Computers & Structures Vol. 40, No. 2, pp. 203–209, 1991
Printed in Great Britain.

0045-7949/91 $3.00 + 0.00
Pergamon Press plc

ON THE TREATMENT OF INEQUALITY CONSTRAINTS ARISING FROM CONTACT CONDITIONS IN FINITE ELEMENT ANALYSIS

A. L. ETEROVIC and K. J. BATHE
Massachusetts Institute of Technology, Department of Mechanical Engineering, Room 3-356, Cambridge, MA 02139, U.S.A.

Abstract—Existing methods for the analysis of contact problems deal with the inequality constraints arising from contact conditions by means of an implicit iteration on all constraints. This paper presents a formulation for contact problems with friction for large deformations where all inequality constraints are enforced explicitly. A robust solution technique for the resulting system of nonlinear equations can then be used. This approach admits the use of line search procedures to enlarge the region of convergence.

1. INTRODUCTION

Much attention has been devoted over the past years to the development of solution methods for the analysis of contact problems. Since the earlier studies [1–4], a number of different approaches have been proposed and researched.

Contact problems range from frictionless contact in small-strain elastic analysis, to contact with friction in general large-strain inelastic analysis. Although conceptually related, these cases differ significantly in the way they may be formulated and solved.

Problems that admit an energy functional, such as general hyperelastic behavior with contact conditions that do not include frictional effects, can be formulated as unilaterally constrained optimization problems. In this case a number of optimization methods with global convergence properties are available [5–7], and algorithms based on mathematical programming have been proposed [8].

The situation is much more complex when the problem includes inelastic material behavior or nonconservative frictional conditions are present. In this case global convergence is much harder to achieve.

Among the most commonly used methods for contact analysis are the Lagrange multiplier method [9–12], the penalty method [13–18], the perturbed Lagrangian method [19–21] and the augmented Lagrangian method [22].

The penalty method has the advantage that the contact constraints are taken into account with no increase in the number of degrees of freedom. Indeed, using the displacement-based finite element formulation the number of unknowns equals the number of displacement degrees of freedom, independent of the number of constraints.

In the classical Lagrange multiplier method contact tractions are considered as additional degrees of freedom. The perturbed Lagrangian method can be considered as a generalization of the Lagrange multiplier method where an additional term involving the contact tractions is added to the variational equations. The classical Lagrange multiplier method is then obtained as a limiting case, while the penalty method is recovered by solving for the contact tractions and eliminating them from the equilibrium equations.

It should be noted that in both the penalty method and the perturbed Lagrangian method an approximate solution to the original problem is obtained, which depends on an arbitrary penalty parameter, and in practice this parameter can drastically affect the results.

The augmented Lagrangian method is a widely used method in optimization where a function of displacements and Lagrange multipliers is constructed so that its minimum corresponds to the solution of the constrained problem.

In order to apply the above procedures to contact inequality constraints, methods are used in which in each iteration only those constraints that are active contribute to the incremental equations. There are two undesirable consequences of this approach. First, the quadratic convergence of Newton's method may be seriously affected if there are frequent changes of the active set, and second, few global convergence results are available.

It has been observed that the local convergence properties of the Newton iteration are hard to achieve, and emphasis has been given to the use of consistent tangent stiffness matrices for nonlinear contact analysis [20, 23, 24]. However, for fully quadratic convergence the tangent stiffness matrix has to satisfy regularity conditions that are being violated by the most commonly used definitions of the gap or interpenetration functions.

The objective of this paper is to present a method for finite element analysis of contact problems in which all inequality constraints are enforced directly by means of additional equations,. These equations, together with the equilibrium equations, form a system of nonlinear equations in the displacements and contact tractions. The Newton method can then be applied, and the region of convergence can be improved by line-searching techniques. The method is particularly suited for nonconservative problems such as inelastic large-strain analysis with frictional conditions.

This paper is expository in nature, numerical tests of the procedure will be presented in a forthcoming communication.

2. CONTINUUM FORMULATION

We consider a system of N bodies in Euclidean space subjected to a history of loads such that arbitrary displacements and strains are developed, and varying contact and frictional conditions are established. The bodies are allowed to be either deformable or rigid.

2.1. *Equilibrium equations*

Let T be the relevant time interval. Let ${}^0V$ be the position of the system at time $t_0 \in T$. Then ${}^0V = \bigcup_{I=1}^{N} {}^0V^I$, where ${}^0V^I$ denotes the position of body I at time t_0. Let $\mathbf{x}$ denote the motion of the system, i.e. $\mathbf{x}$ is defined on ${}^0V \times T$ and $\mathbf{x} = \mathbf{x}(\mathbf{x}_0, t)$ gives the position at time t of the material point that at time t_0 was at $\mathbf{x}_0$. Let $V = \mathbf{x}({}^0V, t)$ be the position of the system at time t and let $V^I = \mathbf{x}({}^0V^I, t)$ be the position of body I at time t. We denote by ${}^0\partial V$ and $\partial V = \mathbf{x}({}^0\partial V, t)$ the boundaries of ${}^0V$ and V, respectively.

Our purpose is to derive a variational formulation for the motion of the system that takes into account the contact conditions. We start with the field equations corresponding to the balance law of linear momentum. For all $t \in T$ we have

$$\operatorname{div} \mathbf{T} + \rho(\mathbf{b} - \ddot{\mathbf{u}}) = \mathbf{0}, \quad \text{on } V \tag{1}$$

$$\mathbf{Tn} = \mathbf{f}, \quad \text{on } \partial V \tag{2}$$

where $\mathbf{T}$ is the Cauchy stress tensor, ρ is the density per unit current volume, $\mathbf{b}$ is the vector of body forces per unit mass, $\ddot{\mathbf{u}}$ is the acceleration vector, $\mathbf{n}$ is the outward unit normal to the boundary of V and $\mathbf{f}$ is the vector of surface tractions. The divergence operator div involves derivatives with respect to current coordinates and when acting on second order tensors it is defined by the property $\mathbf{c} \cdot \operatorname{div} \mathbf{T} = \operatorname{div}(\mathbf{T}^T\mathbf{c})$ for all constant vectors $\mathbf{c}$.

