ be divided into sub-intervals and let us concentrate on one such interval, $[\theta_\ell, \theta_{\ell+1}]$. At time θ_ℓ the vectors u, p_{n}, p_{t} and p_{o} are assumed to be known. We use the notation $\bar{u}$ for the known displacement vector and a similar notation for the contact displacements, i.e. $\bar{w}_{\text{n}} = C_{\text{n}}\bar{u}$, $\bar{w}_{\text{t}} = C_{\text{t}}\bar{u}$ and $\bar{w}_{\text{o}} = C_{\text{o}}\bar{u}$; we omit the bar on these variables for the unknown variables at time $\theta_{\ell+1}$. In particular, with w_{it} and w_{io} denoting the unknown tangential contact displacements at time $\theta_{\ell+1}$ and using the backward Euler time discretization, the approximation for the time derivatives become

$$\dot{w}_{it}(\theta_{\ell+1}) \approx \frac{w_{it} - \bar{w}_{it}}{\theta_{\ell+1} - \theta_\ell}, \quad \dot{w}_{io}(\theta_{\ell+1}) \approx \frac{w_{io} - \bar{w}_{io}}{\theta_{\ell+1} - \theta_\ell} \tag{7}$$

The variational inequality (6) may now be stated in a time discretized setting: find $(p_{it}, p_{io}) \in \mathscr{F}(\mu_i p_{in})$ such that

$$(w_{it} - \bar{w}_{it})(q_{it} - p_{it}) + (w_{io} - \bar{w}_{io})(q_{io} - p_{io}) \leqslant 0, \quad \forall (q_{it}, q_{io}) \in \mathscr{F}(\mu_i p_{in}) \tag{8}$$

The *discrete*, *time-incremental*, *three-dimensional*, *small-displacement*, *elastic-body frictional contact problem* is now given by (1)–(3) along with the constitutive laws (5) and (8). The numerical solution of a sequence of these problems is the main focus of this paper. The theory is developed for one particular increment, or time interval.

3. TWO SOLUTION METHODS: GENERAL DISCUSSION

Digressing from the discussion of the contact problem, we present two general solution methods for systems of non-linear equations. Although equation solving is a classical subject in applied mathematics, the methods presented herein are new entries into the field of mathematical programming. The motivation for adopting these methods for solving the contact problems has been discussed in the Introduction of this paper. Since these methods are quite new in the mechanics literature, we summarize in this section the main features of the methods and their principal convergence properties. In the next two sections, we will show how the contact problem can be cast in the required forms to which the methods become applicable and we will present some convergence results.

Let $H : \Re^n \to \Re^n$ be a given mapping from the finite-dimensional Euclidean space $\Re^n$ into itself and let Ω be a closed subset of $\Re^n$. We consider the problem of solving the system of equations

with constraints, denoted CE (H,Ω):

$$H(z)=0, \quad z\in\Omega \tag{9}$$

The two methods to be described below differ from the classical methods for solving smooth, unconstrained systems of equations, such as those described in References 22 and 23, in two major aspects. The B-differentiable Newton method, proposed in Reference 17, solves the unconstrained equation (i.e., (9) with $\Omega=\Re^n$) with H being a *Bouligand differentiable* function (as opposed to a Fréchet differentiable function in the classical case). The potential reduction interior point method, proposed in Reference 18, solves the constrained equation (9) with H being a Fréchet differentiable function and Ω a proper subset of $\Re^n$ with a non-empty interior (as opposed to Ω being $\Re^n$ in the classical case). Details of the methods follow. In the discussion below, we use H_{N} to denote the mapping associated with B-differentiable Newton method and H_{I} to denote the mapping associated with the interior-point method.

3.1. A B-differentiable Newton method

A function $H_{\mathrm{N}}:\Re^n\to\Re^n$ is said to be *B-differentiable* (B for Bouligand) at a point $z\in\Re^n$ if H_{N} is Lipschitz continuous in a neighbourhood of z and directionally differentiable at z. The directional derivative of H_{N} at z along a direction $d\in\Re^n$ is denoted by the standard notation $H'_{\mathrm{N}}(z;d)$ and also called the B-derivative of H_{N} at z along d. A B-differentiable function H_{N} at a point need not be F-differentiable there (F for Fréchet). The fundamental difference between the B-differentiability and (classical) F-differentiability is the absence of linearity in the directional derivative $H'_{\mathrm{N}}(z;d)$ in the second argument (although this derivative is, by its definition, still positively homogeneous in d). If H_{N} is F-differentiable at z with $\nabla H_{\mathrm{N}}(z)$ denoting the $n\times n$ Jacobian matrix at z, then $H'_{\mathrm{N}}(z;d)=\nabla H_{\mathrm{N}}(z)d$. The function H_{N} is said to be *B-differentiable* if it is B-differentiable at all points in its domain.

The classical Gauss–Newton method[22] for solving the smooth equation $H_{\mathrm{N}}(z)=0$ has been extended to solving a B-differentiable equation. In the classical method, H_{N} is assumed to be F-differentiable. At each iteration of the method, a system of linear equations, called the *Newton equation* and obtained by linearizing the equation at the current iterate z^k, is solved to obtain a search direction $\mathrm{d}z^k$. This direction $\mathrm{d}z^k$ provides descent for the merit function

$$\Theta(z)\equiv\tfrac{1}{2}H_{\mathrm{N}}(z)^{\mathrm{T}}H_{\mathrm{N}}(z), \quad z\in\Re^n \tag{10}$$

a one-dimensional line search (typically of 'Armijo type') is performed on Θ starting at z^k and searching along $\mathrm{d}z^k$. The next iterate z^{k+1} is then obtained as the vector $z^k+\tau_k\mathrm{d}z^k$ for a suitable step length $\tau_k>0$ that yields sufficient decrease in Θ from its current value $\Theta(z^k)$.

In essence, the extension of the classical Newton method to a B-differentiable equation consists simply of the replacement of the linear Newton equation by an equation obtained from the use of the B-derivative instead of the F-derivative. In general, the resulting 'directional Newton equation' (11) is no longer linear.

Below is a step-by-step description of the B-differentiable Newton method for solving the equation $H_{\mathrm{N}}(z)=0$. For more details of the method and its justification, see Reference 17.

Description of Algorithm BN

Step 1 (*Initialization*). Let β, σ and ε be given scalars with $\beta\in(0,1)$, $\sigma\in(0,\frac{1}{2})$ and $\varepsilon>0$ small. Set $k=0$. Let z^k be given.

Step 2 (*Direction generation*). Solve the directional Newton equation to obtain the direction $\mathrm{d}z^k$:

$$H_{\mathrm{N}}(z^k) + H'_{\mathrm{N}}(z^k; \mathrm{d}z^k) = 0 \tag{11}$$

Step 3 (*Step size determination*). Let $\tau_k \equiv \beta^{m_k}$, where m_k is the smallest non-negative integer m for which the following decrease criterion holds:

$$\Theta(z^k + \beta^m \mathrm{d}z^k) \leqslant (1 - 2\sigma\beta^m)\Theta(z^k) \tag{12}$$

Set $z^{k+1} \equiv z^k + \tau_k \mathrm{d}z^k$.

Step 4 (*Termination check*). If $\Theta(z^{k+1}) \leqslant \varepsilon$, terminate with z^{k+1} as an approximate zero of H_{N}. Otherwise, return to Step 2 with $k \leftarrow k + 1$.

The following theorem is a global convergence result for the above algorithm. It summarizes the limiting properties of an infinite sequence of iterates $\{z^k\}$ produced by the algorithm. See Reference 17 for a proof and expanded discussion of this theorem; in particular, some detailed treatment is given in the reference concerning the application of the algorithm for solving complementarity problems and variational inequalities. The concepts of the strong F-derivative and quadratic convergence used in the theorem are well known and can be found, e.g. in Reference 22.

Theorem 1. Let $H_{\mathrm{N}} : \Re^n \to \Re^n$ be a B-differentiable function. Suppose that the directional Newton equation (11) *is always solvable. Let z^∞ be an accumulation point of the sequence $\{z^k\}$ produced by the B-differentiable Newton method. The following two statements hold:*

(a) *If Θ has a strong F-derivative at z^∞ and there exist a neighbourhood $\mathcal{N}$ of z^∞ and a positive scalar c such that for all $z \in \mathcal{N}$ and all $v \in \Re^n$, $\|H'_{\mathrm{N}}(z; v)\| \geqslant c\|v\|$, then $H_{\mathrm{N}}(z^\infty) = 0$.*

(b) *If z^∞ is the limit of $\{z^k\}$, H_{N} has a strong F-derivative at z^∞, the Jacobian matrix $\nabla H_{\mathrm{N}}(z^\infty)$ is non-singular, the step size $\tau_k = 1$ for all k sufficiently large, and there exist a neighbourhood $\mathcal{N}'$ and a positive scalar c' such that for all z and z' in $\mathcal{N}'$ and all $v \in \Re^n$ with unit length,*

$$\|H'_{\mathrm{N}}(z; v) - H'_{\mathrm{N}}(z'; v)\| \leqslant c' \|z - z'\|$$

then $\{z^k\}$ converges to z^∞ Q-quadratically, i.e.

$$\limsup_{k \to \infty} \frac{\|z^{k+1} - z^\infty\|}{\|z^k - z^\infty\|^2} < \infty$$

Certain deficiencies, both theoretical and practical, exist in the above algorithm and its convergence result. In applications to complementarity problems (of which the one arising from the contact problem described in Section 2 is a special case), these deficiencies have been remedied by various improved algorithms. Indeed, there has been a flurry of recent advances in solution methods for complementarity problems, all of which are potentially applicable for solving the contact problem; see Reference 24. In this paper, we focus our attention on the above basic B-differentiable Newton method and the interior point method to be described next; the former because of its direct applicability to contact problem formulations of the 'augmented Lagrangian' type (see Section 4) and the latter because of its well-documented effectiveness for solving linear programs and monotone complementarity problems.

The main practical deficiency is that (11) is generally not a linear equation system, and in large-scale problems this may limit the effectiveness. However, numerical tests show that for the frictional contact problems treated in this paper, (11) can be substituted for a related linear system of equations without effecting the convergence of the method, see Section 4.

There is a principal deficiency associated with the above convergence theorem; namely, the assumed F-differentiability of H_N at the limit vector z^∞. Although several modified Newton methods for non-smooth equations developed subsequently to Reference 17 (such as those presented in References 25 and 26) have resolved this theoretical deficiency, our experience with the basic B-differentiable Newton method as described above suggests that this method is well suited for solving the contact problem; as seen from the numerical results in Section 6. Thus we have not attempted to adopt any of the more advanced non-smooth Newton methods in our study.

As a related development, we mention two recent papers by Leung *et al.*[27] and Chen *et al.*[28] These authors propose smoothing techniques for the non-smooth equations; the resulting smoothed equations are then solved by the classical damped Newton method. It turns out that a great deal of research efforts in the complementarity community has in recent years been devoted to smoothing methods (see e.g. References 29–31); it would be interesting to investigate how some of the latter research can benefit the numerical solution of the contact problems. Such an investigation is beyond the scope of this paper.

