Int. J. Solids Structures Vol. 22, No. 12, pp. 1377 1398, 1986
Printed in Great Britain.

0020 7683/86 $3.00+.00
Pergamon Journals Ltd.

GENERAL CONTACT BOUNDARY CONDITIONS AND THE ANALYSIS OF FRICTIONAL SYSTEMS

A. KLARBRING
Linköping Institute of Technology, Department of Mechanical Engineering, Division of Solid Mechanics and Strength of Materials, Linköping, Sweden

(*Received* 11 *February* 1985; *in revised form* 15 *November* 1985)

Abstract—General frictional systems, i.e. solid bodies interacting by contact forces, are investigated and variational formulations are presented. Following recent ideas, general contact boundary conditions are formulated using the notions of subdifferentials and generalized gradients.

1. INTRODUCTION

Questions concerning the interaction of machine elements by frictional forces are frequently raised in modern design situations. Nevertheless, the application of analytical methods to such problems has been limited. The reason for this is perhaps that a theoretical model, describing the general behaviour of *frictional systems* (i.e. an assembly of solid bodies interacting by normal and tangential contact forces), has only recently begun taking shape. This development has consisted of two branches: firstly, applications of ideas and concepts from the classical theory of plasticity to frictional behaviour[1–4]; and, secondly, applications and extensions of ideas from the theory of variational inequalities[5–10] to contact problems with friction.

Unilaterality is a property common to all problems connected with frictional systems. That is, while most treatments of mechanical systems involve introduction of linear relations as a first approximation of non-linear phenomena, the physical phenomena of contact and friction are usually approximated by non-differentiable relations, which do not admit linearization. Therefore, such problems have to be treated by mathematics that does not rely on the existence of classical two-side derivatives, but instead admits a one-side, or unilateral analysis. The notion of the subdifferential[11–13] provides a generalization of differentiation that admits the use of convex, non-differentiable "superpotentials" when describing unilateral phenomena[13, 14]. It suggests that one should look for "convexification" instead of linearization when treating non-linear problems. Convexity is usually naturally present in mechanics and is in some sense implied by a "stable" mechanical behaviour. However, a number of problems of practical importance do not have this property. In such cases, the notion of the generalized gradient, developed in Ref. [15], and introduced into mechanics by Panagiotopoulos[16], can replace the subdifferential; i.e. non-convex superpotentials can be admitted. In this paper we will, by means of convex and non-convex superpotentials, treat some problems connected with frictional systems. Therefore, in Section 2 we present some mathematical concepts related to subdifferentials and generalized gradients which will hopefully make the presentation reasonably self-contained. Notice also that non-convex superpotentials have recently been used to study problems of finite elastoplasticity[35, 36].

The similarity between frictional and elasto-plastic behaviour is obvious from experimental investigations[17]. Nevertheless, the level of development reached by the theory of plasticity is hardly matched by any theory of friction. Almost all investigations of contact problems in solid mechanics assume either negligible frictional effects, perfect adhesion of Coulomb's eighteenth century law of dry friction. However, some attempts to use more general laws of friction have been made. Duvaut[6], and Oden and Pires[7] have introduced "non-local and non-linear" friction laws for which theoretical and numerical investigations have been carried out[8, 9]. Their arguments mainly deal with consistency of the theoretical model and possibility of numerical treatment. Another line of research has been inspired by the above-mentioned similarity with elasto-plastic behaviour[1–3]. These ideas were recently extended to Curnier[4], who proposed a model of friction similar to that for standard generalized materials[18]. In Section 3 of this paper we follow Curnier and his theory is given a more general mathematical setting; this is achieved through the use of possibly non-convex superpotentials. Moreover, in the same way as we describe Curnier's law of friction, here called an elasto-resistance law, we define static laws of contact. They are usually applicable when force and displacement components normal to the contact surface are under consideration.

In order to apply the introduced contact boundary conditions in an analysis of an actual frictional system we have to consider the behaviour of the deformable bodies involved. We thereby assume them to be linear elastic. Furthermore, in order to stress the general structure of the theory, we will assume the displacement fields to belong to a finite dimensional vector space. Such a representation can be obtained by a finite element approximation of infinite dimensional displacement fields. In Section 4 we obtain relations between elements of contact boundary spaces, i.e. the subspaces of the displacement and force spaces where the nonlinearity of the problem occurs. The general contact boundary conditions are combined with these relations in Section 5. It is seen that the variational formulations of the considered problems are of inequality type, due to the unilateral character of contact and friction behaviour. Furthermore, they are hemi-variational inequalities and not, what is more familiar, variational inequalities; this is due to the presence of non-convex superpotentials. The hemi-variational inequality is a type of inequality which has recently been introduced by Panagiotopoulos[16] as a direct consequence of the definition of the generalized gradient.