Let $\bar{\mathbf{u}}$ be any (virtual) displacement field defined on ${}^0V \times T$. Multiplying (1) by $\bar{\mathbf{u}}$ and integrating over V we obtain

$$\sum_{I=1}^{N} \int_{V^I} [\operatorname{div} \mathbf{T} + \rho(\mathbf{b} - \ddot{\mathbf{u}})] \cdot \bar{\mathbf{u}} \, \mathrm{d}V = 0. \tag{3}$$

Recall the following special form of the divergence theorem

$$\int_V \bar{\mathbf{u}} \cdot \operatorname{div} \mathbf{T} \, \mathrm{d}V = \int_{\partial V} \mathbf{n} \cdot \mathbf{T}^T\bar{\mathbf{u}} \, \mathrm{d}S - \int_V \mathbf{T} \cdot \operatorname{grad} \bar{\mathbf{u}} \, \mathrm{d}V, \tag{4}$$

where for any two second order tensors $\mathbf{A}$, $\mathbf{B}$ with components A_{ij}, B_{ij} we write $\mathbf{A} \cdot \mathbf{B} = \Sigma_{i,j} A_{ij} B_{ij}$. The gradient operator grad involves derivatives with respect to current coordinates and when acting on vectors it is defined by the property $(\operatorname{grad} \bar{\mathbf{u}})^T\mathbf{c} = \operatorname{grad}(\bar{\mathbf{u}} \cdot \mathbf{c})$ for all constant vectors $\mathbf{c}$.

Using (2) and (4), eqn (3) can be written as

$$\sum_{I=1}^{N} \left\{ \int_{V^I} \mathbf{T} \cdot \operatorname{grad} \bar{\mathbf{u}} \, \mathrm{d}V - \int_{V^I} \rho(\mathbf{b} - \ddot{\mathbf{u}}) \cdot \bar{\mathbf{u}} \, \mathrm{d}V - \int_{\partial V^I} \mathbf{f}^I \cdot \bar{\mathbf{u}} \, \mathrm{d}S \right\} = 0, \tag{5}$$

where $\mathbf{f}^I = \mathbf{Tn}^I$ and $\mathbf{n}^I$ is the outward unit normal to ∂V^I.

We denote by ∂V^{Id} that part of the current surface ∂V^I where displacements are prescribed, and by ∂V^{If} that part of ∂V^I where surface tractions are prescribed. If $\tilde{\mathbf{u}}^I$ and $\tilde{\mathbf{f}}^I$ are respectively the prescribed displacement and surface traction vectors, we require that for all $t \in T$

$$\mathbf{u}^I = \tilde{\mathbf{u}}^I \quad \text{on } \partial V^{Id} \tag{6}$$

$$\mathbf{f}^I = \tilde{\mathbf{f}}^I \quad \text{on } \partial V^{If}. \tag{7}$$

Finally, denote by ∂V^{Ic} that part of ∂V^I where contact can occur during the motion. Note that for each time t, ∂V^{Id}, ∂V^{If}, and ∂V^{Ic} are disjoint, and their union is all of ∂V^I.

We require $\bar{\mathbf{u}}$ to vanish on ∂V^{Id}. Using this fact and eqn (7) we can rewrite (5) as

$$\sum_{I=1}^{N} \left\{ \int_{V^I} \mathbf{T} \cdot \operatorname{grad} \bar{\mathbf{u}} \, \mathrm{d}V - \int_{V^I} \rho(\mathbf{b} - \ddot{\mathbf{u}}) \cdot \bar{\mathbf{u}} \, \mathrm{d}V - \int_{\partial V^{If}} \tilde{\mathbf{f}}^I \cdot \bar{\mathbf{u}} \, \mathrm{d}S \right\} - \sum_{I=1}^{N} \int_{\partial V^{Ic}} \mathbf{f}^I \cdot \bar{\mathbf{u}} \, \mathrm{d}S = 0. \tag{8}$$

Let $\mathbf{X} = \operatorname{Grad} \mathbf{x}$ be the deformation gradient, where Grad is the gradient operator with respect to the coordinates at time t_0. Then

$$\operatorname{Grad} \bar{\mathbf{u}} = (\operatorname{grad} \bar{\mathbf{u}})\mathbf{X}. \tag{9}$$

If $\mathrm{d}S$ is a surface element at time t with unit normal $\mathbf{n}$ and ${}^0\mathrm{d}S$ is the corresponding surface element at time t_0 with unit normal ${}^0\mathbf{n}$ then

$$\mathbf{n} \, \mathrm{d}S = (\det \mathbf{X})\mathbf{X}^{-T}\, {}^0\mathbf{n}\, {}^0\mathrm{d}S.$$

