PERGAMON

International Journal of Solids and Structures 39 (2002) 2323–2341

INTERNATIONAL JOURNAL OF SOLIDS and STRUCTURES

www.elsevier.com/locate/ijsolstr

A semi-smooth Newton method for elasto-plastic contact problems

Peter W. Christensen *

Department of Mechanical Engineering, Linköping University, SE–581 83 Linköping, Sweden

Received 6 January 2001; received in revised form 14 November 2001

Abstract

In this paper we reformulate the frictional contact problem for elasto-plastic bodies as a set of unconstrained, non-smooth equations. The equations are semi-smooth so that Pang's Newton method for B-differentiable equations can be applied. An algorithm based on this method is described in detail. An example demonstrating the efficiency of the algorithm is presented. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Frictional contact; Elasto–plasticity; Semi-smooth equations; Newton method; Radial return

1. Introduction

The design of efficient and robust algorithms for frictional contact problems is an important task in computational solid mechanics. In Klarbring (1992, 1993) the discrete, time-incremental, linear elastic, frictional contact problem was reformulated as a set of unconstrained equations. The reason for using an unconstrained reformulation is to enable Newton methods to be used for solving the equations. Since the equations obtained are non-smooth, a standard Newton method is not applicable. However, it was observed that the equations are at least B-differentiable, so that Pang's Newton method for B-differentiable equations (Pang, 1990) may be used to solve the frictional contact problem. This method was implemented in Strömberg (1997) (including wear), and subsequently in Christensen et al. (1998), and showed excellent performance and robustness. Later, in Christensen and Pang (1998), it was shown that the non-smooth reformulation of the constitutive laws of frictional contact is even semi-smooth, a stronger property than B-differentiability. A general Newton algorithm for constrained semi-smooth equations was developed, which, for unconstrained equations has Pang's Newton method as a special case. Using the semi-smoothness, the convergence results for Pang's Newton method could be strengthened.

In this work we will extend the method in Christensen et al. (1998) and Christensen and Pang (1998) to elasto-plastic frictional contact problems. Only J_2-plasticity will be considered, i.e. the von Mises yield function together with an associated flow rule is assumed. We will use the radial return method to obtain

* Tel.: +46-13-281188; fax: +46-13-281101.
E-mail address: petch@ikp.liu.se (P.W. Christensen).

0020-7683/02/$ - see front matter
PII: S0020-7683(02)00086-0

the Gauss point stresses as functions of the displacements. The standard way this method is employed for plasticity problems, is to insert the stresses into the equilibrium equations and solve these using a standard Newton method, possibly using a heuristic line search in order to improve the robustness, see e.g. Simo and Hughes (1998). However, the stresses are not smooth functions of the displacements, so strictly speaking a standard Newton method cannot be applied. We will, however, pay attention to the non-smooth character of the plasticity equations and use a non-smooth Newton method to solve the governing equations. In Christensen (2002) it was recognized, for the case of linear hardening, that the radial return method yields the stresses as piecewise smooth functions of the displacements. Since piecewise smooth functions are semi-smooth, the plasticity problem may be solved using a semi-smooth Newton method. Since also the contact equations are semi-smooth, it follows that the frictional contact problem for elasto-plastic bodies may be described by semi-smooth equations. Thus, Pang's Newton method is applicable for this problem as well.

The paper is organized as follows: In the next section we present the governing equations for quasi-static, small-strain, rate-independent J_2-plasticity, and the small-displacement frictional contact problem. Time and space discretizations are performed in order to formulate the discrete, time-incremental, elasto-plastic frictional contact problem. In Section 3 the laws of contact are rewritten as a set of unconstrained semi-smooth equations. Section 4 presents a modified semi-smooth Newton method to solve the problem. A numerical example is presented in Section 5, which demonstrates the efficiency of our algorithm.

2. Problem formulation

Consider two two-dimensional bodies loaded in plane strain, $\mathscr{A}$ and $\mathscr{B}$, see Fig. 1. Body $\mathscr{A}$ is a fixed rigid foundation, whereas body $\mathscr{B}$ is elasto-plastic. We assume that all strains and displacements are small. Our aim is to determine stresses, displacements, strains and contact tractions as functions of time for a prescribed loading history when inertia effects may be neglected. The fact that body $\mathscr{A}$ is considered rigid is for notational simplicity only; under the assumption of small displacements there is no significant mathematical difference between problems where both bodies are flexible and those where one of the bodies is rigid. Although we will present the theory for the case of plane strain, generalization to three-dimensional problems is straightforward.

Let body $\mathscr{B}$ occupy the region $\Omega \subset \mathbb{R}^2$ and divide the boundary $\partial\Omega$ in three disjoint parts S_t, S_u and S_c. The body is subjected to prescribed tractions $\tilde{t}$ on S_t, on S_u the displacements are fixed, and S_c represents the

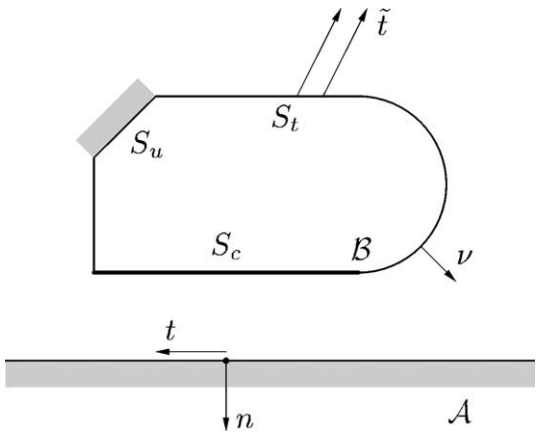

Fig. 1. The two bodies considered.

potential contact surface. To simplify notation, we assume that there are no non-zero prescribed displacements (however, in the numerical example in Section 5, some displacements are in fact prescribed). For each point on the boundary of the foundation $\mathscr{A}$, we define an inward normal direction n and a tangential direction t. The location of a point x is represented in a fixed Cartesian frame, and the components of x are denoted x_i, $i = 1, 2$. For all tensor expressions in component form, the summation convention will be adopted for repeated indices.

2.1. Equilibrium equations

Under the assumption of no body forces, the equilibrium equations for body $\mathscr{B}$ are written as

$$\frac{\partial \sigma_{ij}}{\partial x_j} = 0 \quad \text{in } \Omega, \tag{1}$$

$$u_i = 0 \quad \text{on } S_u, \qquad \sigma_{ij} v_j = \tilde{t}_i \quad \text{on } S_t, \qquad \sigma_{ij} v_j = -p_i \quad \text{on } S_c, \tag{2}$$

where $i, j = 1, 2$, σ is the Cauchy stress, u is the displacement field, v is the outward unit normal of body $\mathscr{B}$ and p is the contact traction. As a consequence of the small-displacement assumption, the surface of the foundation and the contact surface may be regarded as parallel, i.e. $v \approx n$ (Kikuchi and Oden, 1988).

2.2. Constitutive laws of plasticity

Body $\mathscr{B}$ is assumed to be made of an elasto-plastic material with rate-independent behavior, obeying the von Mises yield criterion with an associate flow rule. For simplicity, we assume only linear isotropic hardening. In the elastic domain, linear isotropic behavior is assumed. Thus, the stress σ is related to the elastic strain ϵ^e through Hooke's law

$$\sigma_{ij} = C^e_{ijkl} \epsilon^e_{kl}, \tag{3}$$

where the elasticity tensor C^e is given by

$$C^e_{ijkl} \equiv \tilde{\lambda} \delta_{ij} \delta_{kl} + \tilde{\mu} (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}), \tag{4}$$

where δ_{ij} denotes Kronecker's delta, and $\tilde{\lambda}$ and $\tilde{\mu}$ are the two Lamé constants. The stress tensor may be decomposed into a deviatoric part, $s_{ij} \equiv (\operatorname{dev} \sigma)_{ij} = \sigma_{ij} - (1/3)\sigma_{kk}\delta_{ij}$, and a dilatational part $d_{ij} \equiv (\operatorname{dil} \sigma)_{ij} = (1/3)\sigma_{kk}\delta_{ij}$. From (3) and (4) we obtain

$$d_{ij} = \tilde{\kappa} \epsilon^e_{kk} \delta_{ij}, \tag{5}$$

$$s_{ij} = 2\tilde{\mu} e^e_{ij}, \tag{6}$$

where e^e is the deviatoric elastic strain, and $\tilde{\kappa} \equiv \tilde{\gamma} + (2/3)\tilde{\mu}$ is the bulk modulus. The total strain ϵ is written as the sum of the elastic strain ϵ^e and the plastic strain ϵ^p:

$$\epsilon_{ij} = \epsilon^e_{ij} + \epsilon^p_{ij}, \tag{7}$$

where ϵ is related to u according to

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right). \tag{8}$$

The plastic strain rate is determined from the flow rule

$$\dot{\epsilon}_{ij}^{\mathrm{p}} = \dot{\gamma}\frac{s_{ij}}{||s||}, \tag{9}$$

where the dot denotes differentiation with respect to time, and $||s|| \equiv \sqrt{s_{ij}s_{ij}}$. Plastic flow cannot occur in the elastic domain $\mathscr{E} \equiv \{(\sigma,\alpha) : f(\sigma,\alpha) < 0)\}$, where α is a hardening variable and f is the yield function

$$f(\sigma,\alpha) \equiv ||s|| - \sqrt{\frac{2}{3}}K(\alpha).$$

For linear isotropic hardening, K is written

$$K(\alpha) \equiv \sigma_y + k\alpha, \tag{10}$$

where σ_y is the yield stress and $k \geqslant 0$. The yield function and the flow parameter $\dot{\gamma}$ satisfy a set of complementarity conditions

$$f(\sigma,\alpha) \leqslant 0, \qquad \dot{\gamma} \geqslant 0, \qquad f(\sigma,\alpha)\dot{\gamma} = 0. \tag{11}$$