3.2. An interior point method

Motivated by the popularity of the family of interior point methods in the mathematical programming community, we become interested in applying one such method for solving the contact problem described in Section 2. In what follows, we will describe the potential reduction interior point method proposed in Reference 18 for solving the constrained equation (9). Parallel to the notation H_N used in the previous subsection, we use H_I for the mapping H in this subsection. The blanket assumptions are as follows:

(A1) Ω is a closed subset of $\Re^n$ with a non-empty interior, denoted int(Ω); H_I is continuously differentiable in int(Ω);
(A2) H_I is partitioned as

$$H_I(z) = \begin{pmatrix} F(z) \\ G(z) \end{pmatrix}$$

where $F : \Re^n \to \Re^{n_1}$ and $G : \Re^n \to \Re^{n_2}$ with $n_1 + n_2 = n$, such that

$$\Omega_{++} \equiv \{z \in \Omega:\ G(z) > 0\} = \{z \in \text{int}(\Omega):\ G(z) > 0\} \neq \emptyset$$

(A3) the Jacobian matrix $\nabla H_I(z)$ is non-singular for all $z \in \Omega_{++}$.

Before giving the details of the interior point method, we summarize the essential ideas behind it. Initiated at a vector in the set Ω_{++}, the method generates a sequence of vectors lying in this set. At each iteration, a perturbed Newton equation is set up at the current iterate $z^k \in \Omega_{++}$; the goal of the perturbation is to facilitate the membership of the next iterate in Ω_{++}. It turns out that the solution obtained from the perturbed Newton equation is a descent direction for the following

merit function: for a given scalar $\zeta > n_2$,

$$\psi(z) \equiv \zeta \log(F(z)^{\mathrm{T}} F(z) + e^{\mathrm{T}} G(z)) - \sum_{i=1}^{n_2} \log G_i(z) \quad \text{for } z \in \Omega_{++} \tag{13}$$

where e is the n_2-vector of all ones. The next iterate z^{k+1} is obtained as the vector $z^k + \tau_k \mathrm{d}z^k$ for a suitable step length $\tau_k > 0$ that yields sufficient decrease in ψ from its current value $\psi(z^k)$.

Below is a step-by-step description of the potential reduction interior point method for solving the constrained equation (9). For more details of the method and its justification, see Reference 18.

Description of Algorithm IP

Step 1 (*Initialization*). Let scalars $\zeta > n_2$; $\alpha, \hat{\tau} \in (0,1)$; $\bar{\sigma} \in [0,1)$; $\xi > 0$; and $\rho \in (0,1)$ be given. Choose any $z^0 \in \Omega_{++}$ and $\sigma_0 \in [0, \bar{\sigma}]$. Set $k = 0$.

Step 2 (*Direction generation*). Solve the system of linear equations:

$$\nabla H_{\mathrm{I}}(z^k)\, \mathrm{d}z = -H_{\mathrm{I}}(z^k) + \begin{pmatrix} 0 \\ \sigma_k \beta_k e \end{pmatrix} \tag{14}$$

where $\beta_k = e^{\mathrm{T}} G(z^k)/n_2$, to obtain the search direction $\mathrm{d}z^k$.

Step 3 (*Step size determination*). Determine the scalar

$$\delta_k \equiv \sup\{\tau\colon\ z^k + \tau' \,\mathrm{d}z^k \in \Omega_{++} \text{ for all } \tau' \in (0, \tau)\}$$

set $\delta_k' \equiv \min(\hat{\tau}\delta_k, \xi)$. Let m_k be the smallest non-negative integer m such that

$$\psi(z^k + \delta_k' \rho^m \,\mathrm{d}z^k) - \psi(z^k) \leqslant -\alpha \delta_k' \rho^m (1 - \sigma_k)(\zeta - n_2)$$

Set $z^{k+1} \equiv z^k + \delta_k' \rho^{m_k} \,\mathrm{d}z^k$.

Step 4 (*Termination check*). Terminate if $\Theta(z^{k+1}) \leqslant \varepsilon$. Otherwise, pick any $\sigma_{k+1} \in [0, \bar{\sigma}]$ and return to Step 2 with $k \leftarrow k + 1$.

Note that the function Θ, constructed from the B-differentiable function used in the Newton method, is used in the termination check. This is in order to make numerical comparisons between the two methods compatible.

The convergence of the above iterative algorithm is summarized in the following theorem. This result contains a rate of convergence of a sequence of iterates produced by the algorithm; for the complete proof of this theorem, see Reference 32.

Theorem 2. Under assumptions (A1)–(A3), *let* $\{z^k\}$ *be an infinite sequence generated by the above algorithm. The following statements hold*:

(1) *the sequence of iterates* $\{z^k\}$ *is well defined and contained in* Ω_{++};
(2) *the sequence* $\{H_{\mathrm{I}}(z^k)\}$ *is bounded and every accumulation point of* $\{z^k\}$, *if it exists, is a solution of the constrained equation* (9);

(3) *if in addition,*

(A4) *for all scalars* $\alpha > \beta > 0$ *and* $\gamma > 0$, *the level set below is bounded,*

$$\{z \in \Omega\colon\ \|F(z)\| \leqslant \gamma,\ \ \beta e \leqslant G(z) \leqslant \alpha e\}$$

then $\lim_{k\to\infty} H_{\mathrm{I}}(z^k) = 0$.

Finally, assume that $\lim_{k\to\infty} \sigma_k = 0$. *If* z^∞ *is an accumulation point of* $\{z^k\}$ *and* H_{I} *is continuously differentiable near* z^∞ *with* $\nabla H_{\mathrm{I}}(z^\infty)$ *non-singular, then* $\{z^k\}$ *converges to* z^∞ *superlinearly, i.e.*

$$\lim_{k\to\infty} \frac{\|z^{k+1} - z^\infty\|}{\|z^k - z^\infty\|} = 0$$

if and only if $\lim_{k\to\infty} \tau_k = 1$.

4. FORMULATION AS A B-DIFFERENTIABLE SYSTEM

Originating in the field of non-linear programming,[33–35] the augmented Lagrangian approach was independently developed for frictional contact problems by Alart and Curnier,[11] De Saxce and Feng[13] and Simo and Laursen.[12] Other references include References 3, 14 and 15. In Strömberg[9] such a formulation is used for solving fretting problems. The essential idea of the original augmented Lagrangian approach for solving inequality constrained optimization problems is to penalize the inequality constraints by using an augmented Lagrangian term in the objective function, thus obtaining an 'equivalent' equality constrained optimization formulation for the given inequality constrained problem. A difficulty arises when this approach is extended to frictional contact problems. This is due to the absence of a natural minimization principle associated with these problems. The authors of the cited references have circumvented the difficulties with various proposals to deal with the inequality constraints in the frictional contact problems. In particular, Alart and Curnier[11] stated a '*quasi*'-minimization problem where p_{in} in $\mathscr{F}(\mu_i p_{in})$ is kept fixed.

The present formulation is strongly related to, and indeed initially inspired by, the paper by Alart and Curnier.[11] However, our formulation contains certain simplifications of the formulation of that paper and the derivation is based directly on the Euclidean projector, without relying on any auxiliary minimization principle and/or other indirect techniques. Furthermore, as first observed in References 3 and 14, augmented Lagrangian-like formulations of the frictional contact problem lead to a system of B-differentiable equations.

In what follows, we shall present the system of B-differentiable equations that describe the contact problem stated in Section 2. As in References 3, 9 and 14, our derivation below is based directly on the Euclidean projector.

Firstly, equations (1) and (3) are put together to yield

$$Ku + C_{\mathrm{n}}^{\mathrm{T}} p_{\mathrm{n}} + C_{\mathrm{t}}^{\mathrm{T}} p_{\mathrm{t}} + C_{\mathrm{o}}^{\mathrm{T}} p_{\mathrm{o}} = f^{\mathrm{ext}} \tag{15}$$

Next, the variational inequality (5) is rewritten in an obvious manner as: for any scalar $r > 0$

$$\min(p_{\mathrm{n}}, r(g - C_{\mathrm{n}} u)) = 0 \tag{16}$$

where the minimum operator is applied componentwise to its two arguments. When writing (8) as an equation we have to take care of the fact that (8) is defined only for $p_{in} \geqslant 0$. One possibility

to deal with this is to replace p_{in} with $(p_{in} + r(C_{in}u - g_i))_+$, which is justified because (16) is equivalent to

$$p_{in} = (p_{in} + r(C_{in}u - g_i))_+ \tag{17}$$

where $x_+ \equiv \max(0,x)$ for any vector x. Alternatively, we could also use $(p_{in})_+$ as a substitute for p_{in} in (8). Using this latter substitution, we write for $i = 1, \ldots, n_c$,

$$\mathscr{F}_i \equiv \mathscr{F}(\mu_i(p_{in})_+) = \{(q_{it}, q_{io}) \in \Re^2 \colon q_{it}^2 + q_{io}^2 \leqslant (\mu_i(p_{in})_+)^2\}$$

Let Π_i denote the Euclidean projector onto the planar disk $\mathscr{F}_i$. We then obtain the following equivalent formulation of (8) as a non-smooth equation: for any $r > 0$,

$$\begin{aligned}(p_{it}, p_{io}) &= \Pi_i(p_{it}(r), p_{io}(r)) \\ &= \begin{cases} (p_{it}(r), p_{io}(r)) & \text{if } (p_{it}(r), p_{io}(r)) \in \mathscr{F}_i \\ \dfrac{\mu_i(p_{in})_+}{\sqrt{(p_{it}(r))^2 + (p_{io}(r))^2}}\ (p_{it}(r), p_{io}(r)) & \text{if } (p_{it}(r), p_{io}(r)) \notin \mathscr{F}_i \end{cases}\end{aligned}$$

or more compactly,

$$(p_{it}, p_{io}) = \min\left(\frac{\mu_i (p_{in})_+}{\sqrt{(p_{it}(r))^2 + (p_{io}(r))^2}}, 1\right) (p_{it}(r), p_{io}(r)) \tag{18}$$

where $0/0$ is defined to be 1 in the last expression, and

$$p_{it}(r) \equiv p_{it} + r(C_{it}u - \bar{w}_{it}) \quad \text{and} \quad p_{io}(r) \equiv p_{io} + r(C_{io}u - \bar{w}_{io})$$

More generally, one can, of course, write down (16) and (18) using a possibly distinct $r_i > 0$ for each contact node $i = 1, \ldots, n_c$. This fact is used when the method is implemented, see Section 6.1. However, here we consider the settings of (16) and (18) with just one scalar $r > 0$ for simplicity.

Concatenating the equations (15), (16) and (18), we define the following function for any fixed $r > 0$,

$$H_N(u, p_t, p_o, p_n) \equiv \begin{pmatrix} Ku + C_n^T p_n + C_t^T p_t + C_o^T p_o - f^{ext} \\ -\min(p_n, r(g - C_n u)) \\ \left(-\begin{pmatrix} p_{it} \\ p_{io} \end{pmatrix} + \Pi_i \begin{pmatrix} p_{it}(r) \\ p_{io}(r) \end{pmatrix}\right)_{i=1}^{i=n_c} \end{pmatrix} \tag{19}$$

The B-differentiability of this function follows from two observations. One, the min function is clearly B-differentiable with an obvious B-derivative; two, the projection map Π_i is also B-differentiable, by (18); see also Reference 9, Appendix. There are two essential differences in this formulation as compared to that of Alart and Curnier.[11] The first one is the use of $(p_{in})_+$ instead of $(p_{in} + r(C_{in}u - g_i))_+$. The second is that we do not substitute (17) and (18) into the first line of (19). Each of these modifications simplifies the formulation but still keeps the property of B-differentiability. For frictionless contact problems the substitution of (17) into (19) is preferable if one also normalizes (16) with the penalty term r, since then the Newton equation (11) possesses symmetric qualities. However, for problems including friction this step is fruitless since in this

case it will not produce symmetry. Also, it turns out that without the substitution, the first line of (19) is linear for all iterations. This is numerically desirable when factoring the matrix related to $\tilde{H}'_{\mathrm{N}}(z;\mathrm{d}z)$, defined below.