Concerning the numerical treatment of problems formulated as hemi-variational inequalities, it has previously been suggested that one should introduce a smoothing of the non-differentiable superpotentials, and thereby obtain a variational equality[19]. We point out that one can in special cases use a different method: if the contact boundary conditions can be expressed using non-negative multipliers, then the problems can be formulated as various forms of linear complementarity problems (LCP)—a mathematical structure known from the area of mathematical programming. Several different solution algorithms are then applicable. A similar mathematical problem also arises when analysing elasto-plastic structures, and extensive investigations have been carried out in this case[20–22]. Several useful analogies between frictional systems and elasto-plastic structures can therefore be stated[23].

2. SOME CONCEPTS FROM CONVEX AND NON-CONVEX OPTIMIZATION

Consider a pair of normed spaces $\{X, X'\}$ placed in separating duality by a bilinear form $\langle y, x\rangle_X$, where $x \in X$ and $y \in X'$. This is a convenient mathematical setting for many problems of mechanics[24] and such a system has been called a mechanical element (the French term in Ref. [24] is "élement mécanique"). For the needs of this paper one could consider X and X' as finite dimensional spaces.

Denote by f a function on X (or on X'), with values in $\bar{R} = [-\infty, +\infty]$. We define a subgradient of f at x, if $f(x)$ is finite, as $y \in X'$ such that

$$f(x') - f(x) \geqslant \langle y, x' - x\rangle_X \,\forall\, x' \in X. \tag{1}$$

The set of all subgradients of f at x is the subdifferential of f at x, which we will denote by $\partial f(x)$. We have

$$\partial f(x) = \{y \in X' : f(x') - f(x) \geqslant \langle y, x' - x\rangle_X \,\forall\, x' \in X\}. \tag{2}$$

Relation (1) could due to eqn (2) be equivalently written as

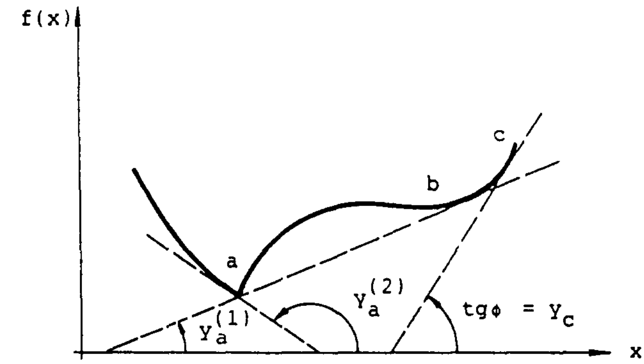

Fig. 1. The subdifferential: $\partial f(x_a) = [y_a^{(2)}, y_a^{(1)}]$, $\partial f(x_c) = y_c$, $\partial f(x) = \varnothing$ for $x_a < x < x_b$.

$$y \in \partial f(x). \tag{3}$$

The notion of the subdifferential is in most cases not useful when f is a non-convex function. As seen in Fig. 1 the subdifferential is then an empty set in cases where ordinary derivatives obviously exist. To overcome this Clarke[15] introduced a local convexification of f in terms of the directional derivative f^0. This original definition of f^0 encompasses functions that are locally Lipschitzian.† It was later extended to lower semi-continuous (lsc) functions‡ by Rockafellar[12]. For the needs of this paper, it is sufficient to be able to define an extension of the subdifferential in cases when f is directionally Lipschitzian.§ This can be achieved, in an unambiguous manner, by the use of the extended Clarke derivative $f^0(x, z)$ of f at x in the z-direction[12]. If $f(x)$ is finite we have

$$f^0(x, z) = \limsup_{\substack{x' \to_f x \\ t \downarrow 0}} \frac{f(x' + tz) - f(x')}{t} \tag{4}$$

where the notation $x' \to_f x$ means that $x' \to x$ and $f(x') \to f(x)$. If f is locally Lipschitzian, $x' \to x$ can be written in place of $x' \to_f x$ and Clark's original definition is re-established.

It could be seen that $f^0(x, z)$ is convex in the second variable and we define the generalized gradient of f at x to be the subdifferential of the function $f^0(x, \cdot)$ at 0; if $f(x)$ is infinite the generalized gradient at x is not defined. Furthermore, if f is convex, the generalized gradient coincides with the subdifferential. Therefore, the former is a generalization of the latter and we will use the notation $\partial f(x)$ for generalized gradient in the sequel. From relation (1) it follows that

$$f^0(x, x' - x) \geqslant \langle y, x' - x \rangle_X \, \forall x' \in X \tag{5}$$

for $y \in \partial f(x)$. A multivalued relation of the form $y \in \partial f(x)$ will be called a subdifferential relation, both in the convex and in the non-convex case.

† f is Lipschitzian on an open set U if there exists a number $\lambda \geqslant 0$ such that

$$|f(x') - f(x)| \leqslant \lambda \|x' - x\| \qquad \forall x', x \in U.$$

f is locally Lipschitzian on an open set U if it is Lipschitzian on some neighbourhood of each $x \in U$.