The hardening parameter α is chosen to be the equivalent plastic strain, so that its evolution is governed by

$$\dot{\alpha} = \dot{\gamma}\sqrt{\frac{2}{3}}. \tag{12}$$

2.3. Constitutive laws of contact

The laws of contact are of two types; normal contact laws and friction laws. The normal contact law used here is that of Signorini:

$$p_{\mathrm{n}} \geqslant 0, \qquad u_{\mathrm{n}} \leqslant g, \qquad p_{\mathrm{n}}(u_{\mathrm{n}} - g) = 0 \quad \text{on } S_{\mathrm{c}}, \tag{13}$$

where the normal contact pressure $p_{\mathrm{n}} = p_i n_i$, the normal displacement $u_{\mathrm{n}} = u_i n_i$ on S_{c}, and g is the initial gap between bodies $\mathscr{A}$ and $\mathscr{B}$. We may rewrite (13) as

$$p_{\mathrm{n}} \in \mathscr{K}_{\mathrm{n}} : (u_{\mathrm{n}} - g)(q_{\mathrm{n}} - p_{\mathrm{n}}) \leqslant 0 \quad \forall q_{\mathrm{n}} \in \mathscr{K}_{\mathrm{n}}, \tag{14}$$

where

$$\mathscr{K}_{\mathrm{n}} \equiv \{p_{\mathrm{n}} : p_{\mathrm{n}} \geqslant 0\}.$$

As the friction law we choose Coulomb's law which can be stated using the principle of maximum dissipation:

$$p_{\mathrm{t}} \in \mathscr{F}(\mu(p_{\mathrm{n}})_+) : \dot{\boldsymbol{u}}_{\mathrm{t}}(q_{\mathrm{t}} - p_{\mathrm{t}}) \leqslant 0 \quad \forall q_{\mathrm{t}} \in \mathscr{F}(\mu(p_{\mathrm{n}})_+), \tag{15}$$

where the tangential contact traction $p_{\mathrm{t}} = p_i t_i$, the tangential displacement $u_{\mathrm{t}} = u_i t_i$ on S_{c}, μ is the friction coefficient,

$$\mathscr{F}(z) \equiv \{p_{\mathrm{t}} : |p_{\mathrm{t}}| \leqslant z\}, \quad \text{for } z \geqslant 0,$$

and $z_+ \equiv \max(0,z)$.

2.4. Time discretization

The strong form of the quasi-static frictional contact problem for elasto-plastic bodies may now be written in the following way. For a given loading history $\tau \mapsto \tilde{t}(\tau)$ on a time interval $\tau \in [0,\mathscr{T}]$, find the

functions $\tau \mapsto u(x,\tau)$, $\tau \mapsto \epsilon(x,\tau)$, $\tau \mapsto \sigma(x,\tau)$, $\tau \mapsto \epsilon^{\mathrm{e}}(x,\tau)$, $\tau \mapsto \epsilon^{\mathrm{p}}(x,\tau)$, $\tau \mapsto \alpha(x,\tau)$, $\tau \mapsto p_{\mathrm{n}}(x_{\mathrm{c}},\tau)$ and $\tau \mapsto p_{\mathrm{t}}(x_{\mathrm{c}},\tau)$ for all $x \in \mathscr{B}$ and $x_{\mathrm{c}} \in S_{\mathrm{c}}$, such that (1)–(3), (7)–(9), (11), (12), (14) and (15) are satisfied for all $\tau \in [0,\mathscr{T}]$.

In order to solve this problem numerically, we have to perform discretizations in space and time. The space discretization will be discussed in Section 2.5, whereas the time discretizations of the constitutive laws of plasticity and frictional contact are discussed next.

2.4.1. Time discretization of the plasticity equations, radial return

Let the time interval $[0,\mathscr{T}]$ be divided into sub-intervals. For each sub-interval, $[\tau_k,\tau_{k+1}]$, say, the variables u, σ, ϵ^{p}, α, p_{n} and p_{t} are assumed to be known at time τ_k. The plastic strain rate is approximated by a backward Euler time discretization according to

$$\dot{\epsilon}^{\mathrm{p}}_{ij}(\tau_{k+1}) \approx \frac{\epsilon^{\mathrm{p}}_{ij} - \bar{\epsilon}^{\mathrm{p}}_{ij}}{\Delta\tau},$$

where $\Delta\tau \equiv \tau_{k+1} - \tau_k$, $\epsilon^{\mathrm{p}}_{ij} \equiv \epsilon^{\mathrm{p}}_{ij}(\tau_{k+1})$ and $\bar{\epsilon}^{\mathrm{p}}_{ij} \equiv \epsilon^{\mathrm{p}}_{ij}(\tau_k)$. Similarly, we obtain for $\dot{\alpha}$ and $\dot{\gamma}$

$$\dot{\alpha}(\tau_{k+1}) \approx \frac{\alpha - \bar{\alpha}}{\Delta\tau}, \qquad \dot{\gamma}(\tau_{k+1}) \approx \frac{\Delta\gamma}{\Delta\tau},$$

where $\Delta\gamma \equiv \gamma(\tau_{k+1}) - \gamma(\tau_k)$.

Insertion of these time discretized expressions into the flow rule (9) and (12) makes it possible to obtain the stress as a function of the displacement using the so-called radial return method (Simo and Hughes, 1998). Eqs. (9) and (12) in time-discretized form become

$$\epsilon^{\mathrm{p}}_{ij} = \bar{\epsilon}^{\mathrm{p}}_{ij} + \Delta\gamma \frac{s_{ij}}{||s||}, \tag{16}$$

$$\alpha = \bar{\alpha} + \sqrt{\frac{2}{3}}\Delta\gamma. \tag{17}$$

An elastic trial deviatoric stress is obtained from (6) and (7) by assuming that no plastic flow occurs during the increment under investigation, i.e. $\epsilon^{\mathrm{p}} = \bar{\epsilon}^{\mathrm{p}}$:

$$s^{\mathrm{tr}}_{ij} \equiv 2\tilde{\mu}(e_{ij} - \bar{\epsilon}^{\mathrm{p}}_{ij}), \tag{18}$$

where e_{ij} is the deviatoric total strain, and we have also used the fact that the plastic strain is deviatoric, cf. (9). The real deviatoric stress may then be obtained from (18) and (16) as

$$s_{ij} = 2\tilde{\mu}(e_{ij} - \epsilon^{\mathrm{p}}_{ij}) = 2\tilde{\mu}(e_{ij} - \bar{\epsilon}^{\mathrm{p}}_{ij}) - 2\tilde{\mu}(\epsilon^{\mathrm{p}}_{ij} - \bar{\epsilon}^{\mathrm{p}}_{ij}) = s^{\mathrm{tr}}_{ij} - 2\tilde{\mu}\Delta\gamma\hat{n}_{ij}, \tag{19}$$

where $\hat{n}_{ij} \equiv s_{ij}/||s||$. It remains to determine the flow parameter $\Delta\gamma$. If $||s^{\mathrm{tr}}|| \leqslant \sqrt{\frac{2}{3}}K(\bar{\alpha})$, then $\Delta\gamma = 0$, as the assumption of elastic response is correct in this case. Otherwise, $\Delta\gamma$ can be obtained by using the condition $\|s\| - \sqrt{\frac{2}{3}}K\left(\bar{\alpha} + \sqrt{\frac{2}{3}}\Delta\gamma\right) = 0$. It follows from (19) that s and s^{tr} are colinear, i.e. $\hat{n}_{ij} = s^{\mathrm{tr}}_{ij}/||s^{\mathrm{tr}}||$, which inserted into (19) gives

$$s_{ij} = (||s^{\mathrm{tr}}|| - 2\tilde{\mu}\Delta\gamma)\hat{n}_{ij}.$$

If this equation is (scalar) multiplied with $\hat{n}_{ij}$, we obtain

$$||s|| = ||s^{\mathrm{tr}}|| - 2\tilde{\mu}\Delta\gamma,$$

where $||s|| = \sqrt{\frac{2}{3}}K\left(\bar{\alpha} + \sqrt{\frac{2}{3}}\Delta\gamma\right)$. Solving for $\Delta\gamma$, we obtain

$$\Delta\gamma = \frac{||s^{\mathrm{tr}}|| - \sqrt{\frac{2}{3}}K(\bar{\alpha})}{2\tilde{\mu} + \frac{2}{3}k}.$$

The total stress may now be obtained as $\sigma_{ij} = d_{ij} + s_{ij}$, where the dilatational stress, d_{ij}, is given in (5):

$$\sigma_{ij} = \begin{cases} \tilde{\kappa}\epsilon_{kk}\delta_{ij} + s^{\mathrm{tr}}_{ij} & \text{if } ||s^{\mathrm{tr}}|| \leqslant \sqrt{\frac{2}{3}}K(\bar{\alpha}), \\ \tilde{\kappa}\epsilon_{kk}\delta_{ij} + \frac{1}{2\tilde{\mu}+\frac{2}{3}k}\left(\frac{2}{3}ks^{\mathrm{tr}}_{ij} + 2\tilde{\mu}\sqrt{\frac{2}{3}}K(\bar{\alpha})\frac{s^{\mathrm{tr}}_{ij}}{||s^{\mathrm{tr}}||}\right) & \text{otherwise.} \end{cases} \tag{20}$$

It is noted that σ is a function of the displacements since s^{tr} depends only on the total deviatoric strain, and the strain in turn is obtained from the displacement from (8).