Letting z denote the generic tuple $(u, p_{\mathrm{t}}, p_{\mathrm{o}}, p_{\mathrm{n}})$, we will in what follows give an expression for the directional derivative $H'_{\mathrm{N}}(z;\mathrm{d}z)$, where $\mathrm{d}z$ denotes the direction vector $(\mathrm{d}u, \mathrm{d}p_{\mathrm{t}}, \mathrm{d}p_{\mathrm{o}}, \mathrm{d}p_{\mathrm{n}})$. For this purpose, we introduce the following index sets which are motivated by the two min functions in (16) and (18):

$$
\begin{aligned}
\mathscr{I}_{<} &= \{i:\ p_{i\mathrm{n}} + r(C_{i\mathrm{n}}u - g_i) < 0\}\\
\mathscr{I}_{=} &= \{i:\ p_{i\mathrm{n}} + r(C_{i\mathrm{n}}u - g_i) = 0\}\\
\mathscr{I}_{>} &= \{i:\ p_{i\mathrm{n}} + r(C_{i\mathrm{n}}u - g_i) > 0\}\\
\mathscr{J} &= \{i:\ p_{i\mathrm{n}} < 0\}\\
\mathscr{J}_{<} &= \{i:\ p_{i\mathrm{n}} > 0,\ (p_{i\mathrm{t}}(r))^2 + (p_{i\mathrm{o}}(r))^2 < (\mu_i p_{i\mathrm{n}})^2\}\\
\mathscr{J}_{=} &= \{i:\ p_{i\mathrm{n}} > 0,\ (p_{i\mathrm{t}}(r))^2 + (p_{i\mathrm{o}}(r))^2 = (\mu_i p_{i\mathrm{n}})^2\}\\
\mathscr{J}_{>} &= \{i:\ p_{i\mathrm{n}} > 0,\ (p_{i\mathrm{t}}(r))^2 + (p_{i\mathrm{o}}(r))^2 > (\mu_i p_{i\mathrm{n}})^2\}\\
\mathscr{K}_{+} &= \{i:\ p_{i\mathrm{n}} = 0 < (p_{i\mathrm{t}}(r))^2 + (p_{i\mathrm{o}}(r))^2\}\\
\mathscr{K}_{0} &= \{i:\ p_{i\mathrm{n}} = p_{i\mathrm{t}}(r) = p_{i\mathrm{o}}(r) = 0\}
\end{aligned}
$$

We then have

$$
H'_{\mathrm{N}}(z;\mathrm{d}z) = \begin{pmatrix}
K\,\mathrm{d}u + C_{\mathrm{n}}^{\mathrm{T}}\,\mathrm{d}p_{\mathrm{n}} + C_{\mathrm{t}}^{\mathrm{T}}\,\mathrm{d}p_{\mathrm{t}} + C_{\mathrm{o}}^{\mathrm{T}}\mathrm{d}p_{\mathrm{o}}\\
(-\mathrm{d}p_{i\mathrm{n}})_{i\in\mathscr{I}_{<}}\\
(-\min(\mathrm{d}p_{i\mathrm{n}}, -rC_{i\mathrm{n}}\,\mathrm{d}u))_{i\in\mathscr{I}_{=}}\\
(rC_{i\mathrm{n}}\,\mathrm{d}u)_{i\in\mathscr{I}_{>}}\\
\begin{pmatrix} -\mathrm{d}p_{i\mathrm{t}} \\ -\mathrm{d}p_{i\mathrm{o}} \end{pmatrix}_{i\in\mathscr{J}}\\
\begin{pmatrix} rC_{i\mathrm{t}}\,\mathrm{d}u \\ rC_{i\mathrm{o}}\,\mathrm{d}u \end{pmatrix}_{i\in\mathscr{J}_{<}}\\
\left[r\begin{pmatrix} C_{i\mathrm{t}}\,\mathrm{d}u \\ C_{i\mathrm{o}}\,\mathrm{d}u \end{pmatrix} + \theta_i^r \begin{pmatrix} p_{i\mathrm{t}}(r) \\ p_{i\mathrm{o}}(r) \end{pmatrix}\right]_{i\in\mathscr{J}_{=}}\\
\left[-\begin{pmatrix} \mathrm{d}p_{i\mathrm{t}} \\ \mathrm{d}p_{i\mathrm{o}} \end{pmatrix} + \dfrac{\mu_i\,\mathrm{d}p_{i\mathrm{n}}}{\sqrt{(p_{i\mathrm{t}}(r))^2 + (p_{i\mathrm{o}}(r))^2}} \begin{pmatrix} p_{i\mathrm{t}}(r) \\ p_{i\mathrm{o}}(r) \end{pmatrix} + R_i^r \begin{pmatrix} \mathrm{d}p_{i\mathrm{t}} + rC_{i\mathrm{t}}\,\mathrm{d}u \\ \mathrm{d}p_{i\mathrm{o}} + rC_{i\mathrm{o}}\,\mathrm{d}u \end{pmatrix}\right]_{i\in\mathscr{J}_{>}}\\
\left[-\begin{pmatrix} \mathrm{d}p_{i\mathrm{t}} \\ \mathrm{d}p_{i\mathrm{o}} \end{pmatrix} + \dfrac{\mu_i(\mathrm{d}p_{i\mathrm{n}})_+}{\sqrt{(p_{i\mathrm{t}}(r))^2 + (p_{i\mathrm{o}}(r))^2}} \begin{pmatrix} p_{i\mathrm{t}}(r) \\ p_{i\mathrm{o}}(r) \end{pmatrix}\right]_{i\in\mathscr{K}_{+}}\\
\left[-\begin{pmatrix} \mathrm{d}p_{i\mathrm{t}} \\ \mathrm{d}p_{i\mathrm{o}} \end{pmatrix} + \min\left(\dfrac{\mu_i(\mathrm{d}p_{i\mathrm{n}})_+}{\sqrt{(\mathrm{d}p_{i\mathrm{t}}(r))^2 + (\mathrm{d}p_{i\mathrm{o}}(r))^2}}, 1\right) \begin{pmatrix} \mathrm{d}p_{i\mathrm{t}} + rC_{i\mathrm{t}}\,\mathrm{d}u \\ \mathrm{d}p_{i\mathrm{o}} + rC_{i\mathrm{o}}\,\mathrm{d}u \end{pmatrix}\right]_{i\in\mathscr{K}_0}
\end{pmatrix} \tag{20}
$$

where for $i \in \mathscr{J}_>$,

$$R_i^r \equiv \frac{\mu_i p_{in}}{\left((p_{it}(r))^2 + (p_{io}(r))^2\right)^{3/2}} \begin{pmatrix} (p_{io}(r))^2 & -p_{it}(r)p_{io}(r) \\ -p_{it}(r)p_{io}(r) & (p_{it}(r))^2 \end{pmatrix}$$

and for $i \in \mathscr{J}_=$,

$$\theta_i^r \equiv \frac{1}{\mu_i p_{in}} \min\left(\mu_i \, \mathrm{d}p_{in} - \frac{p_{it}(r)\,(\mathrm{d}p_{it} + rC_{it}\,\mathrm{d}u) + p_{io}(r)\,(\mathrm{d}p_{io} + rC_{io}\,\mathrm{d}u)}{\mu_i p_{in}},\, 0\right)$$

As indicated in Section 3, a drawback of the B-differentiable Newton method is that (11) is generally not a linear equation. However, as seen from the above, in this particular application, $H'_{\mathrm{N}}(z;\mathrm{d}z)$ and then also (11), is, nevertheless, linear in $\mathrm{d}z$ if $\mathscr{I}_= \cup \mathscr{J}_= \cup \mathscr{K}_+ \cup \mathscr{K}_0$ is empty. Since these index sets are defined by equalities, they can be expected to be empty in the large majority of practical iterations; this phenomenon is indeed observed in the numerical experiments performed below. As a matter of fact, a modification of the B-differentiable Newton method could be developed (for contact problems in particular and complementarity problems in general) in which one only generates iterates that are F-differentiable points of the aggregate mapping H_{N}; see Reference 36. For our purpose, such a modification is not necessary. Specifically, we have chosen to implement the Newton method by using a substitute for the directional derivative in (20) as given below:

$$\tilde{H}'_{\mathrm{N}}(z;\mathrm{d}z) \equiv \begin{pmatrix} K\,\mathrm{d}u + C_{\mathrm{n}}^{\mathrm{T}}\mathrm{d}p_{\mathrm{n}} + C_{\mathrm{t}}^{\mathrm{T}}\mathrm{d}p_{\mathrm{t}} + C_{\mathrm{o}}^{\mathrm{T}}\mathrm{d}p_{\mathrm{o}} \\ (-\mathrm{d}p_{in})_{i \in \mathscr{I}_< \cup \mathscr{I}_=} \\ (rC_{in}\,\mathrm{d}u)_{i \in \mathscr{J}_>} \\ \begin{pmatrix} -\mathrm{d}p_{it} \\ -\mathrm{d}p_{io} \end{pmatrix}_{i \in \mathscr{J} \cup \mathscr{K}_+ \cup \mathscr{K}_0} \\ \begin{pmatrix} rC_{it}\,\mathrm{d}u \\ rC_{io}\,\mathrm{d}u \end{pmatrix}_{i \in \mathscr{J}_< \cup \mathscr{J}_=} \\ \left[-\begin{pmatrix} \mathrm{d}p_{it} \\ \mathrm{d}p_{io} \end{pmatrix} + \frac{\mu_i\,\mathrm{d}p_{in}}{\sqrt{(p_{it}(r))^2 + (p_{io}(r))^2}} \begin{pmatrix} p_{it}(r) \\ p_{io}(r) \end{pmatrix} + R_i^r \begin{pmatrix} \mathrm{d}p_{it} + rC_{it}\,\mathrm{d}u \\ \mathrm{d}p_{io} + rC_{io}\,\mathrm{d}u \end{pmatrix} \right]_{i \in \mathscr{J}_>} \end{pmatrix} \tag{21}$$

and solve instead of (11) the linear equation

$$H_{\mathrm{N}}(z^k) + \tilde{H}'_{\mathrm{N}}(z^k;\mathrm{d}z^k) = 0 \tag{22}$$

This modification of the algorithm has never caused any convergence difficulties. A similar experience is confirmed by the comments in Alart and Curnier.[11]

It appears that the expression (20) for the directional derivative $H'_{\mathrm{N}}(z^k;\mathrm{d}z)$ has not previously been given in the literature; its derivation is not entirely trivial. The term involving R_i^r disappears in two-dimensional contact problems; moreover, the expression (21) simplifies substantially for these problems. These simplifications are included in the implementation of the BN algorithm for solving two-dimensional contact problems to be reported in the last section.