‡ f is lower semi-continuous on U if

$$f(x) \leqslant \liminf_{x' \to x} f(x') \, \forall x \in U.$$

§ f is directionally Lipschitzian at x if there exists a vector z such that

$$\limsup_{\substack{x' \to_f x \\ z' \to z \\ t \downarrow 0}} \frac{f(x' + tz') - f(x')}{t} < \infty.$$

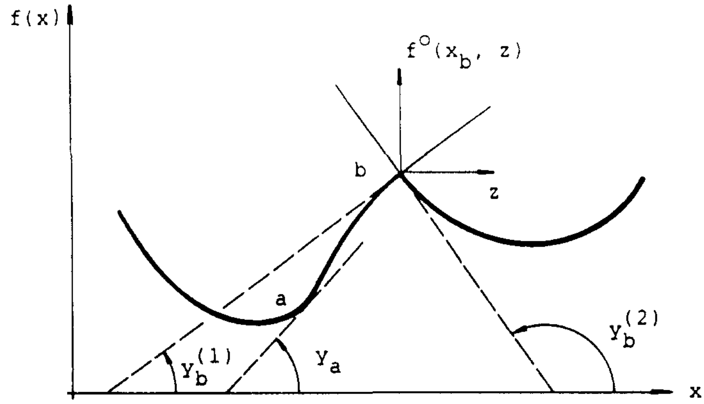

Fig. 2. The generalized gradient: $\partial f(x_a) = y_a$, $\partial f(x_b) = [y_b^{(2)}, y_b^{(1)}]$.

It should be noted that if the Gâteaux derivative $Df(x)$ of f exists and is continuous, then $\partial f(x) = \{Df(x)\}$. This theory therefore constitutes a proper generalization of concepts developed for smooth functions in the calculus of variations.

In the following the notion of normal cones will be of interest. We assume that X equals R^n and consider the distance function d_C of the non-empty closed set $C \subset R^n$

$$d_C(x) = \min\{|x - c| : c \in C\}. \tag{6}$$

The normal cone to C at $x \in C$, denoted by $N_C(x)$, is defined as the closure of the set

$$\{y \in R^n : ty \in \partial d_C(x) \text{ for some } t > 0\}. \tag{7}$$

Of special interest is the case when

$$C = \{x \in R^n : \phi_i(x) \leqslant 0, i = 1, \ldots, m\} \tag{8}$$

where ϕ_i are continuously differentiable real-valued functions. Let B denote the set of indices such that $\phi_i \leqslant 0$ is satisfied as an equality. If there exists a $z \in R^n$ such that

$$\langle z, \nabla\phi_i(x)\rangle_{R^n} < 0 \quad \forall i \in B \tag{9}$$

then

$$N_C(x) =$$

$$\left\{ y \in R^n : y = \sum_{i=}^{m} \lambda_i \nabla\phi_i(x), \quad \lambda_i \geqslant 0, \quad \phi_i(x) \leqslant 0, \quad \lambda_i\phi_i(x) = 0, \quad i = 1, \ldots, m \right\}. \tag{10}$$

We further introduce the concept of an indicator function ψ_C of the closed set $C \subset X$

$$\psi_C(x) = \begin{cases} 0 \text{ if } x \in C \\ \infty \text{ if } x \notin C. \end{cases} \tag{11}$$

It can be shown that

$$N_C(x) = \partial \psi_C(x). \tag{12}$$

If a function f belongs to the class of convex, lsc and proper† functions[11–13], its conjugated function

$$f^c(y) = \sup_{x \in X} \{\langle y, x \rangle_X - f(x)\} \tag{13}$$

may be of interest. It is, in this case, convex, lsc, and proper. Furthermore, $(f^c)^c = f$. The conjugated function introduces a method of "inversion" of relation (3), since the following relations are equivalent:

$$y \in \partial f(x) \tag{14}$$

$$x \in \partial f^c(y) \tag{15}$$

$$f(x) + f^c(y) = \langle y, x \rangle_X. \tag{16}$$

The conjugated function of the indicator functions is the support function

$$f^c(y) = \sup_{x \in C} \langle y, x \rangle_X. \tag{17}$$

3. CONTACT BOUNDARY CONDITIONS

When solid bodies interact by contact forces, complex physical phenomena occur in the interface between them[25]. On a continuum mechanical scale these phenomena display themselves through resistance against normal penetration and tangential relative movement. Such interaction is described by contact boundary conditions in the form of phenomenological contact laws, i.e. relations between (relative) displacements and forces‡ occurring on the contact surfaces. On the basis of the theory presented in the previous section three different classes of such laws will be considered in this section.

The first one consists of static laws, which are generally assumed to govern the behaviour of components normal to the contact surface, when a small displacement assumption is applied.

The second one consists of laws which can be used to describe a variety of non-reversible phenomena, such as those of viscosity or plasticity type. A proper name seems to be elasto-resistance laws. They can be assumed to govern the behaviour of components tangential to the contact surface. Using a terminology widely used in plasticity the static laws would be called holonomic and the elasto-resistance laws non-holonomic.