2.4.2. Time discretization of Coulomb's law

The tangential velocity, $\dot{\boldsymbol{u}}_{\mathrm{t}}$, is also approximated by a backward Euler time discretization:

$$\dot{\boldsymbol{u}}_{\mathrm{t}}(\tau_{k+1}) \approx \frac{u_{\mathrm{t}} - \bar{\boldsymbol{u}}_{\mathrm{t}}}{\Delta\tau},$$

where $u_{\mathrm{t}} \equiv u_{\mathrm{t}}(\tau_{k+1})$ and $\bar{\boldsymbol{u}}_{\mathrm{t}} \equiv u_{\mathrm{t}}(\tau_k)$. Coulomb's law (15) may then be written in time discretized form as

$$p_{\mathrm{t}} \in \mathscr{F}(\mu(p_{\mathrm{n}})_+) : (u_{\mathrm{t}} - \bar{\boldsymbol{u}}_{\mathrm{t}})(q_{\mathrm{t}} - p_{\mathrm{t}}) \leqslant 0 \quad \forall q_{\mathrm{t}} \in \mathscr{F}(\mu(p_{\mathrm{n}})_+).$$

2.5. Finite element approximation

A variational formulation of the incremental elasto-plastic frictional contact problem with the displacements and contact tractions as variables reads as follows: Given $\bar{\epsilon}$, $\bar{\alpha}$, $\bar{\boldsymbol{u}}_{\mathrm{t}}$ and $\tilde{\boldsymbol{t}}(\tau_{k+1})$, find u, p_{n} and p_{t} that satisfy

$$\int_\Omega \frac{\partial v_i}{\partial x_j}\sigma_{ij}\,\mathrm{d}V + \int_{S_{\mathrm{c}}} v_i p_i\,\mathrm{d}S - \int_{S_{\mathrm{t}}} v_i \tilde{t}_i\,\mathrm{d}S = 0 \quad \forall v \in \mathscr{V}, \tag{21}$$

$$\int_{S_{\mathrm{c}}} (u_{\mathrm{n}} - g)(q_{\mathrm{n}} - p_{\mathrm{n}})\,\mathrm{d}S \leqslant 0 \quad \forall q_{\mathrm{n}} \in \mathscr{K}_{\mathrm{n}}, \tag{22}$$

$$\int_{S_{\mathrm{c}}} (u_{\mathrm{t}} - \bar{\boldsymbol{u}}_{\mathrm{t}})(q_{\mathrm{t}} - p_{\mathrm{t}})\,\mathrm{d}S \leqslant 0 \quad \forall q_{\mathrm{t}} \in \mathscr{F}(\mu(p_{\mathrm{n}})_+), \tag{23}$$

where

$$\mathscr{V} \equiv \{v : v = 0 \quad \text{on } S_u\},$$

and $\sigma = \sigma(u)$ satisfies (20).

The finite element discretization of the first and last terms in (21) is standard and will not be discussed here, see (1987) for a detailed account. The integrals over the potential contact surface are approximated as

$$\int_{S_{\mathrm{c}}} h\,\mathrm{d}S \approx \sum_{m=1}^{n_{\mathrm{c}}} I^m h(x^m), \tag{24}$$

where I^m are weighting factors and n_{c} is the number of integration points x^m. We choose x^m to coincide with the finite element nodes of body $\mathscr{B}$ on S_{c}, i.e. n_{c} is the number of nodes on S_{c}, and let the weights I^m be determined by the trapezoidal rule for linear elements and Simpson's rule for quadratic elements. Application of the space discretization results in the following discretized version of the equilibrium equations (21):

$$H^{\mathrm{eq}}(U, P_{\mathrm{n}}, P_{\mathrm{t}}) \equiv F^{\mathrm{int}} + C_{\mathrm{n}}^{\mathrm{T}} P_{\mathrm{n}} + C_{\mathrm{t}}^{\mathrm{T}} P_{\mathrm{t}} - F^{\mathrm{ext}} = 0, \tag{25}$$

where the vector of nodal displacements is denoted $U \in \mathbb{R}^{n_d}$, and n_d is the total number of degrees of freedom in the finite element mesh. The matrices $C_n \in \mathbb{R}^{n_c \times n_d}$ and $C_t \in \mathbb{R}^{n_c \times n_d}$ are transformation matrices relating the normal, U_n, and tangential displacements, U_t, for the contact nodes, respectively:

$$U_n = C_n U, \qquad U_t = C_t U.$$

The normal and tangential contact forces are denoted P_n and P_t, respectively:

$$P_n \equiv [P_n^1, \ldots, P_n^{n_c}]^T, \qquad P_n^m \equiv I^m p_n(x^m), \qquad m = 1, \ldots, n_c,$$

$$P_t \equiv [P_t^1, \ldots, P_t^{n_c}]^T, \qquad P_t^m \equiv I^m p_t(x^m), \qquad m = 1, \ldots, n_c,$$

where T denotes the transpose of a matrix. The forces F^{int} and F^{ext} are obtained from

$$F^{int} \equiv \mathop{\mathbf{A}}_{e=1}^{n_e} f_e^{int}, \qquad F^{ext} \equiv \mathop{\mathbf{A}}_{e=1}^{n_e} f_e^{ext},$$

where $\mathbf{A}$ denotes the finite element assembly operator. The element external load f_e^{ext} is obtained from the traction $\tilde{t}$ in a standard manner, see Hughes (1987), and the element internal load is written

$$f_e^{int} \equiv \sum_{i=1}^{n_{i,e}} (\overline{B}^{i,e})^T \sigma^{i,e} w^{i,e} J^{i,e},$$

where $n_{i,e}$ is the number of Gauss points in element e, and $\overline{B}^{i,e}$ is the B-bar matrix evaluated at Gauss point i (in natural coordinates) of element e. The B-bar matrix replaces the standard strain displacement matrix in order to prevent locking due to the incompressibility induced by plastic deformation. In our implementation we have used the mean dilatation formulation; see Hughes (1987) for an explicit expression of the B-bar matrix. The approximated stress obtained from (20) at Gauss point i of element e is denoted $\sigma^{i,e} = [\sigma_{11}^{i,e}, \sigma_{22}^{i,e}, \sigma_{33}^{i,e}, \sigma_{12}^{i,e}]^T$, $w^{i,e}$ is the integration point weight at Gauss point i of element e, and $J^{i,e}$ is the determinant of the Jacobian for the transformation from the global system to the local system of element e evaluated at Gauss point i. For each Gauss point i of element e, the strain is determined from the element displacement vector $U^e \in \mathbb{R}^{n_{de}}$, where n_{de} is the number of degrees of freedom for element e, as

$$\epsilon^{i,e} = \overline{B}^{i,e} U^e,$$

where $\epsilon^{i,e} = [\epsilon_{11}^{i,e}, \epsilon_{22}^{i,e}, \epsilon_{33}^{i,e}, 2\epsilon_{12}^{i,e}]^T$.

The approximation of Signorini's law (22) takes the form

$$(U_n - G)^T (Q_n - P_n) \leqslant 0 \quad \forall Q_n \in \mathscr{K}_n^h, \tag{26}$$

where $G \equiv [g^1, \ldots, g^{n_c}]^T$, with $g^m \equiv g(x^m)$, and $\mathscr{K}_n^h \equiv \{P_n : P_n^m \geqslant 0, m = 1, \ldots, n_c\}$. Coulomb's law (23) is approximated as

$$(U_t - \overline{U}_t)^T (Q_t - P_t) \leqslant 0 \quad \forall Q_t \in \mathscr{F}^h(\mu(P_n)_+), \tag{27}$$

where $\mathscr{F}^h(\mu(P_n)_+) \equiv \{Q_t : |Q_t^m| \leqslant \mu(P_n^m)_+, m = 1, \ldots, n_c\}$. The discrete, time-discretized, elasto-plastic frictional contact problem is now governed by (25)–(27). Our next aim is to rewrite (26) and (27) as equalities so that a Newton method can be applied to solve the problem.

3. Formulation as a system of semi-smooth equations

The reformulation of the constitutive laws of frictional contact as a set of unconstrained equations has been thoroughly described in Christensen et al. (1998) and Christensen and Pang (1998) for the three-dimensional case. First, we note that Signorini's law (26) is equivalent to

$$H_{\mathrm{n}}(U,P_{\mathrm{n}}) \equiv \min(P_{\mathrm{n}}, \rho_{\mathrm{n}} \circ (G - U_{\mathrm{n}})) = 0, \tag{28}$$

where $\rho_{\mathrm{n}} = [\rho_{\mathrm{n}}^1,\ldots,\rho_{\mathrm{n}}^{n_{\mathrm{c}}}]^{\mathrm{T}}$ is a vector of given positive scalars, $a \circ b$ is the vector whose components are $a_i b_i$ (no summation), and the minimum operator is applied componentwise. Coulomb's law (27) is easily rewritten as

$$(P_{\mathrm{t}} + r \circ (U_{\mathrm{t}} - \overline{U}_{\mathrm{t}}) - P_{\mathrm{t}})^{\mathrm{T}}(Q_{\mathrm{t}} - P_{\mathrm{t}}) \leqslant 0 \quad \forall Q_{\mathrm{t}} \in \mathscr{F}^h(\mu(P_{\mathrm{n}})_+),$$

where $r = [r^1,\ldots,r^{n_{\mathrm{c}}}]^{\mathrm{T}}$ is a vector of given positive scalars. This variational inequality may be written in terms of a projection as, see Hiriart-Urruty and Lemaréchal (1993)

$$H_{\mathrm{t}}(U,P_{\mathrm{n}},P_{\mathrm{t}}) \equiv P_{\mathrm{t}} - \Pi(P_{\mathrm{t}}(r)) = 0, \tag{29}$$

where $P_{\mathrm{t}}(r) \equiv P_{\mathrm{t}} + r \circ (U_{\mathrm{t}} - \overline{U}_{\mathrm{t}})$, and $\Pi(y)$ denotes the Euclidian projection of y onto $\mathscr{F}^h(\mu(P_{\mathrm{n}})_+)$. The projection $\Pi(P_{\mathrm{t}}(r)) = \{\Pi^m(P_{\mathrm{t}}^m(r^m)), m = 1,\ldots,n_{\mathrm{c}}\}$, and for each $m \in \{1,\ldots,n_{\mathrm{c}}\}$

$$\Pi^m\left(P_{\mathrm{t}}^m(r^m)\right) = \begin{cases} P_{\mathrm{t}}^m(r^m) & \text{if } |P_{\mathrm{t}}^m(r^m)| \leqslant \mu(P_{\mathrm{n}}^m)_+, \\ \mu(P_{\mathrm{n}}^m)_+ \frac{P_{\mathrm{t}}^m(r^m)}{|P_{\mathrm{t}}^m(r^m)|} & \text{otherwise,} \end{cases}$$

which in turn can be written more compactly as

$$\Pi^m(P_{\mathrm{t}}^m(r^m)) \equiv \min\left(\frac{\mu(P_{\mathrm{n}}^m)_+}{|P_{\mathrm{t}}^m(r^m)|}, 1\right) P_{\mathrm{t}}^m(r^m),$$

where 0/0 is defined to be 1. Our discrete, time-incremental, elasto-plastic frictional contact problem is thus equivalent to the following unconstrained system of non-smooth equations:

$$0 = H(U,P_{\mathrm{n}},P_{\mathrm{t}}) \equiv \begin{pmatrix} F^{\mathrm{int}} + C_{\mathrm{n}}^{\mathrm{T}} P_{\mathrm{n}} + C_{\mathrm{t}}^{\mathrm{T}} P_{\mathrm{t}} - F^{\mathrm{ext}} \\ \min(P_{\mathrm{n}}, \rho_{\mathrm{n}} \circ (G - U_{\mathrm{n}})) \\ P_{\mathrm{t}} - \Pi(P_{\mathrm{t}}(r)) \end{pmatrix}. \tag{30}$$

In Christensen and Pang (1998) it was proven that H_{n} and H_{t} are semi-smooth mappings. A function $H : \mathbb{R}^n \to \mathbb{R}^n$ is semi-smooth at $\bar{z} \in \mathbb{R}^n$ if it is Lipschitz continuous in a neighborhood of $\bar{z}$, directionally differentiable at $\bar{z}$, and the directional derivative is sufficiently well behaved in that it satisfies the following condition (Qi, 1993):

$$\lim_{h \to 0} \frac{H'(\bar{z}+h;h) - H'(\bar{z};h)}{||h||} = 0,$$

where $H'(\bar{z};h)$ is the directional derivative at $\bar{z}$ in the direction h.