Multiplying both sides by $\mathbf{T}$ we have

$$\mathbf{f}\,\mathrm{d}S = \mathbf{X}\mathbf{S}\,{}^0\mathbf{n}\,{}^0\mathrm{d}S = {}^0\mathbf{f}\,{}^0\mathrm{d}S, \qquad (10)$$

where $\mathbf{S} = (\det \mathbf{X})\mathbf{X}^{-1}\mathbf{T}\mathbf{X}^{-T}$ is the second Piola–Kirchhoff stress tensor and ${}^0\mathbf{f} = \mathbf{X}\mathbf{S}\,{}^0\mathbf{n}$ is the surface traction per unit reference area. Combining (9) and (10) we write the referencial form of (8) as

$$\sum_{I=1}^{N}\left\{\int_{{}^0V^I} \mathbf{S}\cdot\mathbf{X}^T\,\mathrm{Grad}\,\bar{\mathbf{u}}\,{}^0\mathrm{d}V - \int_{{}^0V^I} {}^0\rho(\mathbf{b}-\ddot{\mathbf{u}})\cdot\bar{\mathbf{u}}\,{}^0\mathrm{d}V - \int_{{}^0\partial V^I_f} {}^0\bar{\mathbf{f}}^I\cdot\bar{\mathbf{u}}\,{}^0\mathrm{d}S\right\} - \sum_{I=1}^{N}\int_{\partial V^{Ic}} \mathbf{f}^I\cdot\bar{\mathbf{u}}\,\mathrm{d}S = 0, \qquad (11)$$

where ${}^0\rho$ is the density per unit reference volume. This is of course the total Lagrangian formulation of the principle of virtual work [25]. The last term in (11) is written in terms of the current surface area for convenience. In the following section we analyze this term in more detail.

2.2. *Contact conditions*

Let ∂V^{IJ} be that part of ∂V^{Ic} where contact with body J can occur during the motion. Then $\partial V^{Ic} = \cup_{J=1}^{N}\partial V^{IJ}$. Let $\mathbf{f}^{IJ}$ be the vector of surface tractions on body I due to contact with body J. By Newton's law of action and reaction, $\mathbf{f}^{JI} = -\mathbf{f}^{IJ}$. Hence that part of the last term of (11) which corresponds to the surfaces ∂V^{IJ} and ∂V^{JI} can be written as

$$-\int_{\partial V^{IJ}} \mathbf{f}^{IJ}\cdot\bar{\mathbf{u}}^I\,\mathrm{d}S - \int_{\partial V^{JI}} \mathbf{f}^{JI}\cdot\bar{\mathbf{u}}^J\,\mathrm{d}S = \int_{\partial V^{IJ}} \mathbf{f}^{IJ}\cdot\bar{\mathbf{u}}^{IJ}\,\mathrm{d}S, \qquad (12)$$

where $\bar{\mathbf{u}}^I = \bar{\mathbf{u}}|_{V^I}$, the restriction of $\bar{\mathbf{u}}$ to V^I, $\bar{\mathbf{u}}^J = \bar{\mathbf{u}}|_{V^J}$, and $\bar{\mathbf{u}}^{IJ} = \bar{\mathbf{u}}^J - \bar{\mathbf{u}}^I$.

We call each pair of surfaces ∂V^{IJ} and ∂V^{JI} with $I \neq J$ a 'contact pair'. It is convenient to call ∂V^{IJ} the 'contactor surface' and ∂V^{JI} the 'target surface'. Thus the right-hand side of (12) can be interpreted as the virtual work that the contact tractions produce over the virtual relative displacements on contact pair IJ.

In what follows we analyze the right-hand side of (12). For the sake of clarity, we do not carry the superindices I and J over the new variables to be defined for the contact pair. Also, we specialize the derivations to three-dimensional analysis; equations for the two-dimensional case can then be inferred.

Let $\mathbf{n}$ be the unit outward normal to ∂V^{JI} and let $\mathbf{s}_1$ and $\mathbf{s}_2$ be vectors such that $\{\mathbf{n}, \mathbf{s}_1, \mathbf{s}_2\}$ form an orthonormal basis. We decompose the unknown contact traction $\mathbf{f}^{IJ}$ acting on ∂V^{IJ} into normal and tangential components relative to ∂V^{JI} according to

$$\mathbf{f}^{IJ} = \lambda\mathbf{n} + \mathbf{t} \qquad (13)$$

$$\mathbf{t} = t_1\mathbf{s}_1 + t_2\mathbf{s}_2, \qquad (14)$$

where

$$\lambda = \mathbf{f}^{IJ}\cdot\mathbf{n} \qquad (15)$$

$$t_1 = \mathbf{f}^{IJ}\cdot\mathbf{s}_1 \qquad (16)$$

$$t_2 = \mathbf{f}^{IJ}\cdot\mathbf{s}_2. \qquad (17)$$

With these definitions the right-hand side of (12) reads

$$\int_{\partial V^{IJ}} (\lambda\mathbf{n} + t_1\mathbf{s}_1 + t_2\mathbf{s}_2)\cdot\bar{\mathbf{u}}^{IJ}\,\mathrm{d}S. \qquad (18)$$

Next we analyze the contact conditions. First, no interpenetration should occur throughout the motion. This condition can be formally stated as

$$V^I \cap V^J = \emptyset, \quad \forall t \in T, \qquad (19)$$

where $\emptyset$ denotes the empty set. Second, the normal contact tractions can only be compressive.

Consider a point $\mathbf{x}$ on ∂V^{IJ} and let $\mathbf{y}^*(\mathbf{x}, t)$ be a point on ∂V^{JI} satisfying

$$\|\mathbf{x} - \mathbf{y}^*(\mathbf{x}, t)\| = \min\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in \partial V^{JI}\}. \qquad (20)$$

The (signed) distance from $\mathbf{x}$ to ∂V^{JI} is then given by

$$g(\mathbf{x}, t) = [\mathbf{x} - \mathbf{y}^*(\mathbf{x}, t)]\cdot\mathbf{n}(\mathbf{y}^*(\mathbf{x}, t)). \qquad (21)$$

We call g the 'gap function' for the contact pair IJ. Note that when the solution $\mathbf{y}^*(\mathbf{x}, t)$ to the minimization in (20) is unique, the gradient of the gap function is given by

$$\nabla g(\mathbf{x}, t) = \mathbf{n}(\mathbf{y}^*(\mathbf{x}, t)), \qquad (22)$$

i.e. the gradient of the gap function is precisely equal to the normal to the target surface at the solution point $\mathbf{y}^*(\mathbf{x}, t)$.