5. CONSTRAINED EQUATION FORMULATION

The idea of using an interior point method for solving frictional contact problems occurs more recently than the augmented Lagrangian approach. In Reference 19, the interior point method described in Section 3.2 is applied to a rigid-body frictional contact problem. In this rigid-body application, a four-sided pyramid is employed to approximate the quadratic friction cone and the resulting approximated contact model is solved by the interior point method. In what follows, we explain how a two-dimensional equivalence of the contact problem described in Section 2 can be cast as a constrained equation solvable by the interior point method. The reason why we focus only on two-dimensional problems is because there has been some significant numerical difficulties with the application of the interior-point method for solving three-dimensional problems. We have investigated a separate approach for the latter problems.[8]

For two-dimensional problems, the time discretized form of Coulomb's law may be expressed as

$$\begin{aligned} p_{i\mathrm{t}} \in \mathscr{F}_{2\mathrm{d}}(\mu_i p_{i\mathrm{n}})\colon (w_{i\mathrm{t}} - \bar{w}_{i\mathrm{t}})(q_{i\mathrm{t}} - p_{i\mathrm{t}}) \leqslant 0, \quad & \forall q_{i\mathrm{t}} \in \mathscr{F}_{2\mathrm{d}}(\mu_i p_{i\mathrm{n}}) \\ \mathscr{F}_{2\mathrm{d}}(\gamma) \equiv \{q_{i\mathrm{t}} \in \Re\colon |q_{i\mathrm{t}}| \leqslant \gamma\} \quad & \text{for } \gamma \in \Re_+ \end{aligned} \tag{23}$$

This discretized maximum dissipation principle is satisfied if and only if there exist slack variables s_i^+ and s_i^- and multipliers λ_i^+ and λ_i^- such that for all $i = 1, \ldots, n_{\mathrm{c}}$,

$$\begin{aligned} s_i^+ - \mu_i p_{i\mathrm{n}} + p_{i\mathrm{t}} &= 0 \\ s_i^- - \mu_i p_{i\mathrm{n}} - p_{i\mathrm{t}} &= 0 \\ w_{i\mathrm{t}} - \bar{w}_{i\mathrm{t}} &= \lambda_i^+ - \lambda_i^- \\ s_i^+ \lambda_i^+ = 0, \quad s_i^- \lambda_i^- &= 0 \\ s_i^+, s_i^-, \lambda_i^+, \lambda_i^- &\geqslant 0 \end{aligned} \tag{24}$$

In what follows, we write $a \circ b$ to denote the Hadamard product of the two vectors a and b; i.e. $a \circ b$ is the vector whose components are equal to the product of the corresponding components of a and b. Introducing slacks v_{n} for the inequality:

$$C_{\mathrm{n}} u - g \leqslant 0$$

and aggregating equations (2), (4), (15) and (24), we define the set

$$\Omega \equiv \Re^{n_u} \times \Re^{n_{\mathrm{c}}} \times \Re_+^{6n_{\mathrm{c}}}$$

and the mapping

$$H_{\mathrm{l}}(u, p_{\mathrm{t}}, p_{\mathrm{n}}, v_{\mathrm{n}}, s, \lambda) \equiv \begin{pmatrix} Ku + C_{\mathrm{t}}^{\mathrm{T}} p_{\mathrm{t}} + C_{\mathrm{n}}^{\mathrm{T}} p_{\mathrm{n}} - f^{\mathrm{ext}} \\ v_{\mathrm{n}} + C_{\mathrm{n}} u - g \\ -C_{\mathrm{t}} u + L_{\mathrm{t}}^{\mathrm{T}} \lambda + \bar{w}_{\mathrm{t}} \\ s - \mu L_{\mathrm{n}} p_{\mathrm{n}} + L_{\mathrm{t}} p_{\mathrm{t}} \\ p_{\mathrm{n}} \circ v_{\mathrm{n}} \\ \lambda \circ s \end{pmatrix} \tag{25}$$

Here, $\mu = \mathrm{diag}(\mu_1, \ldots, \mu_{n_c}, \mu_1, \ldots, \mu_{n_c})$,

$$L_n = \begin{pmatrix} I \\ I \end{pmatrix}, \quad L_t = \begin{pmatrix} I \\ -I \end{pmatrix}$$

where $I \in \Re^{n_c \times n_c}$ is the identity matrix, and $s \equiv (s^{\pm})$ and $\lambda \equiv (\lambda^{\pm})$. Clearly, the contact problem is equivalent to the constrained equation (9) with the pair (Ω, H_I) defined above.

One can make some interesting comparisons between the two mappings (19) and (25). In the former mapping, the inequalities of the contact problem are cast in terms of the non-smooth functions min and Π_i. In the latter mapping, the inequalities of the contact problem are converted into equations with the introduction of slack variables; the complementarity conditions are retained with the use of the Hadamard product; and the non-negativity of variables are treated as explicit constraints. As we have mentioned before, there are many other ways of dealing with inequalities and complementarity conditions; all of these are applicable to the contact problem. It would be much beyond the scope of this work to investigate each of these alternative ways. Our hope is that the present study will provide the foundation and inspire further research into the numerical solution of frictional contact problems by complementarity methods.

There are two particularly interesting partitions associated with this mapping H_I. One partition has

$$F(u, p_t, p_n, v_n, s, \lambda) \equiv \begin{pmatrix} Ku + C_t^T p_t + C_n^T p_n - f^{ext} \\ v_n + C_n u - g \\ -C_t u + L_t^T \lambda + \bar{w}_t \end{pmatrix} \tag{26}$$

and

$$G(u, p_t, p_n, v_n, s, \lambda) \equiv \begin{pmatrix} s - \mu L_n p_n + L_t p_t \\ p_n \circ v_n \\ \lambda \circ s \end{pmatrix} \tag{27}$$

Associated with this partition, we have

$$\Omega_{++} = \{(u, p_t, p_n, v_n, s, \lambda) \in \Re^{n_u} \times \Re^{n_c} \times \Re^{6n_c}_{++} : s - \mu L_n p_n + L_t p_t > 0\} \tag{28}$$

For this partition, it is easy to obtain an initial tuple $z^0 \equiv (u^0, p_t^0, p_n^0, v_n^0, s^0, \lambda^0)$ belonging to Ω_{++} as required by Algorithm IP.

The other partition has

$$\tilde{F}(u, p_t, p_n, v_n, s, \lambda) \equiv \begin{pmatrix} Ku + C_t^T p_t + C_n^T p_n - f^{ext} \\ v_n + C_n u - g \\ -C_t u + L_t^T \lambda + \bar{w}_t \\ s - \mu L_n p_n + L_t p_t \end{pmatrix} \tag{29}$$

and

$$\tilde{G}(u, p_t, p_n, v_n, s, \lambda) \equiv \begin{pmatrix} p_n \circ v_n \\ \lambda \circ s \end{pmatrix} \tag{30}$$

Associated with the latter partitioning, we have

$$\tilde{\Omega}_{++} \equiv \Re^{n_u} \times \Re^{n_c} \times \Re^{6n_c}_{++} \tag{31}$$

For this partition, it is trivial to obtain an initial tuple $z^0 \equiv (u^0, p_t^0, p_n^0, v_n^0, s^0, \lambda^0)$ belonging to $\tilde{\Omega}_{++}$ as required by Algorithm IP. Note that Ω_{++} is a subset of $\tilde{\Omega}_{++}$.

The difference between these the partitions (F, G) versus $(\tilde{F}, \tilde{G})$ is that in the former partition (F, G), there is a restriction on the friction constraint $(s - \mu L_n p_n + L_t p_t > 0)$; whereas in the latter partition $(\tilde{F}, \tilde{G})$, there is no restriction on this constraint. In both cases, the iterates produced by the interior point algorithm are allowed to lie outside of the friction cone; only the limit points of the iterates, if they exist, will satisfy this cone constraint.

Specializing Theorem 2 to the CE (H_I, Ω) with the partition (F, G) given by (26) and (27), we can state a convergence result of Algorithm IP applied to the two-dimensional contact problem. It is trivial to see that assumptions (A1) and (A2) hold. By postulating that (A3) also holds, we have the following convergence result.

Theorem 3. Let K be a symmetric positive-semidefinite matrix. Assume that the Jacobian matrix $\nabla H_I(u, p_t, p_n, v_n, s, \lambda)$, where H_I is defined by (25), *is non-singular for all tuples* $(u, p_t, p_n, v_n, s, \lambda) \in \Omega_{++}$, *where Ω_{++} is given by* (28). *If the following two conditions hold*:

(a) $[Ku = 0,\ C_n u \leqslant 0] \Rightarrow u = 0$; *and*
(b) *the matrix* $[C_n^T\ \ C_t^T]$ *has linearly independent columns*,

then the interior point algorithm, when applied to the partition (F, G) given by (26) *and* (27), *generates a well-defined sequence*

$$\{(u^k, p_t^k, p_n^k, v_n^k, s^k, \lambda^k)\} \subset \Omega_{++}$$

with the property that

$$\lim_{k \to \infty} H_I(u^k, p_t^k, p_n^k, v_n^k, s^k, \lambda^k) = 0$$

Proof. See the appendix.

6. NUMERICAL EXAMPLES

The algorithms presented in the previous sections are implemented for a two-dimensional isotropic linear elastic body under the plane strain assumption. In order to compare the numerical performance of the two methods, two problems are considered which are solved using different numbers of time increments, different numbers of contact nodes and different friction coefficients.

The first problem is a punch pressed into an elastic half-plane. The elastic half-plane is approximated by a finite element mesh according to Figure 2. The second problem, shown in Figure 3, is a

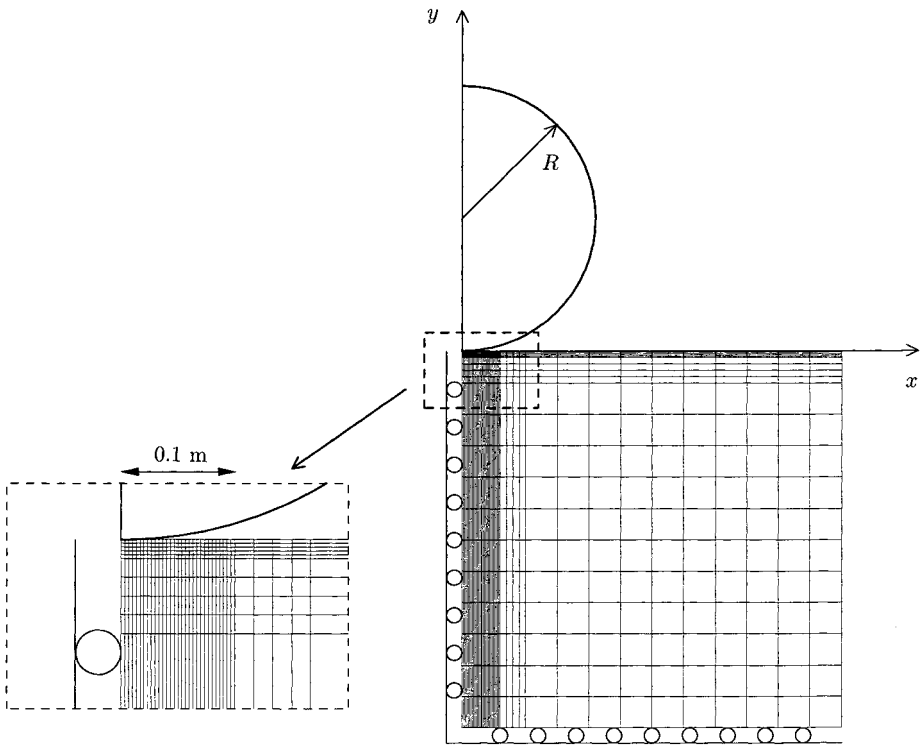

Figure 2. Finite element mesh of the half-plane. The potential contact surface is zoomed in. The location of the punch is represented by the cylinder of radius $R = 100\,\mathrm{m}$. Notice, that the punch is not made to scale with the finite element mesh as the length of the potential contact surface is 0·1 m

unilaterally constrained elastic block which is fixed at one part of its boundary and subjected to prescribed forces. In both problems, the geometries are approximated by four-noded bilinear displacement finite elements with Young's modulus taken to be 210 GPa and Poisson's ratio equals 0·3. The contact stresses are obtained from the contact forces in the same way as in Reference 20, i.e.

$$\sigma_{\mathrm{n}} = M p_{\mathrm{n}}, \quad \sigma_{\mathrm{t}} = M p_{\mathrm{t}} \tag{32}$$

where σ_{n} and σ_{t} are the normal and tangential contact stresses, respectively, and M is a diagonal matrix consisting of the inverse of the integration weights of the numerical integration rule used. The integration rule used in both problems is the trapezoidal rule.