In Christensen (2002), it was recognized that the radial return mapping in (20) is piecewise smooth (Pang and Qi, 1995) which implies that it is also semi-smooth (Chaney, 1990). Consequently, H^{eq} is semi-smooth, and, thus, H is semi-smooth. We remark that for linear elastic contact problems, H^{eq} is an affine function, so that the non-smoothness of H then only stems from the contact equations. In contrast, for the elasto-plastic contact problem studied here, the equilibrium equations are non-smooth as well, since H^{eq} is a non-smooth function of the displacements.

4. A semi-smooth Newton method

During the last 10–15 years, a large number of Newton methods for non-smooth equations have evolved. Naturally, one tries to establish strong convergence results under assumptions which are as weak as possible. For a non-smooth function $F : \mathbb{R}^n \to \mathbb{R}^n$ which is only locally Lipschitz continuous, one cannot establish any convergence results for a Newton method. If F is assumed to be B-differentiable, i.e. locally

Lipschitz continuous and directionally differentiable, then it is possible to design a Newton method with proven convergence properties. In Pang (1990), a damped Newton method for B-differentiable equations was proposed, algorithm BN. Global convergence and local quadratic convergence to strongly F-differentiable (Ortega and Rheinboldt, 1970) solution points was established. In order to be able to prove convergence even to non-smooth solution points, however, F needs to be semi-smooth. We will use Pang's algorithm BN with improved convergence results for semi-smooth equations to solve the semi-smooth system in (30) for a specific time increment. We refer to Christensen and Pang (1998) for a discussion of the convergence results of this algorithm. In short, quadratic convergence to solution points which are strongly F-differentiable with a non-singular Jacobian and a Lipschitz continuous directional derivative, and linear convergence to so-called BD-regular solution points (Qi, 1993) is established. A step-by-step description of algorithm BN is given below. We let z denote the tuple $(U, P_{\mathrm{n}}, P_{\mathrm{t}})$, and $\mathrm{d}z$ the direction $(\mathrm{d}U, \mathrm{d}P_{\mathrm{n}}, \mathrm{d}P_{\mathrm{t}})$.

4.1. Description of algorithm BN

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

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

$$H'(z^k; \mathrm{d}z^k) = -H(z^k). \tag{31}$$

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

$$\Theta(z^k + \tilde{\beta}^m \mathrm{d}z^k) \leqslant (1 - 2\tilde{\sigma}\tilde{\beta}^m)\Theta(z^k), \tag{32}$$

where the merit function Θ is obtained as

$$\Theta(z^k) \equiv \frac{1}{2} H(z^k)^{\mathrm{T}} H(z^k).$$

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

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

Evidently, the only difference between algorithm BN and a standard damped Newton method for smooth equations lies in the generation of a search direction. At differentiable points z^k, the directional derivative satisfies $H'(z^k; \mathrm{d}z^k) = \nabla H(z^k)\,\mathrm{d}z^k$ where $\nabla H(z^k)$ denotes the Jacobian of H at z^k. Thus, at differentiable points, the directional Newton equation (31) is linear in the search direction $\mathrm{d}z^k$. However, at non-differentiable points, where the Jacobian does not exist, the directional derivative is non-linear in the search direction; see Christensen et al. (1998) for the rather complex expression of the directional derivative for the elastic contact problem. Thus, a drawback with algorithm BN is that, for non-differentiable points, a system of non-linear equations has to be solved. For our application, the non-differentiable points are points with $P_{\mathrm{n}}^m = 0$, $P_{\mathrm{n}}^m + \rho_{\mathrm{n}}^m(C_{\mathrm{n}}^m U - g^m) = 0$, $|P_{\mathrm{t}}^m(r^m)| = \mu P_{\mathrm{n}}^m > 0$, and $||s^{\mathrm{tr}}|| = \sqrt{\frac{2}{3}} K(\bar{\alpha})$. Hence, the non-differentiable states are all described by equalities. Consequently, the chance that an iterate ends up at a non-differentiable point is very small. This motivates us to use a simplified version of algorithm BN, where the directional derivative $H'(z^k; \mathrm{d}z^k)$ is replaced with $\widetilde{H}'(z^k; \mathrm{d}z^k) = \Xi(z^k)\mathrm{d}z^k$, so that $\widetilde{H}'(z^k; \mathrm{d}z^k)$ is linear in $\mathrm{d}z^k$, and thus only a system of linear equations has to be solved in order to obtain the search direction $\mathrm{d}z^k$. This modification has previously been used to a very good effect for elastic contact problems in Strömberg (1997), Christensen et al. (1998) and Christensen and Pang (1998), and it is our experience that it works equally well for the elasto-plastic contact problems studied here, see Section 5. If (31) is solved exactly, the

search direction obtained is guaranteed to be a descent direction for the merit function. However, with our simplified directional Newton equation, the search direction may not be a descent direction for the merit function in the event that an iterate does end up at a non-differentiable point. Hence, the line search in (32) might jam. In our implementation of the line search, see the pseudo-code below, we use a lower bound on the step size which prevents the line search from jamming. This implementation has proven to work well in our extensive testing for elastic contact problems and also for the elasto-plastic contact problems in this work. We would like to point out that even if non-differentiable points are hit, e.g. at the first iteration of an increment if the starting point satisfies $P_{\mathrm{n}}^{m}=0$ for some contact node m, the performance of the algorithm does not deteriorate. The main practical reason for adopting this lower bound on the step size is that even for differentiable points, the step size which satisfies (32) might be so small that it is computationally more efficient to accept a larger step size.

Explicitly, the directional derivative of H^{eq} is replaced by the following simplified expression, linear in $\mathrm{d}U$ (and $\mathrm{d}P_{\mathrm{n}}$ and $\mathrm{d}P_{\mathrm{t}}$):

$$\widetilde{H}^{\mathrm{eq}'} \equiv K\,\mathrm{d}U + C_{\mathrm{n}}^{\mathrm{T}}\,\mathrm{d}P_{\mathrm{n}} + C_{\mathrm{t}}^{\mathrm{T}}\,\mathrm{d}P_{\mathrm{t}},$$

where

$$K \equiv \mathop{\mathbf{A}}_{e=1}^{n_e}\sum_{i=1}^{n_{i,e}}(\overline{B}^{i,e})^{\mathrm{T}}C_{\mathrm{ep}}^{i,e}\overline{B}^{i,e}w^{i,e}J^{i,e},$$

and $C_{\mathrm{ep}}^{i,e}$ is the consistent elasto-plastic tangent modulus, obtained by differentiating the stress in (20) with respect to strain. The result is (omitting Gauss point and element number), see Simo and Hughes (1998):

$$C_{\mathrm{ep}} = \begin{cases} \tilde{\kappa}\tilde{1}\tilde{1}^{\mathrm{T}} + 2\tilde{\mu}(I - \frac{1}{3}\tilde{1}\tilde{1}^{\mathrm{T}}) & \text{if } ||s^{\mathrm{tr}}|| \leqslant \sqrt{\frac{2}{3}}K(\bar{\alpha}), \\ \tilde{\kappa}\tilde{1}\tilde{1}^{\mathrm{T}} + 2\tilde{\mu}\beta_1(I - \frac{1}{3}\tilde{1}\tilde{1}^{\mathrm{T}}) - 2\tilde{\mu}\beta_2\hat{n}\hat{n}^{\mathrm{T}} & \text{otherwise,} \end{cases}$$

where

$$\beta_1 \equiv 1 - \frac{2\tilde{\mu}\Delta\gamma}{||s^{\mathrm{tr}}||}, \qquad \beta_2 \equiv \frac{2\tilde{\mu}}{2\tilde{\mu} + \frac{2}{3}k} - (1-\beta_1),$$

and $\tilde{1} \equiv [1,1,1,0]^{\mathrm{T}}$, $\hat{n} = [\hat{n}_{11}, \hat{n}_{22}, \hat{n}_{33}, \hat{n}_{12}]^{\mathrm{T}}$, $s^{\mathrm{tr}} = [s_{11}^{\mathrm{tr}}, s_{22}^{\mathrm{tr}}, s_{33}^{\mathrm{tr}}, s_{12}^{\mathrm{tr}}]^{\mathrm{T}}$, and $I \equiv \mathrm{diag}[1,1,1,\frac{1}{2}]$. Thus, for the non-differentiable points where $||s^{\mathrm{tr}}|| = \sqrt{\frac{2}{3}}K(\bar{\alpha})$, C_{ep} is simply put to the Hooke's (elasticity) matrix.