Using definition (21), the conditions for normal contact can be restated as

$$g \geqslant 0, \qquad \lambda \geqslant 0, \qquad g\lambda = 0. \qquad (23)$$

We shall assume that Coulomb's law of friction holds pointwise on the contact surface (although more representative friction laws are clearly desirable [26–29]), and define the functions h and $\dot{\mathbf{u}}^{IJ}$ on ∂V^{IJ} by

$$h(\mathbf{x}, t) = \mu\lambda(\mathbf{x}, t) - \|\mathbf{t}(\mathbf{x}, t)\| \qquad (24)$$

$$\dot{\mathbf{u}}^{IJ}(\mathbf{x}, t) = \dot{\mathbf{u}}^J(\mathbf{y}^*(\mathbf{x}, t), t) - \dot{\mathbf{u}}^I(\mathbf{x}, t), \qquad (25)$$

where μ is the coefficient of friction between surfaces ∂V^{IJ} and ∂V^{JI}, $\dot{\mathbf{u}}^I(\mathbf{x}, t)$ is the velocity of point $\mathbf{x}$ at time t and $\dot{\mathbf{u}}^J(\mathbf{y}^*(\mathbf{x}, t), t)$ is the velocity of the point with position $\mathbf{y}^*(\mathbf{x}, t)$ at time t. $\dot{\mathbf{u}}^{IJ}(\mathbf{x}, t)$ is then the velocity at time t of the point with position $\mathbf{y}^*$ relative to point $\mathbf{x}$.

In view of these definitions Coulomb's law of friction states that

$$h \geqslant 0 \text{ and} \quad (26)$$

$$h > 0 \text{ implies } \dot{\mathbf{u}}^{IJ} = \mathbf{0} \text{ while} \quad (27)$$

$$h = 0 \text{ implies } \dot{\mathbf{u}}^{IJ} = \gamma \mathbf{t} \text{ for some } \gamma \geqslant 0. \quad (28)$$

Equations (26)–(28) are equivalent to

$$h \geqslant 0, \quad \gamma \geqslant 0, \quad h\gamma = 0 \quad (29)$$

$$\lambda(\dot{\mathbf{u}}^{IJ} - \gamma \mathbf{t}) = \mathbf{0}. \quad (30)$$

Note the similarity between eqns (23) for normal contact and (29) for frictional conditions.

We summarize below the contact conditions for further reference

$$g \geqslant 0, \quad \lambda \geqslant 0, \quad g\lambda = 0 \quad (31)$$

$$h \geqslant 0, \quad \gamma \geqslant 0, \quad h\gamma = 0 \quad (32)$$

$$\lambda(\dot{\mathbf{u}}^{IJ} - \gamma \mathbf{t}) = \mathbf{0} \quad (33)$$

$$\forall x \in \partial V^{IJ} \quad \text{and} \quad \forall t \in T.$$

Equations (31)–(33) can be interpreted by considering the following cases:

1. *No contact*: If $g > 0$, the equality in (31) implies $\lambda = 0$. It follows from definition (24) and the first inequality in (32) that $\|\mathbf{t}\| = 0$. When there is no contact, all contact tractions must be zero. Equation (33) is trivially satisfied.
2. *Sticking contact*: If $\lambda > 0$ and $h > 0$, the equalities in (31) and (32) imply $g = 0$ and $\gamma = 0$. It follows from eqn (33) that $\dot{\mathbf{u}}^{IJ} = \mathbf{0}$. When there is contact and the contact friction force $\mathbf{t}$ has norm less than the frictional resistance $\mu\lambda$ there is no relative motion.
3. *Sliding contact*: If $\lambda > 0$ and $h = 0$, the equality in (31) implies $g = 0$ and definition (24) implies $\|\mathbf{t}\| = \mu\lambda$. It follows from eqn (33) that $\dot{\mathbf{u}}^{IJ} = \gamma\mathbf{t}$. When there is contact and the contact friction force $\mathbf{t}$ has norm equal to the frictional resistance $\mu\lambda$, the motion of body J relative to body I must be in the direction of $\mathbf{t}$, the friction force on body I due to body J.

To put conditions (31)–(33) into a variational framework, we proceed as follows. Let w be a real valued function of two variables such that the solutions of equation $w(x, y) = 0$ satisfy $x \geqslant 0$, $y \geqslant 0$, and $xy = 0$. Functions with these properties have, for example, been used in [30, 31] in the context of complementarity problems and inequality constrained optimization. Using such a function, eqns (31)–(32) can now be written as

$$w(g, \lambda) = 0 \quad (34)$$

$$w(h, \gamma) = 0. \quad (35)$$

Let $\bar{\lambda}$, $\bar{\gamma}$, $\bar{t}_1$ and $\bar{t}_2$ be variations of λ, γ, t_1 and t_2, respectively. Multiplying (34) by $\bar{\lambda}$, (35) by $\bar{\gamma}$, and (33) by $\bar{t}_1\mathbf{s}_1 + \bar{t}_2\mathbf{s}_2$, adding, and integrating over ∂V^{IJ} we obtain

$$\int_{\partial V^{IJ}} [w(g, \lambda)\bar{\lambda} + w(h, \gamma)\bar{\gamma} + \lambda(\dot{\mathbf{u}}^{IJ} \cdot \mathbf{s}_1 - \gamma t_1)\bar{t}_1 + \lambda(\dot{\mathbf{u}}^{IJ} \cdot \mathbf{s}_2 - \gamma t_2)\bar{t}_2]\, \mathrm{d}S = 0. \quad (36)$$

Equations (11), (18), and (36), together with appropriate constitutive equations give a complete variational formulation for the motion of the system.