The final class consists of incremental laws. They apply when an incremental description of the problem is used. In contrast to the two foregoing types of contact boundary relations, which generally corresponds to different physical events, incremental laws describes both reversible and non-reversible force–displacement relations. Some static laws can although be described by elasto-resistance laws[13].

Note that the laws in this section are assumed to hold pointwisely on the contact surface. They will be extended to global relations between contact boundary spaces in Section 5. Furthermore, indices N and T chiefly refer to normal and tangential force and displacement components, although, for example, static laws need not always relate normal ones, as it is seen in Example 2.

3.1. *Static laws*

The notation X_p will, throughout this paper, denote a finite dimensional vector space

† f is proper if $f(x) > -\infty$ for all x, and $f(x) < \infty$ for at least one x.
‡ The term force denotes a force density if contact extends over a hole region.

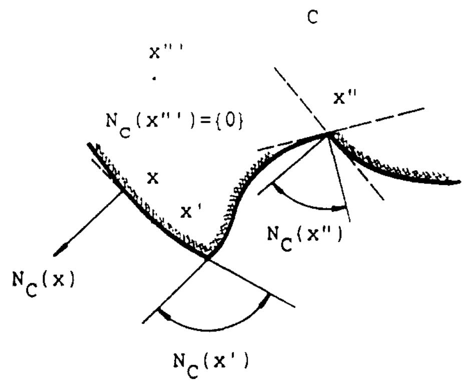

Fig. 3. Normal cones to the closed set C.

of dimension p and $\langle\cdot,\cdot\rangle_{X_p}$ denotes dual pairing between this space and its dual X_p'. If a contact surface displacement w_N belongs to X_r, then $\langle P_N, w_N\rangle_{X_r}$ represents the work done by the contact force $P_N \in X_r'$. The following subdifferential relation between the contact force and displacement represents a static contact law

$$w_N \in \partial J_N(-P_N) \tag{18}$$

where J_N is a convex or non-convex superpotential. Relation (18) was first given by Moreau in 1968 for the convex case and by Panagiotopoulos in 1983 for the non-convex case.

In the case where J_N is convex, lsc and proper, the conjugated function J_N^c can be introduced in order to obtain the inverse relation of (18). The minus sign in relation (18) is used for consistency with the concept of potential force–displacement relations[13], i.e. the case when J_N is differentiable.

Example 1

Assume that w_N, $P_N \in R$ are governed by the following relations:

$$\begin{aligned} &\text{if } w_N - g < 0 \text{ then } P_N = 0 \\ &\text{if } w_N - g \geqslant 0 \text{ then } P_N = -Q_N(w_N - g) \end{aligned} \tag{19}$$

where Q_N describes the "stiffness" of the contact surface and g measures the contact "gap" in the undeformed state. This law can be described using a subdifferential relation if we introduce the convex set $C_N = \{-P_N \in R: P_N \leqslant 0\}$ and denote by Ψ_{C_N} the indicator function of C_N. Relations (19) are then equivalent to

$$w_N - g + Q_N^{-1}P_N \in \partial\Psi_{C_N}(-P_N) = N_{C_N}(-P_N) \tag{20}$$

where the equality follows from eqn (12). It further follows from relation (10) that

$$N_{C_N}(-P_N) = \{v \in R: v = -\lambda, \lambda \geqslant 0, P_N \leqslant 0, \lambda P_N = 0\} \tag{21}$$

and we obtain the following relation, which will prove to be useful in the sequel:

$$w_N = g - Q_N^{-1}P_N - \lambda, \qquad \lambda \geqslant 0, \qquad P_N \leqslant 0, \qquad \lambda P_N = 0. \tag{22}$$

Example 2

A simplified form of Coulomb's law of dry friction can be described using a holonomic relation[10]: the force normal to the contact surface is considered known and the non-reversible character of friction is ignored.

Assume that w_T, $P_T \in R^2$ are governed by the following relations:

$$\begin{aligned} &\text{if } |P_T| < \mu |R_N| \text{ then } w_T = 0 \\ &\text{if } |P_T| = \mu |R_N| \text{ then } w_T = -\lambda P_T, \lambda \geqslant 0 \end{aligned} \tag{23}$$

where μ is the coefficient of friction and R_N is the known normal contact force. Let Ψ_{C_T} denote the indicator function of the convex set $C_T = \{-P_T \in R^2 : |P_T| \leqslant \mu |R_N|\}$. Relations (23) are then equivalent to

$$w_T \in \partial \Psi_{C_T}(-P_T) \tag{24}$$

which can be shown by using the identification between $\partial \Psi_{C_T}(-P_T)$ and the normal cone $N_{C_T}(-P_T)$, as shown in Example 1. ■

3.2. *Elasto-resistance laws*

Here we will present a model that is capable of describing a variety of non-reversible contact boundary relations. It is a development of concepts normally used in the theory of plasticity and it uses ideas that have recently been introduced by Curnier[4].