Before presenting the approximations of H_{n}' and H_{t}', we note that the contact equations obviously only depend on the displacements of the contact nodes, U_{c}, and not on the displacements of the other nodes, U_{o}. It is therefore advantageous to eliminate $\mathrm{d}U_{\mathrm{o}}$ from the equation

$$\widetilde{H}^{\mathrm{eq}'}(z^k; \mathrm{d}z^k) = -H^{\mathrm{eq}}(z^k), \tag{33}$$

by means of static condensation, before determining $(\mathrm{d}U_{\mathrm{c}}, \mathrm{d}P_{\mathrm{n}}, \mathrm{d}P_{\mathrm{t}})$. First, we note that we may write $U_{\mathrm{n}} = C_{\mathrm{n}}U = \overline{C}_{\mathrm{n}}U_{\mathrm{c}}$, where $\overline{C}_{\mathrm{n}} \in \mathbb{R}^{n_{\mathrm{c}} \times 2n_{\mathrm{c}}}$ is obtained from C_{n} by deleting (zero) columns corresponding to nodes not on the contact surface. Introducing the matrix $\overline{C}_{\mathrm{t}}$ in a similar manner, (33) may be written

$$\begin{pmatrix} K_{\mathrm{o,o}} & K_{\mathrm{o,c}} \\ K_{\mathrm{o,c}}^{\mathrm{T}} & K_{\mathrm{c,c}} \end{pmatrix}\begin{pmatrix} \mathrm{d}U_{\mathrm{o}} \\ \mathrm{d}U_{\mathrm{c}} \end{pmatrix} + \begin{pmatrix} 0 \\ \overline{C}_{\mathrm{n}}^{\mathrm{T}}\mathrm{d}P_{\mathrm{n}} \end{pmatrix} + \begin{pmatrix} 0 \\ \overline{C}_{\mathrm{t}}^{\mathrm{T}}\mathrm{d}P_{\mathrm{t}} \end{pmatrix} = -\begin{pmatrix} H_{\mathrm{o}}^{\mathrm{eq}} \\ H_{\mathrm{c}}^{\mathrm{eq}} \end{pmatrix}. \tag{34}$$

Eliminating $\mathrm{d}U_{\mathrm{o}}$ from the above, results in

$$K_{\mathrm{c}}\,\mathrm{d}U_{\mathrm{c}} + \overline{C}_{\mathrm{n}}^{\mathrm{T}}\,\mathrm{d}P_{\mathrm{n}} + \overline{C}_{\mathrm{t}}^{\mathrm{T}}\,\mathrm{d}P_{\mathrm{t}} = -F_{\mathrm{c}}, \tag{35}$$

where

$$\begin{aligned} K_c &= K_{c,c} - K_{o,c}^T K_{o,o}^{-1} K_{o,c}, \\ F_c &= H_c^{eq} - K_{o,c}^T K_{o,o}^{-1} H_o^{eq}. \end{aligned}$$

Once dU_c has been calculated (see below), dU_o is obtained as

$$dU_o = -K_{o,o}^{-1}(H_o^{eq} + K_{o,c}\, dU_c). \tag{36}$$

Note that (34) is not to be interpreted literally; there is no need to partition the nodes with contact nodes last. In fact, if a profile (skyline) solver (Felippa, 1975) is used to perform the condensation this would increase the size of the profile with larger memory requirement and longer computing times as a result.

Returning to the simplified directional derivatives of the mappings H_n and H_t, we first let $\overline{C}_n^m$ denote row m of $\overline{C}_n$, $\overline{C}_t^m$ row m of $\overline{C}_t$, and define the following sets:

$$\begin{aligned} \mathscr{I}_{\leqslant} &\equiv \{m : P_n^m + \rho_n^m(\overline{C}_n^m U_c - g^m) \leqslant 0\}, \\ \mathscr{I}_{>} &\equiv \{m : P_n^m + \rho_n^m(\overline{C}_n^m U_c - g^m) > 0\}, \\ \mathscr{J}_{\leqslant} &\equiv \{m : P_n^m > 0,\ |P_t^m(r^m)| \leqslant \mu P_n^m\}, \\ \mathscr{J}_{>} &\equiv \{m : P_n^m > 0,\ |P_t^m(r^m)| > \mu P_n^m\}, \\ \mathscr{K} &\equiv \{m : P_n^m \leqslant 0\}, \end{aligned}$$

where

$$P_t^m(r^m) \equiv P_t^m + r^m(\overline{C}_t^m U_c - \overline{C}_t^m \overline{U}_c).$$

Let z_c denote the tuple (U_c, P_n, P_t) and dz_c the direction (dU_c, dP_n, dP_t). The simplified expression for directional derivatives of the condensed equilibrium equations together with the contact equations reads

$$\widetilde{H}_c'(z_c; dz_c) \equiv \begin{pmatrix} K_c\, dU_c + \overline{C}_n^T\, dP_n + \overline{C}_t^T\, dP_t \\ (dP_n^m)_{m\in\mathscr{I}_{\leqslant}} \\ (-\rho_n^m \overline{C}_n^m\, dU_c)_{m\in\mathscr{I}_{>}} \\ (dP_t^m)_{m\in\mathscr{K}} \\ (-r^m \overline{C}_t^m\, dU_c)_{m\in\mathscr{J}_{\leqslant}} \\ \left[dP_t^m - \mu\, \mathrm{sgn}(P_t^m(r^m))\, dP_n^m\right]_{m\in\mathscr{J}_{>}} \end{pmatrix}, \tag{37}$$

where $\mathrm{sgn}(y)$ denotes the sign of y, i.e. $+1$ if $y > 0$, -1 if $y < 0$ and 0 if $y = 0$. We refer to Christensen et al. (1998) and Christensen and Pang (1998) for the exact expressions of H_n' and H_t'. In every iteration, we solve the following system of linear equations for (dU_c, dP_n, dP_t):

$$\widetilde{H}_c'(z_c; dz_c) = -H_c(z_c), \tag{38}$$

where $H_c(z_c) \equiv [F_c(z_c), H_n(z_c), H_t(z_c)]^T$. The solvability of this equation (for the elastic case) is discussed in Christensen and Pang (1998); if K_c is positive definite and μ is sufficiently small, then (38) has a unique solution.

For convenience, we next summarize our algorithm to solve the discrete, time-incremental, elasto-plastic frictional contact problem in pseudo-code. In the description of the radial return method, algorithm RR, subscripts indicating the Gauss point number have been dropped, and $\mathrm{Tr}\, w = w_{11} + w_{22} + w_{33}$. The computation of the consistent tangent modulus C_{ep} is of course only performed prior to solving (38) in each iteration, and not for every line search. In the description of the modified B-differentiable Newton method, algorithm BNM, subscript I indicates the Gauss point number, of which there are n_I in the whole finite element mesh.

Algorithm RR (*radial return*)

Element displacements U^{e}, $\bar{\alpha}$ and $\bar{\epsilon}^{\mathrm{p}}$ given.
Compute total strain:
$\epsilon = \overline{B}^{\mathrm{e}} U^{\mathrm{e}}$
Compute trial stress:
$e = \epsilon - \frac{1}{3}\mathrm{Tr}\,\epsilon\tilde{1}$
$s^{\mathrm{tr}} = 2\tilde{\mu}I(e - \bar{\epsilon}^{\mathrm{p}})$
Check yield condition:
$f^{\mathrm{tr}} = ||s^{\mathrm{tr}}|| - \sqrt{\frac{2}{3}}K(\bar{\alpha})$
if $f^{\mathrm{tr}} \leqslant 0$ **then**
$\sigma = s^{\mathrm{tr}} + \tilde{\kappa}\mathrm{Tr}\,\epsilon\tilde{1}$
$\alpha = \bar{\alpha}$
$\epsilon^{\mathrm{p}} = \bar{\epsilon}^{\mathrm{p}}$
$C_{\mathrm{ep}} = \tilde{\kappa}\tilde{1}\tilde{1}^{\mathrm{T}} + 2\tilde{\mu}(I - \frac{1}{3}\tilde{1}\tilde{1}^{\mathrm{T}})$
else
Calculate flow parameter $\Delta\gamma$:
$\Delta\gamma = \frac{f^{\mathrm{tr}}}{2\tilde{\mu}+\frac{2}{3}k}$
Calculate $\hat{\boldsymbol{n}}$ and update the hardening parameter:
$\hat{\boldsymbol{n}} = \frac{s^{\mathrm{tr}}}{||s^{\mathrm{tr}}||}$
$\alpha = \bar{\alpha} + \sqrt{\frac{2}{3}}\Delta\gamma$
Update plastic strain and stress:
$\epsilon^{\mathrm{p}} = \bar{\epsilon}^{\mathrm{p}} + \Delta\gamma I^{-1}\hat{\boldsymbol{n}}$
$\sigma = s^{\mathrm{tr}} + \tilde{\kappa}\mathrm{Tr}\,\epsilon\tilde{1} - 2\tilde{\mu}\Delta\gamma\hat{\boldsymbol{n}}$
Calculate the consistent elasto-plastic tangent modulus:
$\beta_1 = 1 - \frac{2\tilde{\mu}\Delta\gamma}{||s^{\mathrm{tr}}||}$
$\beta_2 = \frac{2\tilde{\mu}}{2\tilde{\mu}+\frac{2}{3}k} - (1 - \beta_1)$
$C_{\mathrm{ep}} = \tilde{\kappa}\tilde{1}\tilde{1}^{\mathrm{T}} + 2\beta_1\tilde{\mu}(I - \frac{1}{3}\tilde{1}\tilde{1}^{\mathrm{T}}) - 2\tilde{\mu}\beta_2\hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^{\mathrm{T}}$
end if

Algorithm BNM (*modified BN method*)

$\tilde{\beta} \in (0,1)$, $\tilde{\sigma} \in (0,\frac{1}{2})$, $\tilde{\varepsilon} > 0$ small, $m_{\max}$, z^0, $\overline{U}_{\mathrm{c}}$, $\bar{\epsilon}_I^{\mathrm{p}}$ and $\bar{\alpha}_I$, $I = 1,\ldots,n_I$ given.
$k = 0$
Get stresses for each Gauss point I:
$\sigma_I(U^k) = \mathrm{RR}(U^k, \bar{\alpha}_I, \bar{\epsilon}_I^{\mathrm{p}})$
Calculate error:
$\Theta(z^k) = \frac{1}{2}H(z^k)^{\mathrm{T}}H(z^k)$
where H is obtained from (30).
while $\Theta(z^k) > \tilde{\varepsilon}$ **do**
Perform static condensation to the contact surface to obtain K_{c} and F_{c} in (35).
Solve the linear Newton equation (38) to obtain the direction $\mathrm{d}z_{\mathrm{c}}^k$:
$\widetilde{H}_{\mathrm{c}}'(z_{\mathrm{c}};\mathrm{d}z_{\mathrm{c}}) = -H_{\mathrm{c}}(z_{\mathrm{c}})$
where $\widetilde{H}_{\mathrm{c}}'(z_{\mathrm{c}};\mathrm{d}z_{\mathrm{c}})$ and $H_{\mathrm{c}}(z_{\mathrm{c}})$ are given in (37), (38), (35), (28) and (29).
Get $\mathrm{d}U_{\mathrm{o}}$ from (36).
$m = 0$