3. FINITE ELEMENT FORMULATION

To attempt a solution of the contact problem formulation (18) and (36) we partition the time interval T into a sequence of time steps. In what follows we restrict our attention to the associated quasistatic problem. Inertia forces corresponding to $\ddot{\mathbf{u}}$ are neglected and the relative interface velocity $\dot{\mathbf{u}}^{IJ}$ at time $t + \Delta t$ is approximated as $\Delta\mathbf{u}^{IJ}/\Delta t$, where $\Delta\mathbf{u}^{IJ}$ is the change in the relative interface displacement from time t to time $t + \Delta t$. With these simplifications, given the solution at time t, we seek displacements and contact tractions at time $t + \Delta t$ that satisfy

$$\sum_{I=1}^{N} \left\{ \int_{{}^0V^I} \mathbf{S} \cdot \mathbf{X}^T \operatorname{Grad} \bar{\mathbf{u}}\, {}^0\mathrm{d}V - \int_{{}^0V^I} {}^0\rho\mathbf{b} \cdot \bar{\mathbf{u}}\, {}^0\mathrm{d}V - \int_{{}^0\partial V^I_f} {}^0\mathbf{f}^I \cdot \bar{\mathbf{u}}\, {}^0\mathrm{d}S \right\} - \sum_{I=1}^{N} \int_{\partial V^I_c} \mathbf{f}^I \cdot \bar{\mathbf{u}}\, \mathrm{d}S = 0 \quad (37)$$

$$\int_{\partial V^{IJ}} [w(g, \lambda)\bar{\lambda} + w(h, \gamma)\bar{\gamma} + \lambda(\mathbf{s}_1 \cdot \Delta\mathbf{u}^{IJ}/\Delta t - \gamma t_1)\bar{t}_1 + \lambda(\mathbf{s}_2 \cdot \Delta\mathbf{u}^{IJ}/\Delta t - \gamma t_2)\bar{t}_2]\, \mathrm{d}S = 0. \quad (38)$$

Given a finite element discretization of ${}^0V$, we approximate the integrals over ∂V^{IJ} in (37) and (38) using the values of the integrands on the surface nodes. Although other integration formulas can be used, this choice has the attractive property that the contact conditions are enforced exactly at each contactor node.

Let $k = 1, \ldots, m$ be an index for all the contactor nodes in the system. Recall definition (21) of the gap function. The minimization over the target surface involved in computing the gap function is now performed over the 'discretized' target surface. This surface is divided into segments defined by the surface nodes. These 'target segments' may or may not correspond to the boundaries of the finite elements on the target body. For each contactor node k, where $k = 1, \ldots, m$, let kt denote the point on the target surface at which the minimum is attained and denote by $\{\mathbf{n}_k, \mathbf{s}_{1k}, \mathbf{s}_{2k}\}$ the normal and tangent vectors to the target surface at this point.

Let $k1, \ldots, kp$ be the nodes that define the target segment on which point kt lies. Denote by

$$\hat{\mathbf{x}}_k^T = (\mathbf{x}_k^T, \mathbf{x}_{k1}^T, \ldots, \mathbf{x}_{kp}^T) \quad \text{and} \quad \hat{\mathbf{u}}_k^T = (\mathbf{u}_k^T, \mathbf{u}_{k1}^T, \ldots, \mathbf{u}_{kp}^T)$$

the vectors containing the nodal point positions and displacements of nodes k, $k1, \ldots, kp$, respectively. We denote by $g_k(\hat{\mathbf{u}}_k)$ and $\mathbf{n}_k(\hat{\mathbf{u}}_k)$ the corresponding (discretized) gap function and normal vector.

The positions and displacements of points on the target segment k are obtained by isoparametric interpolation from the positions and displacements of the nodes $k1, \ldots, kp$, respectively. Therefore the position of point kt is given by

$$\mathbf{x}_{kt} = \sum_{r=1}^{p} \beta_{kr} \mathbf{x}_{kr} \tag{39}$$

for some coefficients β_{kr} with

$$\sum_{r=1}^{p} \beta_{kr} = 1.$$

The relative virtual displacement vector and the increment in the relative displacement vector between the target point and the contactor node are then given by

$$\bar{\mathbf{u}}_k^{IJ} = \sum_{r=1}^{p} \beta_{kr} \bar{\mathbf{u}}_{kr} - \bar{\mathbf{u}}_k \tag{40}$$

$$\Delta\mathbf{u}_k^{IJ} = \sum_{r=1}^{p} \beta_{kr} \Delta\mathbf{u}_{kr} - \Delta\mathbf{u}_k, \tag{41}$$

where $\bar{\mathbf{u}}_k$ and $\Delta\mathbf{u}_k$ are the virtual displacement and the displacement increment vectors for the contactor node k, respectively, and $\bar{\mathbf{u}}_{kr}$ and $\Delta\mathbf{u}_{kr}$ are the virtual displacement and the displacement increment vectors for node kr, respectively.