As was the case with the displacements w_N, when static laws were considered, the contact displacements w_T under consideration will be assumed to belong to a finite dimensional vector space, i.e. $w_T \in X_s$. This implies that the time rate of change $\dot{w}_T$ of w_T also belongs to X_s. The dual pairing $\langle P_T, \dot{w}_T \rangle_{X_s}$ represents the power of the contact force $P_T \in X_s'$ if the rate of the contact displacement is $\dot{w}_T$.

Assume that the displacement $w_T \in X_s$ can be decomposed into a reversible part w_T^A and a non-reversible one w_T^S[4], so that

$$w_T = w_T^A + w_T^S. \tag{25}$$

Following the ideas of Curnier[4], we extend the notion of standard generalized materials[18] to contact boundary relations. That is, we assume the existence of a set of state variables $(w_T^A, \alpha) \in X_s \times X_t$, where α are internal parameters. Associated to these variables is the set $(-P_T, A) \in X_s' \times X_t'$ of force variables

$$-P_T = Q_T w_T^A \tag{26}$$

$$A = H\alpha \tag{27}$$

where Q_T and H are linear transformations. The rate equation, for the non-reversible displacement rate $\dot{w}_T^S$ and the rate of the internal parameters $\dot{\alpha}$, is given by a law of "standard generalized friction"

$$(\dot{w}_T^S, -\dot{\alpha}) \in \partial J_T(-P_T, A; P_N). \tag{28}$$

The superpotential J_T is assumed to depend on $P_N \in X_r'$, which is shown in relation (28). Subdifferentiation is performed with respect to variables standing in front of ";". Relations (25)–(28) represent the most general form of an elasto-resistance law.

The role played by $P_N \in X_r'$ in these laws is notable; P_N-dependent superpotentials are introduced without P_N being an internal force. This makes the elasto-resistance laws

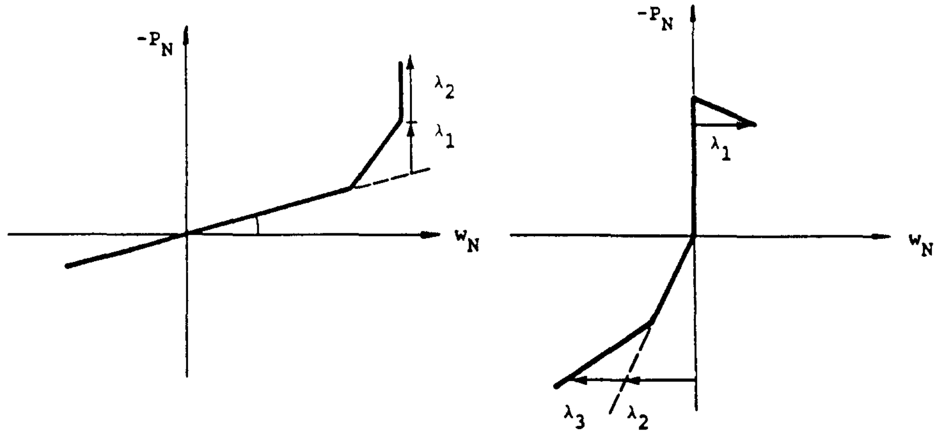

Fig. 4. Piecewise linear static laws expressed by non-negative multipliers.

correspond to non-associated plastic stress–strain relations and many qualitative features of elasto-plastic boundary value problems, where such material laws are present, should therefore be expected to be found also for frictional problems.

In the case where J_T is convex, lsc and proper for a prescribed P_N, a conjugated function J_T^c can be introduced

$$(-P_T, A) \in \partial J_T^c(\dot{w}_T^S, -\dot{\alpha}; P_N). \tag{29}$$

When the theory is rate independent, J_T becomes the indicator function Ψ_{C_T} of a closed set $C_T(P_N) \subset X_s' \times X_t'$, depending on P_N. That is

$$(\dot{w}_T^S, -\dot{\alpha}) \in \partial \Psi_{C_T(P_N)}(-P_T, A) = N_{C_T(P_N)}(-P_T, A) \tag{30}$$

where, as before, the equality follows from eqn (12).

Example 3

Assume that

$$C_T(P_N) = \{(-P_T, A) \in R^s \times R^t : \phi_i(-P_T, A; P_N) \leqslant 0, i = 1, \ldots, q\} \tag{31}$$

where $\phi_i : R^r \times R^s \times R^t \to R$ is continuously differentiable on the space $R^s \times R^t$ for a prescribed P_N. Let us denote by B the set of indices i such that $\phi_i \leqslant 0$ is satisfied as an equality. Assume that there exists a $z \in R^s \times R^t$ such that

$$\langle z, \nabla \phi_i(-P_T, A; P_N) \rangle_X < 0 \qquad \forall i \in B \tag{32}$$

where, in this case, $\langle \cdot, \cdot \rangle_X$ denotes inner product on $R^s \times R^t$ and where ∇ denotes the gradient with respect to $(-P_T, A)$. The normal cone $N_{C_T(P_N)}(-P_T, A)$ can then be represented as

$$\begin{aligned} N_{C_T(P_N)}(-P_T, A) = \{(\dot{w}_T^S, -\dot{\alpha}) \in R^s \times R^t : \\ (\dot{w}_T^S, -\dot{\alpha}) = \sum_{i=1}^{q} \lambda_i \nabla \phi_i(-P_T, A; P_N), \qquad \lambda_i \geqslant 0, \\ \phi_i(-P_T, A; P_N) \leqslant 0, \qquad \lambda_i \phi_i(-P_T, A; P_N) = 0, \qquad i = 1, \ldots, q\}. \end{aligned} \tag{33}$$