Get stresses for displacement $U^k + \mathrm{d}U^k$ for each Gauss point I:
$\sigma_I(U^k + \mathrm{d}U^k) = \mathrm{RR}(U^k + \mathrm{d}U^k, \bar{\alpha}_I, \bar{\epsilon}_I^{\mathrm{p}})$
Armijo line search:
while $\Theta(z^k + \tilde{\beta}^m \mathrm{d}z^k) > (1 - 2\tilde{\sigma}\tilde{\beta}^m)\Theta(z^k) \quad \& \quad m \leqslant m_{\max}$ **do**
$m = m + 1$
Get stresses for displacement $U^k + \tilde{\beta}^m \mathrm{d}U^k$ for each Gauss point I:
$\sigma_I(U^k + \tilde{\beta}^m \mathrm{d}U^k) = \mathrm{RR}(U^k + \tilde{\beta}^m \mathrm{d}U^k, \bar{\alpha}_I, \bar{\epsilon}_I^{\mathrm{p}})$
end while
$z^{k+1} = z^k + \tilde{\beta}^m \mathrm{d}z^k$
$k = k + 1$
end while
Update $\overline{U}_{\mathrm{c}}$, and $\bar{\epsilon}_I^{\mathrm{p}}$ and $\bar{\alpha}_I$ for each Gauss point I, for next increment:
$\overline{U}_{\mathrm{c}} = U_{\mathrm{c}} \qquad \bar{\epsilon}_I^{\mathrm{p}} = \epsilon_I^{\mathrm{p}} \qquad \bar{\alpha}_I = \alpha_I$

5. Numerical example

We have implemented algorithm BNM in our development code μ *Solve* (Christensen, 2000), which is written in modern Fortran 95. The geometry and loads are defined in our graphical Matlab pre-processor, which produces an input file which is read by the Fortran program. Matlab is then automatically called from within the Fortran program to produce plots as requested in the input file. The static condensation to the contact surface in (35) is done using our own skyline solver routines, whereas the non-symmetric system of linear equations in (38) is solved using LAPACK routines (Anderson et al., 1995). The time to perform the static condensation in each iteration dominates the total computing time, so that (38) is solved directly without further eliminating variables in order to reduce the size of the system actually solved, as was done to good effect for elastic contact problems in our previous paper Christensen and Pang (1998).

The parameters of algorithm BNM are set as follows: The line search parameters are $\tilde{\beta} = 0.6$ and $\tilde{\sigma} = 0.2$. In our treatments of elastic contact problems in Christensen et al. (1998) and Christensen and Pang (1998), we used $\tilde{\beta} = 0.9$ and $\tilde{\sigma} = 0.1$, but these parameters are not well suited for elasto-plastic problems as they give a large number of line searches, and, more seriously, often twice the number of iterations. In order to prevent jamming of the algorithm, a maximum number of line searches is set to $m_{\max} = 5$, so that the minimum step size becomes $\tilde{\beta}^{m_{\max}} = 0.0778$.

A guide line for choosing the parameters ρ_{n}^m, $m = 1, \ldots, n_{\mathrm{c}}$, in the reformulation of Signorini's law (28), is to view ρ_{n}^m as a scaling factor between $P_{\mathrm{n}}^m = I^m p_{\mathrm{n}}$ and $g^m - \overline{C}_{\mathrm{n}}^m U_{\mathrm{c}}$. For the example that follows below, the maximum normal contact pressure p_{n} is of the order 10^5, and the maximum gap $g^m - \overline{C}_{\mathrm{n}}^m U_{\mathrm{c}}$ is of the order 10^{-2}. Thus, $\rho_{\mathrm{n}}^m = 10^7 I^m$ represents a good choice. For the approximation of integrals on the contact surface in (24), we use the trapezoidal rule, so that

$$I^m = \sum_{j=1}^{f_m} \frac{l_{mj}}{2}, \quad m = 1, \ldots, n_{\mathrm{c}},$$

where f_m is the number of finite elements adjacent to contact node m, and l_{mj} is the length of the jth element adjacent to contact node m. For our example the lengths of the contact elements vary with the fineness of the mesh from 0.1 to 1 m. For simplicity we have used $\rho_{\mathrm{n}}^m = 10^6$, $m = 1, \ldots, n_{\mathrm{c}}$, and use the same value for the parameters r^m in the reformulation of Coulomb's law (29): $r^m = 10^6$, $m = 1, \ldots, n_{\mathrm{c}}$. For the example we have studied, ρ_{n}^m and r^m may be varied within a wide range without significantly changing the number of iterations (one should keep in mind, though, that the error norm used depends on these parameters, so that

a direct performance comparison is not possible unless an error norm independent on these parameters is used).

The starting point for the first increment of algorithm BNM is $u = 0$ except for the prescribed displacements which are put to their respective prescribed value, $P_{\mathrm{t}}^m = 0$ and $P_{\mathrm{n}}^m = 10^{-10}$, $m = 1, \ldots, n_{\mathrm{c}}$. The reason for not putting $P_{\mathrm{n}}^m = 0$, is that this represents a non-differentiable point, although we would like to stress that the performance of the algorithm would not be affected if $P_{\mathrm{n}}^m = 0$ were chosen. For subsequent increments, the starting point for increment j is set to the solution of increment $j - 1$. The termination tolerance is $\tilde{\varepsilon} = 10^{-10}$, which represents a very severe stopping criterion.

5.1. Plugged sheet

In this example a sheet of dimension 36×20 m^2 with a circular hole of radius 5 m in the middle, is subjected to a prescribed vertical displacement δ of the top and bottom boundaries of the sheet. Only a quarter of the sheet is studied due to symmetry, see Fig. 2. This structure is divided into two regions, which are both discretized by $n \times n$ elements, where $n = 4$, 8, 16 or 32. The sheet is discretized by four-noded bilinear finite elements with Young's modulus 70 MPa and Poisson's ratio 0.2, which implies that $\tilde{\mu} = 29.17$ MPa and $\tilde{\kappa} = 38.89$ MPa. The $\overline{B}$-matrices are calculated using the mean dilatation formulation (Hughes, 1987). Linear isotropic hardening with $\sigma_y = 0.243$ MPa and $k = 2.24$ MPa in (10) is considered. The load varies according to Fig. 2 during (a fictive) time 0 to 8 time units. The load is applied in $4 \times$ inc increments, where inc = 1, 2, 4, or 8. For the case $n = 32$, inc is also put to 16. The circular hole is initially completely filled up by a rigid cylinder of radius 5 m, i.e. the initial gap, $g^m = 0$, $m = 1, \ldots, n_{\mathrm{c}}$. The friction coefficient is $\mu = 0.0$, 0.3 or 0.9.

In Table 1, the average number of line searches, nls, per increment, as well as the average number of iterations, nle, per increment are presented. As seen, for the frictionless case with $n = 32$ and inc = 1, nle = 9.8 and nls = 8.0. In order to determine whether the contact equations or the plasticity equations are the most difficult to satisfy, we put $\sigma_y = 243$ GPa, so that elastic behavior is obtained, which gives nle = 3.3 and nls = 0. If instead, we put $g^m = 0.1$ m, $m = 1, \ldots, n_{\mathrm{c}}$, so that no contact occurs, then nle = 7.8 and nls = 7.8. As expected, $\sigma_y = 243$ GPa and $g^m = 0.1$ m, gives nle = 1 and nls = 0. Since we have always used the line search parameters $\tilde{\beta} = 0.9$ and $\tilde{\sigma} = 0.1$ previously, for elastic contact problems, we present an example illustrating their inefficiency for elasto-plastic problems. For $\mu = 0.9$, $n = 32$ and inc = 1, nle = 14.3 and nls = 15.0, cf. Table 1. If $\tilde{\beta} = 0.9$ and $\tilde{\sigma} = 0.1$ are used, then nle = 29.3 and nls = 237 (minimum step size in line search is put to 0.1). If we put $\sigma_y = 243$ GPa here (elastic behavior), then nle = 7.3 and nls = 6.0 with our new parameters, and nle = 7.0 and nls = 15.0 with $\tilde{\beta} = 0.9$ and $\tilde{\sigma} = 0.1$.

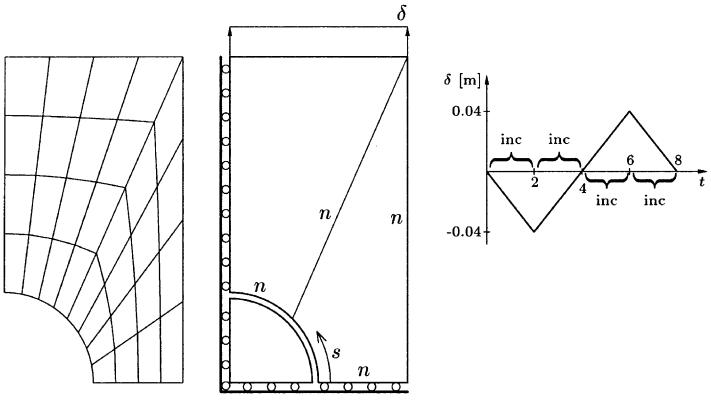

Fig. 2. Geometry for the plugged-sheet problem.