This convention is illustrated in Fig. 1 for a two-dimensional case where the target segment corresponding to contactor node k is defined by nodes $k1$ and $k2$. The target point kt is the closest point of the target segment to the contactor node. The relative virtual displacement vector and the increment in the relative displacement vector between this point and the contactor node are given by

$$\bar{\mathbf{u}}_k^{IJ} = (1-\beta)\bar{\mathbf{u}}_{k1} + \beta\bar{\mathbf{u}}_{k2} - \bar{\mathbf{u}}_k \tag{42}$$

$$\Delta\mathbf{u}_k^{IJ} = (1-\beta)\Delta\mathbf{u}_{k1} + \beta\Delta\mathbf{u}_{k2} - \Delta\mathbf{u}_k, \tag{43}$$

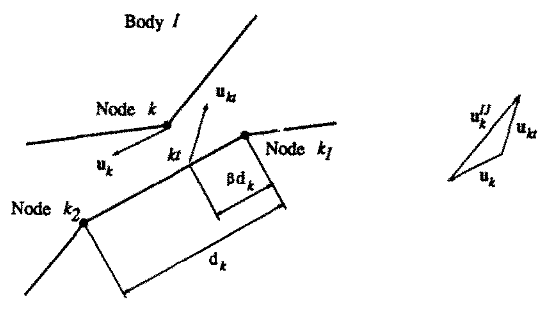

Fig. 1. Example of a two-dimensional contactor node and target segment.

where β is the ratio of the distance from node $k1$ to point kt over the length of the segment.

Let $\tau_k^T = (\lambda_k, \gamma_k, t_{1k}, t_{2k})$ be the vector of contact traction variables corresponding to contactor node k. Denote by $\hat{\tau}^T = (\tau_1^T, \ldots, \tau_m^T)$ the vector of all contact traction variables.

Let $\hat{\mathbf{u}}$ denote the vector of all unknown nodal point displacements. If $\mathbf{F}(\hat{\mathbf{u}})$ is the vector of internal nodal point forces resulting from the internal virtual work term in (37) and $\mathbf{R}$ is the vector of external nodal point forces resulting from the external virtual work terms in (37), we can write the finite element equations corresponding to (37)–(38) as

$$\mathbf{F}(\hat{\mathbf{u}}) + \mathbf{R}_c(\hat{\mathbf{u}}, \hat{\tau}) - \mathbf{R} = \mathbf{0} \tag{44}$$

$$\mathbf{F}_c(\hat{\mathbf{u}}, \hat{\tau}) = \mathbf{0}. \tag{45}$$

The vector $\mathbf{R}_c$ is obtained by assembling for all $k = 1, \ldots, m$ the contactor nodal force vector $\mathbf{R}_k^c$ corresponding to nodes k, $k1, \ldots, kp$, given by

$$\mathbf{R}_k^c = \begin{bmatrix} -(\lambda_k \mathbf{n}_k + t_{1k}\mathbf{s}_{1k} + t_{2k}\mathbf{s}_{2k}) \\ \beta_{k1}(\lambda_k \mathbf{n}_k + t_{1k}\mathbf{s}_{1k} + t_{2k}\mathbf{s}_{2k}) \\ \vdots \\ \beta_{kp}(\lambda_k \mathbf{n}_k + t_{1k}\mathbf{s}_{1k} + t_{2k}\mathbf{s}_{2k}) \end{bmatrix}, \tag{46}$$

where the coefficients β_{kr} have been defined in (39). The vector $\mathbf{F}_c$ can be written as $\mathbf{F}_c^T = \{\mathbf{F}_1^{cT}, \ldots, \mathbf{F}_m^{cT}\}$, where

$$\mathbf{F}_k^c = \begin{bmatrix} w(g_k, \lambda_k) \\ w(h_k, \gamma_k) \\ \lambda_k(\mathbf{s}_1 \cdot \Delta\mathbf{u}_k^{IJ}/\Delta t - \gamma_k t_{1k}) \\ \lambda_k(\mathbf{s}_2 \cdot \Delta\mathbf{u}_k^{IJ}/\Delta t - \gamma_k t_{2k}) \end{bmatrix} \tag{47}$$

and $h_k = \mu\lambda_k - \|\mathbf{t}_k\|$.

Equations (44) and (45) constitute a system of $n + 4m$ nonlinear equations in the $n + 4m$ unknowns $\hat{\mathbf{u}}$, $\hat{\tau}$. It is important to note that, contrary to the standard formulations, (44) and (45) contain explicitly all the inequality constraints associated with the contact conditions. This system is, however, highly nonlinear and a robust solution technique must be used.

The incremental equations corresponding to one iteration of the Newton method can be written as

$$\begin{bmatrix} D_{\hat{u}}\mathbf{F} + D_{\hat{u}}\mathbf{R}_c & D_{\hat{\tau}}\mathbf{R}_c \\ D_{\hat{u}}\mathbf{F}_c & D_{\hat{\tau}}\mathbf{F}_c \end{bmatrix} \begin{bmatrix} \Delta\hat{\mathbf{u}} \\ \Delta\hat{\boldsymbol{\tau}} \end{bmatrix} = \begin{bmatrix} -\mathbf{F} - \mathbf{R}_c + \mathbf{R} \\ -\mathbf{F}_c \end{bmatrix}. \quad (48)$$

The matrix $D_{\hat{u}}\mathbf{F}$ is the usual tangent stiffness matrix not including contact conditions. The matrix $D_{\hat{u}}\mathbf{R}_c$ represents the nonlinear contact stiffness contribution to the displacement degrees of freedom. The matrices $D_{\hat{\tau}}\mathbf{R}_c$ and $D_{\hat{u}}\mathbf{F}_c$ are obtained by assembling for all $k = 1, \ldots, m$ the contactor node matrices $D_{\hat{\tau}_k}\mathbf{R}^c_k$ and $D_{\hat{u}_k}\mathbf{F}^c_k$, which are given by

$$D_{\hat{\tau}_k}\mathbf{R}^c_k = \begin{bmatrix} -\mathbf{n}_k & \mathbf{0} & -\mathbf{s}_{1k} & -\mathbf{s}_{2k} \\ \beta_{k1}\mathbf{n}_k & \mathbf{0} & \beta_{k1}\mathbf{s}_{1k} & \beta_{k1}\mathbf{s}_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{kp}\mathbf{n}_k & \mathbf{0} & \beta_{kp}\mathbf{s}_{1k} & \beta_{kp}\mathbf{s}_{2k} \end{bmatrix} \quad (49)$$