That is

$$\dot{w}_T^S = \sum_{i=1}^{q} \dot{\lambda}_i \nabla\phi_i \tag{34a}$$

$$-\dot{\alpha} = \sum_{i=1}^{q} \dot{\lambda}_i \bar{\nabla}\phi_i \tag{34b}$$

$$\dot{\lambda}_i\phi_i = 0, \qquad \dot{\lambda}_i \geqslant 0, \qquad \phi_i \leqslant 0, \qquad i = 1,\ldots,q. \tag{34c}$$

Here ∇ and $\bar{\nabla}$ denote the gradient with respect to $-P_T$ and A, respectively. All evaluations of ϕ_i, $\nabla\phi_i$ and $\bar{\nabla}\phi_i$ are taken at the point $(-P_T, A; P_N)$.

We obtain the following law when combining eqns (25) and (26) with relations (34a) and (34c):

$$\dot{w}_T = -Q_T^{-1}\dot{P}_T + \sum_{i=1}^{q} \dot{\lambda}_i \nabla\phi_i \tag{35a}$$

$$\dot{\lambda}_i\phi_i = 0, \qquad \dot{\lambda}_i \geqslant 0, \qquad \phi_i \leqslant 0, \qquad i = 1,\ldots,q. \tag{35b}$$

The rates of the constraints are given by

$$\dot{\phi}_i = \langle \nabla\phi_i, -\dot{P}_T \rangle_{R^s} + \langle \bar{\nabla}\phi_i, \dot{A} \rangle_{R^t} + \langle \bar{\bar{\nabla}}\phi_i, \dot{P}_N \rangle_{R^r} \tag{36}$$

where $\bar{\bar{\nabla}}\phi_i$ denotes the gradient of ϕ_i with respect to P_N, assuming its existence. Introducing eqns (27) and (34b) we obtain for the second term in eqn (36)

$$\langle \bar{\nabla}\phi_i, \dot{A} \rangle_{R^t} = -\langle \bar{\nabla}\phi_i, H \sum_{j=1}^{q} \dot{\lambda}_j \bar{\nabla}\phi_j \rangle_{R^t}. \tag{37}$$

Relations (35)–(37) represent an elasto-resistance law in terms of non-negative multipliers, and it is similar to Koiter's generalization of elasto-plastic laws to cases with singular yield surfaces. In Section 6 it will be shown useful for analysis of frictional systems. ■

Example 4

We will investigate Coulomb's classical law of friction. The closed set $C_T(P_N)$ is then given by

$$C_T(P_N) = \{-P_T \in R^2 : \phi = |P_T| - \mu|P_N| \leqslant 0\} \tag{38}$$

where μ is the coefficient of friction and $P_N \in R$. The number of internal variables is thus zero in this case. Since J_T is the indicator function of $C_T(P_N)$, J_T^c is given by the support function

$$\Psi^c_{C_T(P_N)}(\dot{w}_T^S) = \mu|P_N||\dot{w}_T^S|. \tag{39}$$

It is usually assumed that Q_T tends towards infinity and thus, w_T^A equals zero; i.e. $w_T = w_T^S$. The support function (39) is then referred to as the virtual work of the friction forces[6–9].

An extension of Coulomb's law to include internal parameters has been given by

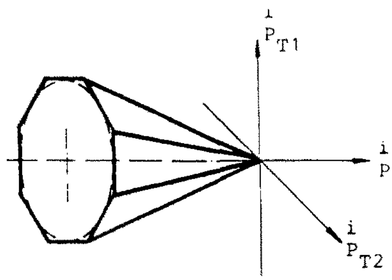

Fig. 5. Piecewise linearization of the closed set, eqn (38), i.e. a piecewise linear Coulomb's friction law.