Table 1
Execution statistics

n	inc	$\mu = 0.0$		$\mu = 0.3$		$\mu = 0.9$	
		nle	nls	nle	nls	nle	nls
4	1	6.8	3.0	8.3	5.0	6.8	2.8
	2	4.0	0.8	5.0	4.3	5.4	7.0
	4	3.1	0.4	3.6	2.1	2.8	1.8
	8	2.8	0.0	3.3	1.3	2.4	0.9
8	1	6.8	4.3	8.3	5.3	8.5	8.5
	2	5.9	4.4	8.0	9.1	6.3	4.5
	4	4.1	0.9	5.1	4.7	4.9	4.6
	8	3.0	0.3	3.7	1.7	3.3	2.0
16	1	8.8	10.0	11.0	11.8	8.8	8.8
	2	6.6	4.9	7.8	6.6	7.9	7.1
	4	6.4	4.9	7.1	7.2	7.6	6.7
	8	4.6	1.9	5.5	4.0	5.5	4.7
32	1	11.3	14.8	12.0	12.5	14.3	15.0
	2	8.5	8.4	8.8	7.5	10.8	11.3
	4	6.8	4.9	7.6	5.1	8.6	9.0
	8	6.8	4.9	7.8	6.8	8.3	8.2
	16	4.8	1.8	5.6	2.9	6.0	4.5

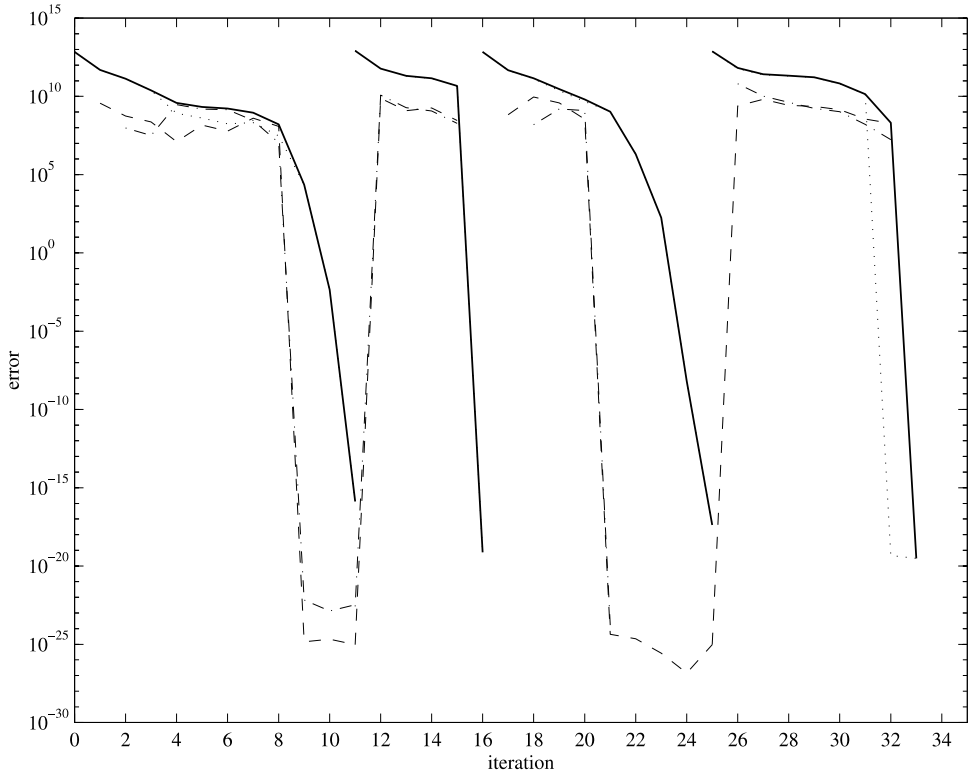

Fig. 3. Total error Θ (—), Θ^{eq} (···), Θ_n (– – –), and Θ_t (–·–·–) for $n = 8$, inc $= 1$ and $\mu = 0.3$.

Thus, the two parameter settings give equal performance for elastic behavior. The variation of the error during the iterations for $\mu = 0.3$, $n = 8$ and inc $= 1$, is shown in Fig. 3. We have found it illustrative to present not only the total error Θ, but also the error corresponding to the equilibrium equations Θ^{eq}, the error of the normal contact equations Θ_{n}, and the error of the friction equations Θ_{t}, where

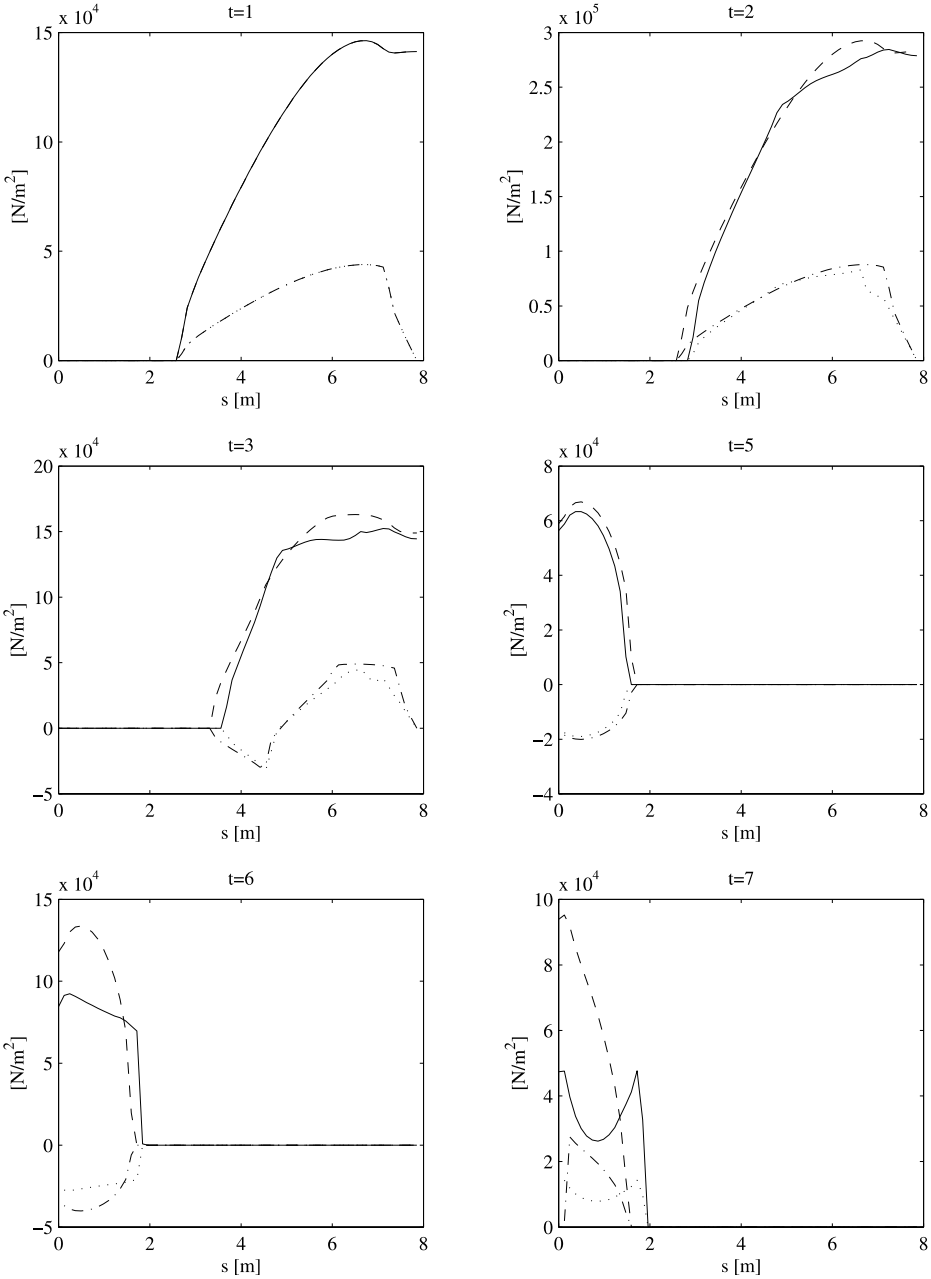

Fig. 4. Contact tractions for various instants in time for $n = 32$, inc $= 16$ and $\mu = 0.3$. Elasto-plastic case: (—): p_{n}; (· · ·): p_t. Elastic case: (—): p_{n}; (–·–·–): p_t. The coordinate s is defined in Fig. 2.

$$\Theta^{\mathrm{eq}} \equiv \frac{1}{2} H^{\mathrm{eq}^{\mathrm{T}}} H^{\mathrm{eq}}, \quad \Theta_{\mathrm{n}} \equiv \frac{1}{2} H_{\mathrm{n}}^{\mathrm{T}} H_{\mathrm{n}}, \quad \Theta_{\mathrm{t}} \equiv \frac{1}{2} H_{\mathrm{t}}^{\mathrm{T}} H_{\mathrm{t}}.$$

As seen in the figure, the total error decreases in each iteration for a given increment. In fact, this is theoretically guaranteed provided no iterates end up at non-differentiable points. The contact tractions for various instants in time are depicted in Fig. 4, for the case $n = 32$, inc $= 16$ and $\mu = 0.3$. In order to visualize the influence of elasto-plastic material behavior, the contact tractions obtained for the elastic case with the same μ are also presented. As the sheet is initially compressed, the time of first yield is $t = 1.625$, which explains why there is no difference between the elastic and elasto-plastic curves for $t = 1$. At $t = 2$ the loading is reversed and the plastic flow stops. At $t = 4$ the prescribed displacement is zero, and the gap is positive at all contact nodes, and, thus, the contact tractions are zero. As the sheet is drawn out, plastic flow starts again at $t = 5.25$. At time $t = 6$ the load is reversed again, and the flow stops. At time $t = 7.75$, the sheet loses contact with the cylinder. Note that the contact pressure for the elasto-plastic case looks very different from the contact pressure for the elastic case for $t = 7$. In Fig. 5, the normal contact pressure for $t \geqslant 6$ is shown. The same entity is plotted for the frictionless case. In Endahl (1985), a similar problem with an infinite elasto-plastic sheet plugged with a smooth elasto-plastic cylinder made of the same material as the sheet, is analyzed. Our plots for the frictionless case show much resemblance to the plots in Endahl (1985, p. 43). In Figs. 6 and 7, the von Mises stress at full compression ($t = 2$) and full tension ($t = 6$) is shown both for the elastic as well as the elasto-plastic case.