Although the quasi-static frictional contact problem stated in Section 2 involves solving for the whole displacement vector at each time increment, it is clear that only the part of the vector corresponding to displacements of the contact nodes has to be found in order to solve the frictional contact problem. This fact has been utilized in the implementations of the algorithms and it is accomplished by first decomposing the equilibrium equations in (1) and (3) in the following way:

$$\begin{pmatrix} K_{\mathrm{c,c}} & K_{\mathrm{c,r}} \\ K_{\mathrm{r,c}} & K_{\mathrm{r,r}} \end{pmatrix} \begin{pmatrix} u_{\mathrm{c}} \\ u_{\mathrm{r}} \end{pmatrix} = \begin{pmatrix} f_{\mathrm{c}}^{\mathrm{ext}} - \bar{C}_{\mathrm{n}}^{\mathrm{T}} p_{\mathrm{n}} - \bar{C}_{\mathrm{t}}^{\mathrm{T}} p_{\mathrm{t}} \\ f_{\mathrm{r}}^{\mathrm{ext}} \end{pmatrix} \tag{33}$$

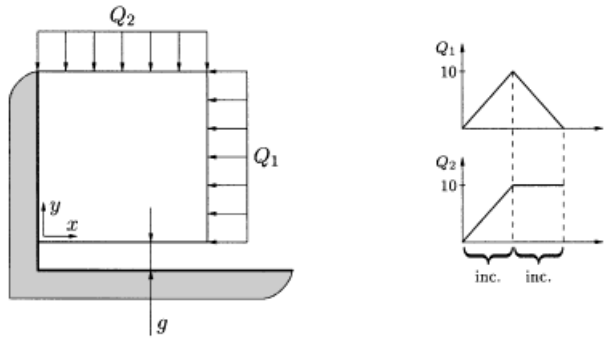

Figure 3. Elastic block subjected to line loads Q_1 and Q_2 given in MN/m

where subscript c represents subvectors formed from elements related to contact nodes and r represents all the other values. Matrices $\bar{C}_\mathrm{n}, \bar{C}_\mathrm{t} \in \Re^{n_\mathrm{c} \times 2n_\mathrm{c}}$ are formed from C_n and C_t by deleting the zero columns corresponding to non-contact nodes. Then, by use of static condensation, one obtains the following equilibrium equation by eliminating u_r:

$$(K_{\mathrm{c,c}} - K_{\mathrm{c,r}} K_{\mathrm{r,r}}^{-1} K_{\mathrm{r,c}}) u_\mathrm{c} = f_\mathrm{c}^{\mathrm{ext}} - K_{\mathrm{c,r}} K_{\mathrm{r,r}}^{-1} f_\mathrm{r}^{\mathrm{ext}} - \bar{C}_\mathrm{n}^\mathrm{T} p_\mathrm{n} - \bar{C}_\mathrm{t}^\mathrm{T} p_\mathrm{t} \tag{34}$$

which corresponds to the nodes of the potential contact surface. In both algorithms, equation (34) is derived in the initialization step. After convergence in each time step, if this is of interest, the total displacement vector can easily be obtained from

$$u_\mathrm{r} = K_{\mathrm{r,r}}^{-1} (f_\mathrm{r}^{\mathrm{ext}} - K_{\mathrm{r,c}} u_\mathrm{c}) \tag{35}$$

6.1. Algorithm BN: Implementation

The algorithm is implemented in Fortran 77 on an HP 9000 Model 712 work station using a double precision arithmetic. The essential step of the algorithm is to find the search direction, i.e. solving equation (22) with $H_\mathrm{N}(z^k)$ and $\tilde{H}'_\mathrm{N}(z^k; \mathrm{d}z^k)$ given by (19) and (21), respectively. This step is achieved by a lower and upper factorization of the matrix associated with the linear equation (22). We note that this factorization must be made each time (22) is solved, since $\tilde{H}'_\mathrm{N}(z^k; \mathrm{d}z^k)$ given by (21) depends on the iterate z^k.

The stability of the algorithm is achieved by the line search procedure in Step 3. The parameters entered into this routine are set to $\beta = 0{\cdot}9$ and $\sigma = 0{\cdot}1$. It has been found that the performance of the line search procedure turns out well with these settings for general problems. The line search is not sensitive to moderate changes of these values. However, it is of major importance to define an additional upper bound on the integer m in this step such that the line search is aborted if β^m gets numerically too small. Otherwise, it is very likely that the algorithm will stall. In our implementation, this upper bound is set to 22, i.e. β^m never becomes smaller than $0{\cdot}1$.

Concerning the termination check, ε is set equal to 10^{-2} in $\Theta(z^{k+1}) \leqslant \varepsilon$. This value is most clearly sufficient to get an accurate solution. Typical values of $\Theta(z^{k+1})$ during iterations are between 10^{15} and 10^{-15}, see Figure 5.

The remaining parameters to be set in the algorithm are the penalty terms r_i, $i = 1, \ldots, n_\mathrm{c}$, appearing in the tribological laws given by (16) and (18). In this work, each r_i is set equal to $10^{12} M_{ii}^{-1}$, where M_{ii}^{-1} is the weighting factors discussed in (32). However, r_i can be chosen

in a range between $10^9 M_{ii}^{-1}$ and $10^{13} M_{ii}^{-1}$ without any major differences in performance of the algorithm. If r_i are set to smaller values, then there might be convergence difficulties. On the contrary, if r_i are set to larger values, then the algorithm might be unstable.

Furthermore, the starting point for each time increment j is taken to be the solution point from the previous time increment $j-1$. The starting point for the first time increment is the same as for the interior point method, see below. However, the performance of the algorithm does not depend strongly on this choice, unlike the interior point method, but this point is chosen in order to get a fair comparison between the two methods.

6.2. *Algorithm IP: Implementation*

The algorithm is implemented in Fortran 90 on a DEC Alpha 200 Model 4/100 work station using double arithmetic precision. MATLAB's sparse matrix solver is invoked when solving the linear system of equations (14). We have thus far not taken advantage of the fact that most of the components of ∇H_{I} are the same for all iterations. Initial testing of the algorithm showed that the partition (F, G) gave slower convergence than the $(\tilde{F}, \tilde{G})$ partition. The latter partition is therefore used in the examples to follow. The following constants are used by the algorithm:

$$\zeta = 10n_2, \quad \sigma_0 = \xi = 1{\cdot}0, \quad \alpha = \rho = 0{\cdot}5, \quad \hat{\tau} = 0{\cdot}9999995$$

where n_2 equals $3n_{\mathrm{c}}$. The performance of the algorithm does not rely heavily on the setting of these parameters. However, the performance is sensitive to the updating of the centring parameter σ, which will be described in the following subsections.

The starting point for the first increment is in all examples

$$\begin{aligned} p_{i\mathrm{n}}^0 &= l, \quad p_{i\mathrm{t}}^0 = 0, \quad v_{i\mathrm{n}}^0 = 1{\cdot}0 \quad \forall i, \ldots, n_{\mathrm{c}} \\ u_i^0 &= 0, \quad s_i^0 = 1{\cdot}0, \quad \lambda_i^0 = 1{\cdot}0 \quad \forall i, \ldots, 2n_{\mathrm{c}} \end{aligned}$$

where l is the length of the elements on the potential contact surface. The input data and the unknowns are scaled in such a way that the forces are calculated in mega-newton and the displacements in 10^{-5} m, respectively.

We have used two different methods for choosing the starting point for the subsequent increments. In the first method we use the solution at increment $j-1$ as starting point for increment j. The second method uses a *safe* starting point which is not close to the boundary of Ω_{++}:

$$p_{i\mathrm{n}}^{0(j)} = \begin{cases} 0{\cdot}1 \max(p_{\mathrm{n}}^{(j-1)}) & \text{if } \max(p_{\mathrm{n}}^{(j-1)}) > 10^{-1} \\ 10^{-2} & \text{otherwise} \end{cases}$$

$$v_{i\mathrm{n}}^{0(j)} = \begin{cases} 0{\cdot}1 \max(v_{\mathrm{n}}^{(j-1)}) & \text{if } \max(v_{\mathrm{n}}^{(j-1)}) > 1 \\ 0{\cdot}1 & \text{otherwise} \end{cases}$$

$$s_i^{0(j)} = \begin{cases} 0{\cdot}1 \max(s^{(j-1)}) & \text{if } \max(s^{(j-1)}) > 10^{-1} \\ 10^{-2} & \text{otherwise} \end{cases}$$

$$\lambda_i^{0(j)} = \begin{cases} 0{\cdot}1 \max(\lambda^{(j-1)}) & \text{if } \max(\lambda^{(j-1)}) > 1 \\ 0{\cdot}1 & \text{otherwise} \end{cases}$$

$$p_{it}^{0(j)} = p_{it}^{(j-1)}$$

$$u_i^{0(j)} = u_i^{(j-1)}$$

where, e.g. $p_{\mathrm{n}}^{(j-1)}$ denotes the p_{n} vector at the final solution of increment $j-1$. When using the solution at increment $j-1$ as starting point for increment j, the line search is aborted whenever $\delta_k' \rho^{m_k}$ is lower than 0·01. If this extra condition is not added, there is a major risk that the algorithm will stall. The termination rule used is the same as that for the Newton method, see the subsection above. If the error norm $\Theta(z^{k+1})$ is smaller than 10^3, then the line search involved is not performed.

6.3. *A punch problem*

In this problem, a rigid cylinder of radius 100 m is pressed a total distance of 0·1 mm into an elastic half-plane. The indentation is divided into either 1, 5, 20 or 100 time steps. 880 elements are used in the discretization of the approximation of the half-plane which has dimension $1 \times 1\,\mathrm{m}^2$, see Figure 2. The potential contact surface is 0·1 m of length and the number of potential contact nodes is 31. The friction coefficients μ_i are taken to be the same for all contact nodes and the common values are 0·1, 0·4 or 0·8. Figure 4 shows the final contact stresses with $\mu = 0{\cdot}4$ using various number of time increments. Note that the solution is quite sensitive to the number of load increments. The solutions show close agreement with those in Reference 37.