Curnier[4]. The first one to describe Coulomb friction by a subdifferential relation was Moreau in 1970. ■

3.3. *Incremental laws*

If we consider, at a specific contact force state, the "response" $\dot{w} \in X_q$ due to a force rate $\dot{P} \in X_q'$, we will obtain what is usually called an incremental relation. Such relations will generally depend on the previous load history. However, if we consider the situation at a particular time instance, superpotentials $\hat{J}_N$ and $\hat{J}_T$ depending on the force rates $\dot{P}_N$ and $\dot{P}_T$ only, can be used to obtain incremental subdifferential contact laws. Hence, the following relations represent incremental laws:

$$\dot{w}N \in \partial \hat{J}_N(-\dot{P}_N) \tag{40}$$

$$\dot{w}_T \in \partial \hat{J}_T(-\dot{P}_T; \dot{P}_N). \tag{41}$$

If $\hat{J}_N[\hat{J}_T]$ is convex, lsc and proper, then we could, by the introduction of the conjugated function $\hat{J}_N^c$ $[\hat{J}_T^c]$, obtain the inverse of relation of (40) [(41)].

Example 5

Let us investigate Coulomb friction when described by an incremental law. Assume that the load history is such that $\phi = |P_T| - \mu|P_N| = 0$ in the present load state. The superpotential $\hat{J}_T$ will then be the indicator function $\Psi_{C_T(\dot{P}_N)}(-\dot{P}_T)$ of the convex set

$$C_T(\dot{P}_N) = \left\{ -\dot{P}_T \in R^2 : \dot{\phi} = \left\langle \frac{P_T}{|P_T|}, \dot{P}_T \right\rangle_{R^2} - \mu \left\langle \frac{P_N}{|P_N|}, \dot{P}_N \right\rangle_R \leqslant 0 \right\}. \tag{42}$$

The normal cone of this set is given by

$$N_{C_T(\dot{P}_N)}(-\dot{P}_T) = \left\{ w_T \in R^2 : w_T = -\dot{\lambda} \frac{P_T}{|P_T|}, \dot{\lambda} \geqslant 0, \dot{\phi}(-\dot{P}_T, \dot{P}_N) \leqslant 0, \dot{\lambda}\dot{\phi}(-\dot{P}_T, \dot{P}_N) = 0 \right\} \tag{43}$$

which introduces an incremental description of Coulomb's friction law in terms of a non-negative multiplier $\dot{\lambda}$. ■

4. FINITE DIMENSIONAL RELATIONS OF LINEAR ELASTICITY

The finite element method permits the traditional continuous tensor fields of mechanics to be replaced by finite dimensional vectors. This representation of the mechanical system generally reflects all properties and features of the physical problem equally well as an infinite dimensional representation. The finite dimensional description is less detailed, but not less general, than the infinite dimensional one. It furthermore makes applicable the mathematical tools of linear algebra. These tools are more manageable than those of functional analysis, required in an infinite dimensional description.

The finite element matrix equations, that describe elasto-plastic continuum problems, have been extensively used for analysis and computation[20–22]. Unilateral boundary value problems have also been treated in this way[14]. In the present work, a slightly more general approach is used; the theory of finite dimensional vector spaces, as described by Besseling[26, 27], is utilized. The advantages of this are that it emphasizes the structure of the theory and that it suggests an abstract formulation of the problem.

Note that for a contact problem of linear elasticity, all non-linearity occurs on the contact boundary. Therefore, it is useful to formulate the problem in terms of elements of contact boundary spaces only. Such a formulation is made possible by equations obtained in this section. These equations relate contact forces, contact surface displacements and the prescribed forces and displacements.

Consider a structure divided into finite elements. A displacement state u of the structure can then be thought of an an element of a finite dimensional vector space $\mathscr{U}$. The forces acting on the structure, could by means of the interpolation functions, be represented by their virtual work as functionals $\mathscr{L}(\mathscr{U}, R)$ on $\mathscr{U}$. These functionals constitute the dual space of $\mathscr{U}$, which we will denote by $\mathscr{F}$. Forces such that the structure is in equilibrium constitute a subspace of $\mathscr{F}$, which is isomorphic to the space of stress states Σ. There exists a one-to-one linear transformation

$$D^{T}: \Sigma \rightarrow \mathscr{F}. \tag{44}$$

It can further be shown that the transpose of D^{T} constitutes a transformation

$$D: \mathscr{U} \rightarrow \mathscr{E} \tag{45}$$

where $\mathscr{E}$ is the space of strains which is isomorphic to a subspace of $\mathscr{U}$. We also assume the existence of a bijective transformation

$$E: \mathscr{E} \rightarrow \Sigma \tag{46}$$

that determines the constitutive behaviour of the structure. Thus, if $u \in \mathscr{U}$, $f \in \mathscr{F}$, $\sigma \in \Sigma$ and $\varepsilon \in \mathscr{E}$ we have the following set of canonical equations:

$$D^{T}\sigma = f \tag{47}$$

$$\varepsilon = Du \tag{48}$$

$$\sigma = E\varepsilon. \tag{49}$$

These equations express equilibrium, compatibility and constitutive behaviour of a linear elastic solid.