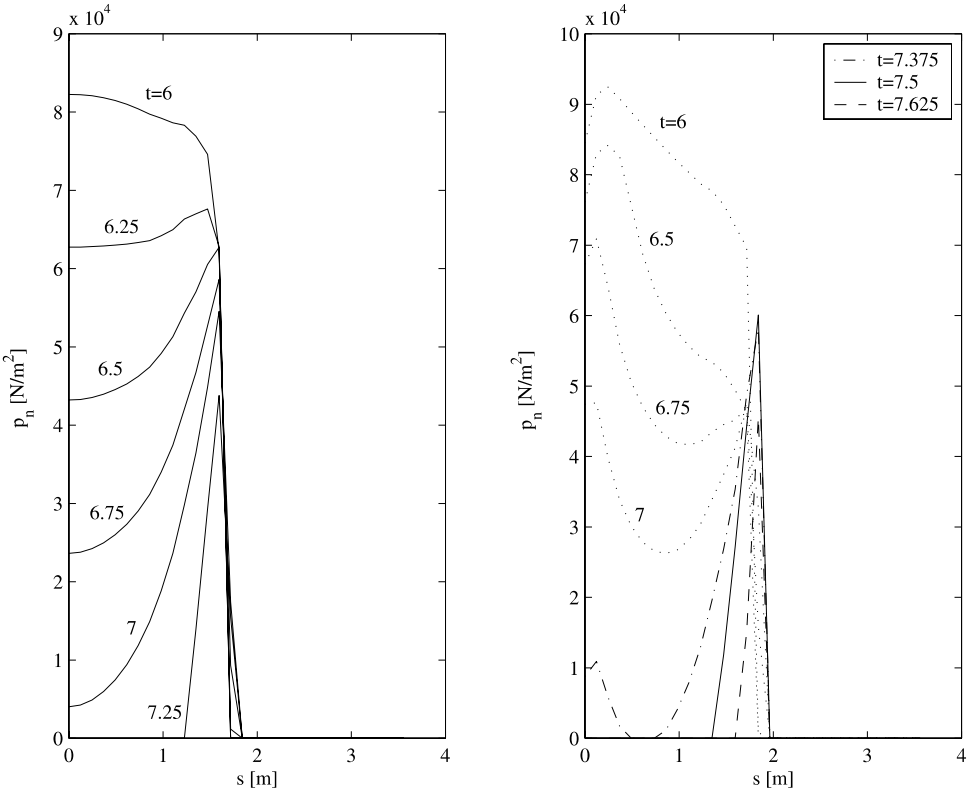

Fig. 5. Normal contact pressure for various instants in time for $n = 32$ and inc $= 16$. Left figure: $\mu = 0$; right figure: $\mu = 0.3$. The coordinate s is defined in Fig. 2.

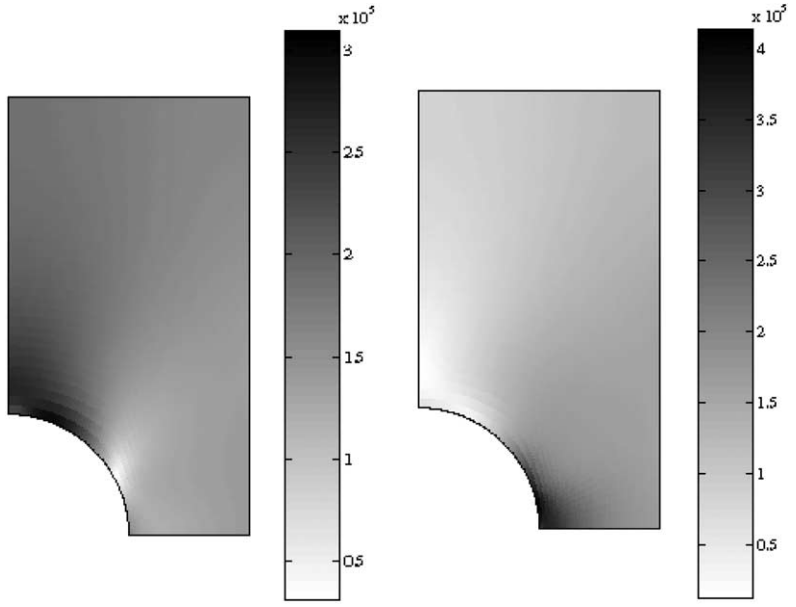

Fig. 6. The von Mises stress in Pa at $t = 2$ (left) and $t = 6$ (right) for $n = 32$, inc $= 16$, $\mu = 0.3$, and elastic behavior.

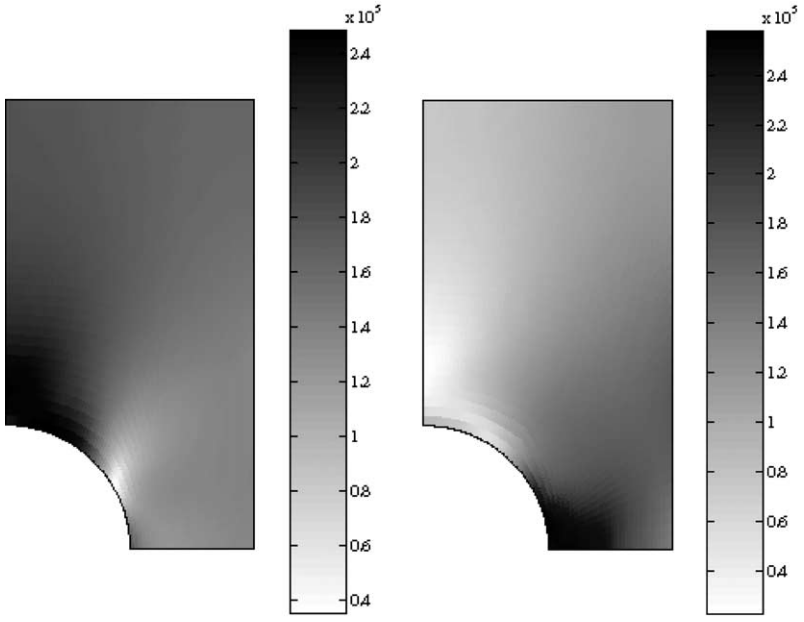

Fig. 7. The von Mises stress in Pa at $t = 2$ (left) and $t = 6$ (right) for $n = 32$, inc $= 16$, $\mu = 0.3$, and elasto-plastic behavior.

6. Conclusions

In this paper we have reformulated the two-dimensional, discrete, time-incremental, elasto-plastic frictional contact problem as a set of unconstrained semi-smooth equations. A damped Newton method with proven global convergence properties for solving these equations is presented. In our implementation, a slight modification of this method is used, where simplifications regarding iterates at non-differentiable points are performed. This modified Newton method shows excellent performance and robustness. Three-dimensional problems can be handled by the same method; see Christensen et al. (1998) and Christensen and Pang (1998) for a semi-smooth reformulation of the constitutive laws of frictional contact in the three-dimensional case.

Finally, we point out that our semi-smooth Newton method with its line search (which is different from the ones usually used in finite element programs) obviously can be used for elasto-plastic problems not involving contact as well.

Acknowledgements

This work was based on research supported by the Center for Industrial Information Technology (CENIIT) and ABB Atom AB.

References

Anderson, E. et al., 1995. LAPACK Users' Guide, second ed. SIAM Publications, Philadelphia.

Chaney, R.W., 1990. Piecewise C^k functions in nonsmooth analysis. Nonlinear Analysis, Theory, Methods and Applications 15, 649–660.

Christensen, P.W., 2000. μ Solve—A Finite Element Program for Nonsmooth Mechanics, in preparation.

Christensen, P.W., 2002. A nonsmooth Newton method for elastoplastic problems. Computer Methods in Applied Mechanics and Engineering, 191, 1189–1219.

Christensen, P.W., Pang, J.S., 1998. Frictional contact algorithms based on semismooth Newton methods. In: Fukushima, M., Qi, L. (Eds.), Reformulation—Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods. Kluwer Academic Publishers, Dordrecht, pp. 81–116.

Christensen, P.W., Klarbring, A., Pang, J.S., Strömberg, N., 1998. Formulation and comparison of algorithms for frictional contact problems. International Journal for Numerical Methods in Engineering 42, 145–173.

Endahl, N., 1985. Elastoplastic Indentation Problems, A Finite Element Study with Applications to Rolling Bearing Technology. Linköping Studies in Science and Technology. Dissertations. No. 128, Linköping University, Linköping.

Felippa, C.A., 1975. Solution of linear equations with skyline-stored symmetric matrix. Computers and Structures 5, 13–29.

Hiriart-Urruty, J.-B., Lemaréchal, C., 1993. Convex Analysis and Minimization Algorithms I. Springer, Berlin.

Hughes, T.J.R., 1987. The Finite Element Method, Linear Static and Dynamic Finite Element Analysis. Prentice–Hall, Englewood Cliffs.

Kikuchi, N., Oden, J.T., 1988. Contact Problems in Elasticity: A Study of Variational Inequalities and Finite Element Methods. SIAM, Philadelphia.

Klarbring, A., 1992. Mathematical programming and augmented Lagrangian methods for frictional contact problem. In: Curnier, A (Ed.), Proceedings Contact Mechanics International Symposium. PPUR, pp. 409–422.

Klarbring, A., 1993. Mathematical programming in contact problems. In: Aliabdali, M.H., Brebbia, C.A. (Eds.), Computational Methods in Contact Mechanics. Computational Mechanics Publications, Southampton, pp. 233–263.

Ortega, J.M., Rheinboldt, W.C., 1970. Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York.

Pang, J.S., 1990. Newton's method for B-differentiable equations. Mathematics of Operations Research 15, 311–341.

Pang, J.S., Qi, L., 1995. A globally convergent Newton method for convex SC^1 minimization problems. Journal of Optimization Theory and Applications 85, 633–648.

Qi, L., 1993. Convergence analysis of some algorithms for solving nonsmooth equations. Mathematics of Operations Research 18, 227–244.

Simo, J.C., Hughes, T.J.R., 1998. Computational Inelasticity. Springer, New York.

Strömberg, N., 1997. An augmented Lagrangian method for fretting problems. European Journal of Mechanics, A/Solids 16, 573–593.