For the interior point method, the centering parameter σ is updated according to

$$\sigma_{k+1} = \begin{cases} 0{\cdot}5\sigma_k & \text{if } \Theta(z^{k+1}) > 10^5 \\ 0{\cdot}1\sigma_k & \text{otherwise} \end{cases} \tag{36}$$

The reason for choosing low σ values and skipping the line search for small errors is that we want to speed up the process by enforcing (almost) pure Newton steps near the solution.

In Table I, the number of iterations and line searches for the Newton method (BN) and the interior point method (IP) are given for different cases of increments and friction coefficients. The notations safe and prev. refer to the different choices of starting points as described above. In the table, nle denotes the average number of linear equations solved per increment and nls denotes the average number of line searches per increment. For all problems studied, nls equals zero for the interior point method when using the safe starting point. In Figure 5, the error norm $\Theta(z^{k+1})$ is plotted versus the number of iterations for the two methods.

From Table I, it is evident that nle is always lower for the Newton method. Furthermore, nle decreases for BN when μ increases, at least for 10 increments or higher. This probably depends on the fact that the Newton method generally converges more rapidly if the sticking area is large. Concerning the interior point method, nle is reasonably constant using the safe method for different numbers of increments, whereas for the prev. point strategy, nle decreases using larger numbers of increments.

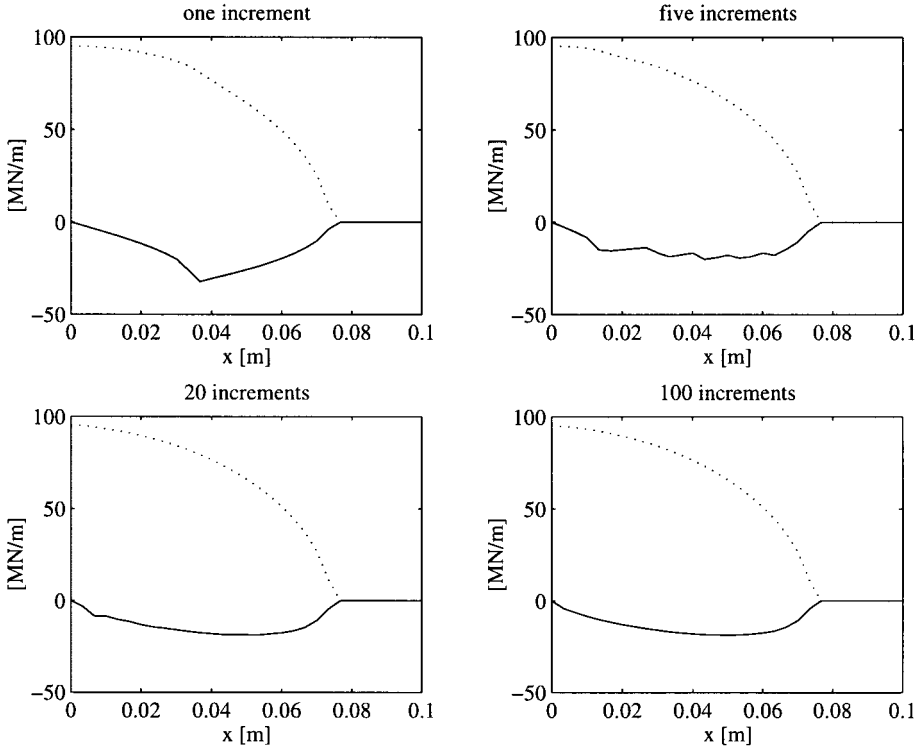

Figure 4. Normal (dotted line) and tangential (solid line) contact stresses in MN/m after full indentation for various numbers of time increments with $\mu = 0{\cdot}4$. The co-ordinate x is defined in Figure 2

Table I. Execution statistics for the punch problem

	$\mu = 0{\cdot}1$					$\mu = 0{\cdot}4$					$\mu = 0{\cdot}8$				
	BN		IP			BN		IP			BN		IP		
			safe	prev.				safe	prev.				safe	prev.	
inc.	nle	nls	nle	nle	nls	nle	nls	nle	nle	nls	nle	nls	nle	nle	nls
1	5·0	7·0	13·0	13·0	0·0	6·0	8·0	13·0	13·0	0·0	5·0	7·0	14·0	14·0	0·0
2	6·5	23·5	13·0	47·5	0·0	6·0	23·0	14·5	58·5	0·0	6·5	24·0	12·5	58·0	0·0
5	6·6	29·4	13·8	40·0	0·0	6·2	29·4	13·2	45·6	0·0	6·4	32·2	12·6	44·0	0·0
10	6·3	35·5	14·0	30·4	0·0	6·0	36·2	12·9	31·9	0·0	5·9	41·4	12·3	33·1	0·0
20	6·5	32·0	15·0	19·3	0·0	5·4	32·6	13·8	21·8	0·0	4·2	33·5	11·9	23·1	0·0
50	5·9	34·6	16·1	13·7	0·0	4·8	35·0	14·7	11·7	0·0	3·7	34·7	12·4	14·5	0·1
100	5·1	25·2	16·9	7·5	0·0	4·3	24·6	15·6	6·5	0·0	2·7	23·3	12·5	7·4	0·2

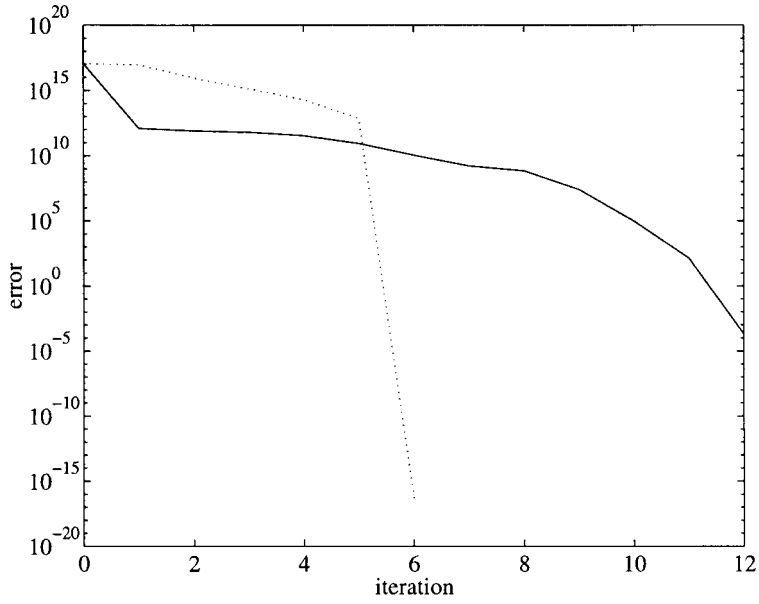

Figure 5. The error norm $\Theta(z^{k+1})$ plotted versus the number of iterations for IP and BN, when inc.$=$ 1 and $\mu = 0{\cdot}4$. The solid line refers to IP and the dotted one to BN

6.4. An elastic block problem

Figure 3 shows an elastic block of dimension $1 \times 1\,\mathrm{m}^2$ which is fixed to a wall, unilaterally constrained to a frictional foundation and subjected to varying line loads. The block is initially at a distance 0·01 mm from the rigid foundation. The line loads Q_1 and Q_2 vary according to the history scheme showed in Figure 3.

This problem is solved using different numbers of time increments, different friction coefficients and different numbers of finite elements. The number of increments, inc., during one load path is either 1, 5, 10, 20 or 50. There are two load paths, see Figure 3. The friction coefficients μ_i are again taken to be the same for all contact nodes and the common values are 0·1, 0·4 or 0·8. Furthermore, the elastic block is discretized by 5×5, 10×10, 20×20, 50×50 or 100×100 elements. Since the whole bottom part of the elastic block is defined as the potential contact surface, the number of contact nodes n_c is consequently either 5, 10, 20, 50 or 100. In Figure 6, the contact stresses, both when Q_1 is maximal (a) and at the end of the loading history (b), are shown.

Concerning the interior point method, the same updating scheme for σ as in the previous example is used for the safe starting point. However, for the other choice of starting point, the method needs to use a more conservative strategy in order to achieve convergence;

$$\sigma_{k+1} = \begin{cases} 0{\cdot}95\sigma_k & \text{if } \Theta(z^{k+1}) > 10^5 \\ 0{\cdot}1\sigma_k & \text{otherwise} \end{cases}$$

For the first increment, though, (36) is used. Unfortunately, although this strategy ensures convergence, it is exceedingly slow for certain increments; sometimes over 200 iterations. Therefore,

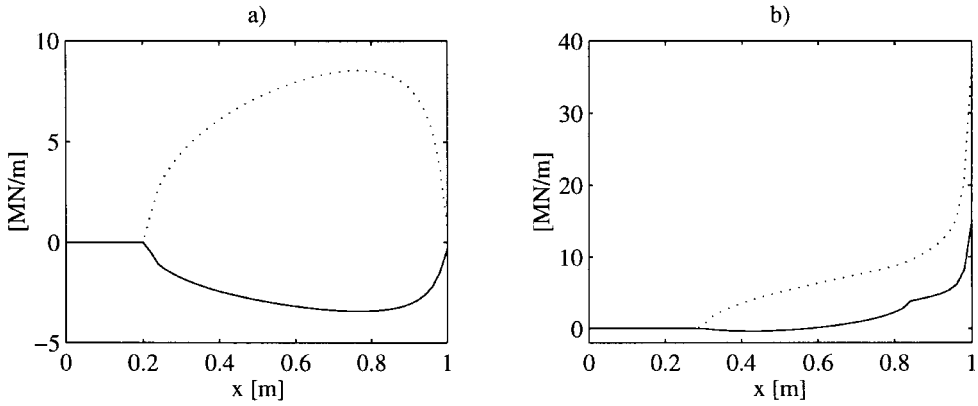

Figure 6. Normal (dotted line) and tangential (solid line) contact stresses when (a) $Q_1 = 10\,\mathrm{MN/m}$ and (b) Q_1 is decreased to zero for the case $n_c = 50$, inc.$= 50$ and $\mu = 0{\cdot}4$. The coordinate x is defined in Figure 3

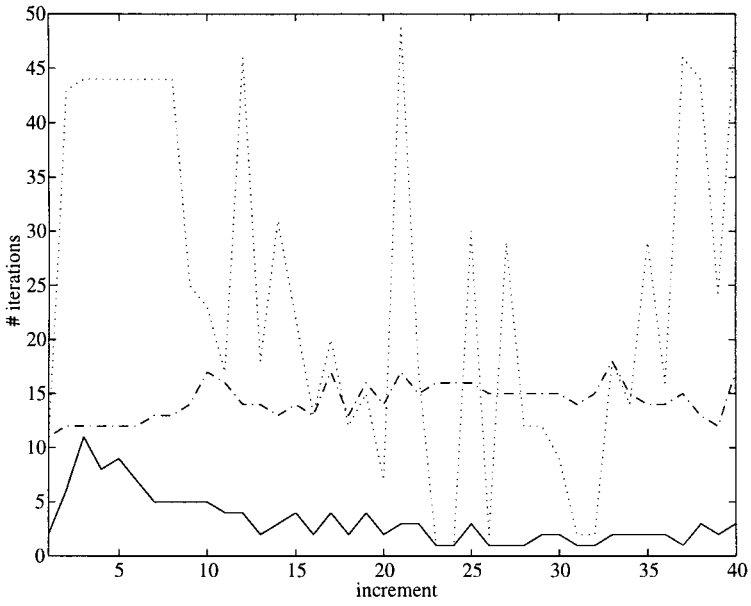

Figure 7. Number of iterations for IP and BN when $n_c = 50$, inc.$= 20$ and $\mu = 0{\cdot}4$. The dotted line refers to the prev., the dashdotted one to the safe starting points in IP and the solid one to BN

whenever the number of iterations reaches 30 we revert back to the safe starting point. In Figure 7, the number of iterations using different starting point strategies for the interior point method are plotted together with the number of iterations using the Newton method.