$$D_{\hat{u}_k}\mathbf{F}^c_k = \begin{bmatrix} w_{,1}(g_k, \lambda_k)\mathbf{n}_k^T & -w_{,1}(g_k, \lambda_k)\beta_{k1}\mathbf{n}_k^T & \ldots & -w_{,1}(g_k, \lambda_k)\beta_{kp}\mathbf{n}_k^T \\ \mathbf{0}^T & \mathbf{0}^T & \ldots & \mathbf{0}^T \\ -\lambda_k\mathbf{s}_{1k}^T/\Delta t & \lambda_k\beta_{k1}\mathbf{s}_{1k}^T/\Delta t & \ldots & \lambda_k\beta_{kp}\mathbf{s}_{1k}^T/\Delta t \\ -\lambda_k\mathbf{s}_{2k}^T/\Delta t & \lambda_k\beta_{k1}\mathbf{s}_{2k}^T/\Delta t & \ldots & \lambda_k\beta_{kp}\mathbf{s}_{2k}^T/\Delta t \end{bmatrix}, \quad (50)$$

where $w_{,1}$ is the partial derivative of w with respect to its first argument and we have used the fact that $\partial g_k/\partial\mathbf{u}_k = \mathbf{n}_k$ and $\partial g_k/\partial\mathbf{u}_{kr} = -\beta_{kr}\mathbf{n}_k$. This follows from (22). Note that the matrices $D_{\hat{\tau}_k}\mathbf{R}^c_k$ and $D_{\hat{u}_k}\mathbf{F}^c_k$ are related by

$$D_{\hat{u}_k}\mathbf{F}^c_k = \mathbf{C}_k(D_{\hat{\tau}_k}\mathbf{R}^c_k)^T \quad (51)$$

where $\mathbf{C}_k$ is the 4×4 diagonal matrix

$$\mathbf{C}_k = \begin{bmatrix} -w_{,1}(g_k, \lambda_k) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \lambda_k/\Delta t & 0 \\ 0 & 0 & 0 & \lambda_k/\Delta t \end{bmatrix}. \quad (52)$$

Matrix $D_{\hat{\tau}}\mathbf{F}_c$ has the form

$$D_{\hat{\tau}}\mathbf{F}_c = \begin{bmatrix} D_{\hat{\tau}_1}\mathbf{F}^c_1 & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ 0 & \ldots & D_{\hat{\tau}_m}\mathbf{F}^c_m \end{bmatrix}, \quad (53)$$

where each of the diagonal blocks has the form

$$D_{\hat{\tau}_k}\mathbf{F}^c_k = \begin{bmatrix} w_{,2}(g_k, \lambda_k) & 0 & 0 & 0 \\ \mu w_{,1}(h_k, \gamma_k) & w_{,2}(h_k, \gamma_k) & -w_{,1}(h_k, \gamma_k)t_{1k}/\|\mathbf{t}_k\| & -w_{,1}(h_k, \gamma_k)t_{2k}/\|\mathbf{t}_k\| \\ (\mathbf{s}_1 \cdot \Delta\mathbf{u}_k^{IJ}/\Delta t - \gamma_k t_{1k}) & -\lambda_k t_{1k} & -\lambda_k\gamma_k & 0 \\ (\mathbf{s}_2 \cdot \Delta\mathbf{u}_k^{IJ}/\Delta t - \gamma_k t_{2k}) & -\lambda_k t_{2k} & 0 & -\lambda_k\gamma_k \end{bmatrix} \quad (54)$$

and $w_{,2}$ is the partial derivative of w with respect to its second argument. Note that nonzero coefficients for $\Delta\gamma_k$ only appear in (54).

Let $p(\hat{\mathbf{u}}, \hat{\boldsymbol{\tau}})$ be defined as

$$p(\hat{\mathbf{u}}, \hat{\boldsymbol{\tau}}) = \tfrac{1}{2}\|\mathbf{F}(\hat{\mathbf{u}}) + \mathbf{R}_c(\hat{\mathbf{u}}, \hat{\boldsymbol{\tau}}) - \mathbf{R}\|^2 + \tfrac{1}{2}c\|\mathbf{F}_c(\hat{\mathbf{u}}, \hat{\boldsymbol{\tau}})\|^2, \quad (55)$$

where c is an appropriate scaling constant. Using $p(\hat{\mathbf{u}}, \hat{\boldsymbol{\tau}})$ as an objective function, line searching can be performed along the direction obtained from the Newton iteration, in order to enlarge the region of convergence of the method.

Note that the coefficient matrix in (48) is nonsymmetric in the rows and columns corresponding to contact.

We may also note that in the solution process initially the contact forces are zero and hence the additional contact equations for $\Delta\hat{\boldsymbol{\tau}}_k$ are added as node k establishes contact. These equations are then kept in the system of equations until convergence. Ideally, in the algorithm, the situation $\lambda_k = 0$ is only reached at convergence for nodes that are not in contact. However, if for some k, λ_k happens to be zero during the iterations prior to convergence, then the coefficient matrices (50) and (54) lead to a singular matrix. In this case the appropriate columns and rows have to be ignored for that iteration.

The full set of conditions under which this may occur will of course depend on the choice of the function w and whether friction is actually included.

4. CONCLUSIONS

A new approach for the treatment of contact conditions in large-strain finite element analysis has been presented. The formulation is attractive in that all inequality constraints corresponding to contact and friction are enforced explicitly. However, the resulting equations are highly nonlinear. Hence robust solution techniques to solve the resulting system of nonlinear equations must be used. A disadvantage is that in the Newton–Raphson iteration the coefficient matrix is in general nonsymmetric (due to friction conditions), and hence efficient techniques dealing with this nonsymmetry must be employed for solution.