Figure 7 shows that the number of iterations of prev. IP changes radically for different increments, in a range between 2 and 48. One should also keep in mind that prev. IP is switched over to safe IP when the number of iterations exceeds 30. The safe IP and the BN produce a more stable behavior. The number of iterations is in a range between 11 and 18 for the safe IP and

Table II. Execution statistics for the elastic block problem

		$\mu = 0{\cdot}1$					$\mu = 0{\cdot}4$					$\mu = 0{\cdot}8$				
		BN		IP			BN		IP			BN		IP		
				safe	prev.				safe	prev.				safe	prev.	
n_c	inc.	nle	nls	nle	nle	nls	nle	nls	nle	nle	nls	nle	nls	nle	nle	nls
5	1	4·0	36·0	9·0	26·5	0·0	4·0	35·5	9·0	18·5	0·0	4·0	32·0	10·5	19·0	0·0
	5	2·4	7·7	9·9	13·3	1·4	2·1	6·6	9·9	12·7	0·8	2·2	8·1	10·8	14·7	1·9
	10	1·8	4·3	9·2	9·8	0·2	1·7	3·8	9·8	8·2	1·1	1·9	6·3	9·6	8·7	0·3
	20	1·5	2·7	9·2	6·6	0·4	1·4	3·0	10·5	5·2	0·4	1·5	3·7	9·5	6·6	1·0
	50	1·2	1·3	10·0	3·4	0·2	1·2	1·2	11·3	3·4	0·3	1·2	1·7	9·7	3·2	0·1
10	1	4·0	36·0	10·0	25·5	2·5	4·5	36·5	11·5	27·0	3·0	3·5	29·5	12·0	22·5	2·5
	5	2·8	14·6	10·4	14·2	0·7	2·6	8·7	10·7	16·5	1·3	2·7	10·5	11·1	16·3	0·4
	10	2·2	9·0	9·8	15·2	1·4	2·0	6·6	10·7	13·4	1·2	2·2	10·4	10·7	11·3	0·8
	20	1·7	5·3	9·7	10·9	1·4	1·6	5·1	11·4	7·3	0·3	1·8	6·6	10·5	8·3	0·6
	50	1·3	1·8	10·5	4·6	0·3	1·2	1·8	12·0	3·6	0·2	1·3	3·0	11·0	4·8	0·6
20	1	4·0	36·0	12·0	27·0	0·0	6·0	38·5	14·0	22·0	0·0	6·0	52·5	13·5	22·0	0·0
	5	4·2	31·3	11·0	26·5	2·9	3·5	20·9	13·2	17·1	1·1	3·4	19·7	12·2	23·6	2·1
	10	3·0	15·9	10·6	16·7	0·6	2·8	18·0	13·2	17·1	1·1	2·7	15·4	12·1	18·0	1·2
	20	2·2	9·0	10·5	12·9	0·7	2·1	13·0	13·7	11·6	1·0	2·1	12·1	12·0	12·2	0·9
	50	1·6	3·2	11·1	7·8	0·6	1·5	4·7	14·8	6·5	0·4	1·6	5·7	12·2	5·8	0·1
50	1	5·0	37·0	14·5	29·5	0·0	6·0	38·0	15·5	31·0	0·0	7·0	45·5	14·5	30·0	2·0
	5	6·2	65·5	12·9	33·9	3·4	5·1	53·0	13·7	32·6	2·5	4·5	36·1	13·8	36·6	3·3
	10	5·0	47·4	12·2	27·5	2·6	4·2	38·5	14·1	28·0	2·3	3·8	23·4	12·7	27·0	2·3
	20	3·6	30·4	12·0	21·9	2·0	3·3	25·8	14·3	24·3	2·3	2·8	19·7	12·7	22·9	2·2
	50	2·3	13·0	12·0	13·7	0·7	2·3	13·3	15·2	15·1	0·9	2·0	11·3	12·8	12·3	0·8
100	1	6·0	21·0	14·5	29·5	0·0	7·0	27·0	17·0	32·0	0·0	8·5	32·5	17·0	33·0	2·0
	5	7·5	85·9	13·4	42·0	4·2	7·1	75·2	15·8	37·6	3·2	6·3	54·8	15·1	44·0	3·8
	10	6·6	81·7	13·2	43·9	3·4	6·2	60·7	15·3	36·0	3·8	5·1	48·8	14·0	37·0	4·0
	20	5·2	61·3	13·2	29·0	2·6	4·5	40·4	15·2	28·7	2·3	3·9	35·8	14·2	29·6	3·5
	50	3·3	28·6	12·9	20·4	1·6	3·2	25·2	16·1	22·5	1·9	2·7	18·5	14·1	20·7	2·8

between 2 and 11 for the BN method. Table II summarizes the execution statistics of running the interior point method and the Newton method for this problem using the different cases of time increments, contact nodes and friction coefficients. It should be said that for this example the frictional behaviour is modeled satisfactorily with just 5 increments using $n_c = 5, 10, 20$, and 20 increments using $n_c = 50, 100$. Keeping this in mind, it is obvious that the prev. IP works worst of the three methods. Then, comparing the safe IP and the BN, it is evident that the BN always has the lowest value on nle. In addition, the linear equation system, which is solved, is of order $4n_c \times 4n_c$ for the BN and $9n_c \times 9n_c$ for the IP method. That is, for BN the number of iterations is smaller than that of IP, as well as the computational effort for each iteration. Thus, in general for the class of two-dimensional frictional problems considered in this work, it is quite evident that numerically the B-differentiable Newton method is to be preferred to the interior point method.

One additional remark: the computational effort in each iteration of the IP method can be significantly reduced by carrying out some careful linear algebraic calculations; these have not been done in this implementation of the method. Nevertheless, the small number of iterations of BN suggests that even with an advanced implementation of IP, it would be quite unlikely for the latter method to surpass the former method for solving the two-dimensional frictional contact problems reported herein.

7. CONCLUDING REMARKS

In this work a Newton method (BN) and an interior point method (IP) have been suggested for solving two-dimensional contact problems with Coulomb friction. The algorithms have been investigated numerically for two problems using different coefficients of friction, different numbers of contact nodes and different numbers of time increments. We have found that:

The number of iterations is generally lower for BN.
The cost of each iteration is lower for BN (less than half of the number of variables compared with IP).
BN is robust. When for IP, the solution at increment $j-1$ is used as starting point for increment j, the algorithm may be very sensitive to different scalings of the equations and different updating schemes of the centring parameter.

However, if the safe starting point is used in IP, we get about the same number of iterations for all increments. This number is fairly moderate and does not seem to increase that much when the size of the problem increases. Therefore, we cannot rule out the possibility that for problems with a *huge* number of potential contact nodes, IP might need less iterations than BN (at least for a one-incremental problem).

The BN algorithm has been applied to three-dimensional friction problems and is working well also; see References 8 and 10. A three-dimensional version of the interior point algorithm has also been studied. However, some major changes have had to be made in order to get the algorithm to work. In fact, for the case when some normal contact forces are zero, the algorithm converges extremely slowly (if at all). In Reference 8, we establish that the equations used in BN are semismooth, and develop a potential reduction Newton method for solving these equations.

We have felt that it could be of importance to compare from a practical point of view the methods presented in this paper with methods presented in commercial finite element programs. As an example we have chosen ABAQUS version 5.3, where one can find contact elements of penalty type and of Lagrangian type. It turned out that only one element, i.e. the GAPUNI element, handle initial gaps correctly so that solutions could be compared. The GAPUNI element is of Lagrangian type and treats the unilateral constraints by a trial and error strategy.

For frictionless contact problems, BN is comparable in CPU-time with the GAPUNI element for the particular problem treated. For frictional contact problems, a straightforward comparison of CPU-time is not appropriate because ABAQUS treats Coulomb friction by a heuristic approach; specifically, when solving for a time increment, the friction force is determined from the solution for the previous increment, i.e. an explicit time discretization approach is attempted. Thus, if we consider CPU-time per increment, then ABAQUS is expected to be faster than, for instance, BN. However, the solutions using ABAQUS with few increments are generally wrong. Consequently,

for problems where the history of contact stresses is important, one can not rely on ABAQUS to yield correct answers as effectively as the MP algorithms; for these problems, the BN algorithm is the most reliable and at least for small friction coefficients, where the implicit discreteization enables large increments, it is also the fastest. As an example of a class of problems where the full load history is of great importance, we mention wear problems for which the sliding history strongly affects the total amount of wear. Finally, in comparing algorithms, one should keep in mind the issue of robustness. The MP methods described in the present paper are supported by provable convergence results; as mentioned in the introduction, such a sound theoretical basis is lacking in the vast majority of methods for solving frictional contact problems discussed in the literature.

APPENDIX

Proof of Theorem 3

According to Theorem 2, it suffices to verify condition (A4). Assume for the sake of contradiction that for some scalars $\alpha > \beta > 0$ and $\gamma > 0$, there exists an infinite sequence

$$\{z^\nu \equiv (u^\nu, p_t^\nu, p_n^\nu, v_n^\nu, s^\nu, \lambda^\nu)\} \subset \Omega$$

satisfying for each ν,

$$\|F(z^\nu)\| \leqslant \gamma, \quad \beta e \leqslant G(z^\nu) \leqslant \alpha e$$

and

$$\lim_{\nu \to \infty} \|z^\nu\| = \infty$$

We write for each ν,

$$\left.\begin{aligned} H_u^\nu &\equiv Ku^\nu + C_t^{\mathrm{T}} p_t^\nu + C_n^{\mathrm{T}} p_n^\nu - f^{\mathrm{ext}} \\ H_n^\nu &\equiv v_n^\nu + C_n u^\nu - g \end{aligned}\right\} \tag{37}$$

$$\left.\begin{aligned} H_t^\nu &\equiv C_t u^\nu - \bar{w}_t - L_t^{\mathrm{T}} \lambda^\nu \\ H_f^\nu &\equiv s^\nu - \mu L_n p_n^\nu + L_t p_t^\nu \end{aligned}\right\} \tag{38}$$

By condition (b) and the boundedness of $\{H_u^\nu\}$, it follows that there exist positive constants δ and L such that for all ν,

$$\|(p_t^\nu, p_n^\nu)\| \leqslant L(\delta + \|u^\nu\|) \tag{39}$$

Multiplying the first equation in (38) by $(p_t^\nu)^{\mathrm{T}}$ and the second equation by $(\lambda^\nu)^{\mathrm{T}}$, and adding, we obtain for each ν,

$$(p_t^\nu)^{\mathrm{T}}(C_t u^\nu - \bar{w}_t) + (s^\nu)^{\mathrm{T}} \lambda^\nu - \mu(\lambda^\nu)^{\mathrm{T}} L_n p_n^\nu = (p_t^\nu)^{\mathrm{T}} H_t^\nu + (\lambda^\nu)^{\mathrm{T}} H_f^\nu$$

Since

$$\alpha e \geqslant G(z^\nu) \equiv \begin{pmatrix} H_f^\nu \\ p_{\mathrm{n}}^\nu \circ v_{\mathrm{n}}^\nu \\ \lambda^\nu \circ s^\nu \end{pmatrix} \geqslant \beta e$$

and p_{n}^ν and λ^ν are positive, we deduce

$$(p_{\mathrm{t}}^\nu)^{\mathrm{T}}(C_{\mathrm{t}} u^\nu - \bar{w}_{\mathrm{t}}) \geqslant (p_{\mathrm{t}}^\nu)^{\mathrm{T}} H_{\mathrm{t}}^\nu - 2\alpha n_{\mathrm{c}} \tag{40}$$

Multiplying the first equation in (37) by $(u^\nu)^{\mathrm{T}}$ and the second equation by $(p_{\mathrm{n}}^\nu)^{\mathrm{T}}$ and subtracting, we obtain

$$(u^\nu)^{\mathrm{T}} H_u^\nu - (p_{\mathrm{n}}^\nu)^{\mathrm{T}} H_{\mathrm{n}}^\nu = (u^\nu)^{\mathrm{T}} K u^\nu + (p_{\mathrm{t}}^\nu)^{\mathrm{T}} C_{\mathrm{t}} u^\nu - (f^{\mathrm{ext}})^{\mathrm{T}} u^\nu + (p_{\mathrm{n}}^\nu)^{\mathrm{T}}(-v_{\mathrm{n}}^\nu + g)$$

We claim that the sequence $\{u^\nu\}$ must be bounded. Otherwise, by this last expression, (40), and (39), it can be shown that any accumulation point u^∞ of the normalized sequence $\{u^\nu / \|u^\nu\|\}$ must satisfy: (i) $u^\infty \neq 0$, (ii) $(u^\infty)^{\mathrm{T}} K u^\infty = 0$, and (iii) $C_{\mathrm{n}} u^\infty \leqslant 0$. By the symmetry and positive semidefiniteness of K, such a vector u^∞ provides a contradiction to assumption (a). Consequently, $\{u^\nu\}$ is bounded. By (39), so is $\{(p_{\mathrm{t}}^\nu, p_{\mathrm{n}}^\nu)\}$. Moreover the second equation in (38) implies that $\{s^\nu\}$ is also bounded. Since $\{H_{\mathrm{n}}^\nu\}$ is bounded, it follows from the second equation in (37) that $\{v_n^\nu\}$ is bounded.

It remains to show that $\{\lambda^\nu\}$ is bounded. Suppose that for some index i, $\{\lambda_i^{+,\nu}\}$ is unbounded. Without loss of generality, we may assume that $\{\lambda_i^{+,\nu}\}$ tends to ∞ as $\nu \to \infty$. Then it follows from the first equation in (38) that $\{\lambda_i^{-,\nu}\}$ also tends to ∞ as $\nu \to \infty$. The fact that $\{s_i^{\pm,\nu} \lambda_i^{\pm,\nu}\}$ is bounded implies $s_i^{\pm,\nu} \to 0$ as $\nu \to \infty$. Since $H_{if}^{\pm,\nu} > 0$ by the definition of the partition (F, G), the last equation in (38) implies that $p_{in}^\nu \to 0$. But this is impossible because $p_{in}^\nu v_{in}^\nu \geqslant \beta > 0$ for all ν and $\{v_{in}^\nu\}$ is bounded. Consequently, $\{\lambda^\nu\}$ is bounded.

In summary, we have shown that the sequence $\{z^\nu\}$ is bounded. But this is a contradiction. Consequently (A4) holds and the theorem is proved. □

Remark. The fact that H_f^ν is kept positive in the algorithm, as dictated by the set Ω_{++} in (28), is an essential element in the above proof. At this time, we do not have an analogous convergence result for the partition $(\tilde{F}, \tilde{G})$ which does not impose this positivity condition.

ACKNOWLEDGEMENTS

The research of Christensen, Klarbring and Strömberg was based on work supported by the Center for Industrial Information Technology (CENIIT), ABB Atom AB and the Swedish Research Council for Engineering Sciences (TFR). The research of Pang was based on work supported by the United States National Science Foundation under grant CCR-9213739 and the Office of Naval Research under grant N00014-93-1-0228. The authors thank Jan Pettersson at Linköping University for helping out with the ABAQUS calculations. We are grateful to Professor Jeffrey Trinkle at Texas A&M University for his constructive comments that have improved the presentation of the paper.

REFERENCES

1. A. Klarbring and J. S. Pang, 'Existence of solutions to discrete semicoercive frictional contact problems', *SIAM J. Optim.*, forthcoming.
2. Z.-H. Zhong and J. Mackerle, 'Static contact problems—a review', *Engng. Comput.*, **9,** 3–37 (1992).
3. A. Klarbring, 'Mathematical programming in contact problems', in M. H. Aliabdali and C. A. Brebbia (eds), *Computational Methods in Contact Mechanics*, Computational Mechanics Publications, Southampton, 1993, pp. 233–263.
4. P. Wriggers, 'Finite element algorithms for contact problems', *Arch. Comput. Meth. Engng.*, **2,** 1–49 (1995).
5. N. Kikuchi and J. T. Oden, *Contact Problems in Elasticity: A study of Variational Inequalities and Finite Elements*, SIAM, Philadelphia, 1988.
6. Z.-H. Zhong, *Finite Element Procedures for Contact-Impact Problems*, Oxford University Press, Oxford, 1993.
7. A. Klarbring, 'Examples of non-uniqueness and non-existence of solutions to quasistatic contact problems with friction', *Ing. Arch.*, **60,** 529–541 (1990).
8. P. W. Christensen and J. S. Pang, 'Frictional contact algorithms based on semismooth Newton methods', in M. Fukushima and L. Qi (eds.), *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, Kluwer Academic Publishers, Dordrecht, Forthcoming.
9. N. Strömberg, 'An augmented Lagrangian method for fretting problems', *European J. Mech. A/Solids*, **16,** 573–593 (1997).
10. N. Strömberg, 'A Newton method for three-dimensional fretting problems', *Internal Report*, *LiTH-IKP-R-950*, Linköping (1997), submitted for publication.
11. P. Alart and A. Curnier, 'A mixed formulation for frictional contact problems prone to Newton like solution methods', *Comput. Meth. Appl. Mech. Engng.*, **92,** 353–375 (1991).
12. J. C. Simo and T. A. Laursen, 'An augmented Lagrangian treatment of contact problems involving friction', *Comput. Struct.*, **42,** 97–116 (1992).
13. G. De Saxce and Z. Q. Feng, 'New inequality and functional for contact with friction: the implicit standard material approach', *Mech. Struct. Mach.*, **19,** 301–325 (1991).
14. A. Klarbring, 'Mathematical programming and augmented Lagrangian methods for frictional contact problem', in A. Curnier (ed.), *Proceedings Contact Mechanics International Symposium*, PPUR, 1992, pp. 409–422.
15. G. Zavarise, P. Wriggers and B. A. Schrefler, 'On augmented Lagrangian algorithms for thermomechanical contact problems with friction', *Int. J. Numer. Meth. Engng.*, **38,** 2929–2949 (1995).
16. A. Heege and P. Alart, 'A frictional contact element for strongly curved contact problems', *Int. J. Numer. Meth. Engng.*, **39,** 165–184 (1996).
17. J. S. Pang, 'Newton's method for B-differentiable equations', *Math. Oper. Res.*, **15,** 311–341 (1990).
18. T. Wang, R. D. C. Monteiro and J. S. Pang, 'An interior point potential reduction method for constrained equations', *Math. Program.*, **74,** 159–196 (1996).
19. J. C. Trinkle, J. S. Pang, S. Sudarsky and G. Lo, 'On dynamic multi-rigid-body contact problems with Coulomb friction', *Z. Angew. Math. Mech.*, **77,** 267–279 (1997).
20. A. Klarbring and G. Björkman, 'A mathematical programming approach to contact problems with friction and varying contact surface', *Comput. Struct.*, **30,** 1185–1198 (1988).
21. G. Holmberg, 'A solution scheme for three-dimensional multi-body contact problems using mathematical programming', *Comput. Struct.*, **37,** 503–514 (1990).
22. J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
23. J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
24. M. C. Ferris and J. S. Pang, *Complementarity and Variational Problems: State of the Art*, SIAM Publications, Philadelphia, 1997.
25. L. Qi and J. Sun, 'A non-smooth version of Newton's method', *Math. Program.*, **58,** 353–368 (1993).
26. D. Ralph, 'Global convergence of damped Newton's method for non-smooth equations via the path search', *Math. Oper. Res.*, **19,** 352–389 (1994).
27. A. Y. T. Leung, G. Chen and W. Chen, 'System of continuously differentiable equations for two and three-dimensional frictional contact problems', manuscript, Department of Civil and Structural Engineering, University of Hong Kong, 1996.
28. W. Chen, A. Y. T. Leung and G. Chen, 'Damped Newton method for solving two and three-dimensional frictional contact problems', manuscript, Research Institute of Engineering Mechanics, Dailan University of Technology, 1996.
29. C. Chen and O. L. Mangasarian, 'A class of smoothing functions for non-linear and mixed complementarity problems', *Comput. Optim. Appl.*, **5,** 97–138 (1996).
30. T. De Luca, F. Facchinei and C. Kanzow, 'A semismooth equation approach to the solution of nonlinear complementarity problems', *Math. Program.*, **75,** 407–439 (1996).
31. S. A. Gabriel and J. J. Moré, 'Smoothing of mixed complementarity problems', in M. C. Ferris and J. S. Pang (eds.), *Complementarity and Variational Problems: State of the Art*, SIAM Publications, Philadelphia, 1997, pp. 105–116.

32. R. D. C. Monteiro and J. S. Pang, 'A potential reduction Newton method for constrained equations', manuscript, Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, March 1997.
33. M. R. Hestenes, 'Multiplier and gradient methods', *J. Opt. Theory Appl.*, **4,** 303–320 (1969).
34. M. J. D. Powell, 'A method for non-linear constraints in minimization problems', in R. Fletcher (ed.), *Optimization*, Academic Press, New York, 1969, pp. 283–298.
35. R. T. Rockafellar, 'Augmented Lagrangian multiplier functions and duality in nonconvex programming', *SIAM J. Control*, **12,** 268–285 (1974).
36. P. T. Harker and J. S. Pang, 'A damped-Newton method for the linear complementarity problem', in E. L. Allgower and K. Georg (eds.), *Computational Solution of Nonlinear Systems of Equations*, Lectures in Applied Mathematics, Vol. 26, American Mathematical Society, Providence, RI, 1990, pp. 265–284.
37. A. Klarbring and G. Björkman, 'Solution of large displacement contact problems with friction using Newton's method for generalized equations', *Int. J. Numer. Meth. Engng.*, **34,** 249–269 (1992).