Numerical tests of the method described in this paper will be presented in a forthcoming communication.

REFERENCES

1. E. A. Wilson and B. Parsons, Finite element analysis of elastic contact problems using differential displacements. *Int. J. Numer. Meth. Engng.* **2**, 387–395 (1970).
2. S. K. Chan and I. S. Tuba, A finite element method for contact problems of solid bodies—part I. Theory and validation. *Int. J. Mech. Sci.* **13**, 615–625 (1971).
3. H. Schäfer, A contribution to the solution of contact problems with the aid of bond elements. *Comput. Meth. Appl. Mech. Engng* **6**, 335–354 (1975).
4. A. Francavilla and O. C. Zienkiewicz, A note on numerical computation of elastic contact problems. *Int. J. Numer. Meth. Engng* **9**, 913–924 (1975).
5. D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd edn. Addison-Wesley, Reading, MA (1984).
6. D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, New York (1982).
7. P. Gill, W. Murray and M. Wright. *Practical Optimization.* Academic Press (1981).
8. A. R. Johnson and C. J. Quigley, Frictionless geometrically non-linear contact using quadratic programming. *Int. J. Numer. Meth. Engng* **28**, 127–144 (1989).
9. J. R. Hughes, R. L. Taylor, J. L. Sackman, A. Curnier and W. Kanoknukulchai, A finite element method for a class of contact-impact problems. *Comput. Meth. Appl. Mech. Engng* **8**, 249–276 (1976).
10. K. J. Bathe and A. B. Chaudhary, On finite element analysis of large deformation frictional contact problems. In *Proceedings of the 7th Invitational Symposium on the Unification of Finite Elements, Finite Differences and Calculus of Variations*, pp. A.165–A.189 (1984).
11. K. J. Bathe and A. B. Chaudhary, A solution method for planar and axisymmetric contact problems. *Int. J. Numer. Meth. Engng* **21**, 65–88 (1985).
12. A. B. Chaudhary and K. J. Bathe, A solution method for static and dynamic analysis of three-dimensional contact problems with friction. *Comput. Struct.* **24**, 855–873 (1986).
13. N. Kikuchi and Y. J. Song, Remarks on relations between penalty and mixed finite element methods for a class of variational inequalities. *Int. J. Numer. Meth. Engng* **15**, 1557–1561 (1980).
14. J. T. Oden, Exterior penalty methods for contact problems in elasticity. In *Nonlinear Finite Element Analysis in Structural Mechanics* (Edited by W. Wunderlich, E. Stein and K. J. Bathe). Springer, New York (1981).
15. J. T. Oden and N. Kikuchi, Finite element methods for constrained problems in elasticity. *Int. J. Numer. Meth. Engng* **18**, 701–725 (1982).
16. N. Kikuchi, A smoothing technique for reduced integration penalty methods in contact problems. *Int J. Numer. Meth. Engng* **18**, 343–350 (1982).
17. J. O. Hallquist, G. L. Goudreau and D. J. Benson, Sliding interfaces with contact-impact in large-scale Lagrangian computations. *Comput. Meth. Appl. Mech. Engng* **51**, 107–137 (1985).
18. J. T. Oden and G. F. Carey, *Finite Elements, Special Problems in Solid Mechanics*, Vol. V. Prentice-Hall, Englewood Cliffs, NJ (1984).
19. J. C. Simo, P. Wriggers and R. L. Taylor, A perturbed Lagrangian formulation for the finite element solution of contact problems. *Comp. Meth. Appl. Mech. Engng* **50**, 163–180 (1985).
20. P. Wriggers and J. C. Simo, A note on tangent stiffness for fully nonlinear contact problems. *Commun. Appl. Numer. Meth.* **1**, 199–203 (1985).
21. B. Nour-Omid and P. Wriggers, A two-level iteration method for solution of contact problems. *Comput. Meth. Appl. Mech. Engng* **54**, 131–144 (1986).
22. P. Wriggers, J. C. Simo and R. L. Taylor, Penalty and augmented Lagrangian formulations for contact problems. In *Proceedings of the International Conference on Numerical Methods in Engineering—Theory and Applications, NUMETA '85* (Edited by J. Middleton and G. N. Pande), pp. 97–106 (1985).
23. P. Wriggers, On consistent tangent matrices for frictional contact problems. In *Proceedings of the International Conference on Numerical Methods in Engineering—Theory and Applications, NUMETA '87*, (Edited by G. N. Pande and J. Middleton) (1987).
24. H. Parisch, A consistent tangent stiffness matrix for three-dimensional non-linear contact analysis. *Int. J. Numer. Meth. Engng* **23**, 1803–1812 (1989).
25. K. J. Bathe, *Finite Element Procedures in Engineering Analysis.* Prentice-Hall, Englewood Cliffs, NJ (1982).
26. E. Rabinowicz, *Friction and Wear of Materials.* John Wiley (1965).
27. J. T. Oden and E. B. Pires, Nonlocal and nonlinear friction laws and variational principles for contact problems in elasticity. *J. Appl. Mech.* **50**, 67–76 (1983).
28. J. T. Oden and J. A. C. Martins, Models and computational methods for dynamic friction phenomena. *Comp. Meth. Appl. Mech. Engng* **52**, 527–634 (1985).
29. P. Wriggers, T. Vu Van and E. Stein, Finite element formulation of large deformation impact-contact problems with friction. *Comput. Struct.* **37**, 319–331 (1990).
30. O. L. Mangasarian, Equivalence of the complementarity problem to a system of nonlinear equations. *J. Appl. Math.* **31**, 89–91 (1976).
31. L. T. Watson and R. T. Haftka, Modern homotopy methods in optimization. *Comput. Meth. Appl. Mech. Engng* **74**, 289–305 (1989).