In order to introduce "boundary conditions" we assume the displacement space $\mathscr{U}$ to be decomposed into the two disjoint subspaces $\mathscr{U}^{0}$ and a direct complement $\mathscr{U}^{c}$. The dual spaces of these spaces constitute a decomposition of $\mathscr{F}$ and they are denoted by $\mathscr{F}^{0}$ and $\mathscr{F}^{c}$. Equations (47) and (48) can then be written

$$D^{0T}\sigma = f^{0} \tag{50}$$

$$D^{cT}\sigma = f^{c} \tag{51}$$

$$\varepsilon = D^{0}u^{0} + D^{c}u^{c} \tag{52}$$

where $u^0 \in \mathscr{U}^0$, $u^c \in \mathscr{U}^c$, $f^0 \in \mathscr{F}^0$ and $f^c \in \mathscr{F}^c$; the transformations D^0 and D^c are restrictions of D to $\mathscr{U}^0$ and $\mathscr{U}^c$, respectively.

Here $\mathscr{U}^0$ will be regarded as a space of prescribed displacements. That is, if $u^0 \in \mathscr{U}^0$, then $u^0 = \bar{u}(t)$, where $\bar{u}(t)$ is a known function of time t. Equations (49) and (50)–(52) result in

$$D^{cT}\sigma = D^{cT}E\varepsilon = D^{cT}ED^c u^c + D^{cT}ED^0\bar{u}(t) = f^c. \tag{53}$$

Now introduce, as a subspace of $\mathscr{U}^c$, the space $\mathscr{W}$ of contact boundary displacements. There exists a projection operator $C_1 : \mathscr{U}^c \to \mathscr{W}$; i.e.

$$w = C_1 u^c \tag{54}$$

where $w \in \mathscr{W}$. It is known that the range space $\mathscr{R}(C_1) = \mathscr{W}$ and the null space $\mathscr{N}(C_1) = \mathscr{U}_2$ form a direct sum of $\mathscr{U}^c$. Furthermore, the prescribed forces $\bar{f}(t)$ are regarded as a functional on $\mathscr{U}^c$; thus, $\bar{f}(t) \in \mathscr{F}^c$. The dual spaces of $\mathscr{W}$ and $\mathscr{U}_2$ are $\mathscr{P}$ and $\mathscr{F}_2$. The contact forces P belong to $\mathscr{P}$ and the restrictions of $\bar{f}(t)$ to $\mathscr{P}$ and $\mathscr{F}_2$ is denoted by $\bar{f}_1(t)$ and $\bar{f}_2(t)$, respectively.

Denote by D_1 and D_2 the restrictions of D^c to $\mathscr{W}$ and $\mathscr{U}_2$, and introduce the notations $K_{11} = D_1^T E D_1$, $K_{12} = D_1^T E D_2$, $K_{22} = D_2^T E D_2$, $f_1^u(t) = -D_1^T E D^0 \bar{u}(t)$ and $f_2^u(t) = -D_2^T E D^0 \bar{u}(t)$. From eqn (53) we then obtain

$$K_{11}w + K_{12}u_2 = P + \bar{f}_1(t) + f_1^u(t) \tag{55}$$

$$K_{12}^T w + K_{22}u_2 = \bar{f}_2(t) + f_2^u(t). \tag{56}$$

If D_2 is one-to-one, then the inverse K_{22}^{-1} exists, and u_2 can be eliminated from eqns (55) and (56). Thus

$$(K_{11} - K_{12}K_{22}^{-1}K_{12}^T)w = P + \bar{f}_2(t) + f_1^u(t) - K_{12}K_{22}^{-1}(\bar{f}_2(t) + f_2^u(t)). \tag{57}$$

This equation states a relation between $P \in \mathscr{P}$, $w \in \mathscr{W}$ and the prescribed forces and displacements.

The operator $(K_{11} - K_{12}K_{22}^{-1}K_{12}^T)$ in eqn (57) may be non-singular and therefore invertible. This possibility, which makes an inverse relation of eqn (57) obtainable, will be investigated in a slightly different setting from what has been used up to now. That is, the space $\mathscr{U}_2$ is assumed to be a subspace of $\mathscr{U}^c$, in contrast to above, not necessarily a direct complement of $\mathscr{W}$. Then a projection operator C_2: $\mathscr{U}^c \to \mathscr{U}_2$ exists; i.e.

$$u_2 = C_2 u^c. \tag{58}$$

Letting the prescribed force $\bar{f}(t)$ belong to $\mathscr{F}_2$, by using a work equivalence a functional $f^c \in \mathscr{F}^c$ on $\mathscr{U}^c$ can be constructed

$$\langle f^c, u^c \rangle_{\mathscr{U}^c} = \langle P, w \rangle_{\mathscr{W}} + \langle \bar{f}(t), u_2 \rangle_{\mathscr{U}_2} \qquad \forall u^c \in \mathscr{U}^c \tag{59}$$

where $\langle \cdot, \cdot \rangle_X$ represents dual pairing between a space X and its dual space. By using eqns (54) and (58) we obtain from eqn (59)

$$f^c = C_1^T P + C_2^T \bar{f}(t). \tag{60}$$

Assuming that D^c is one-to-one, the inverse of $D^{cT}ED^c$ exists and we obtain from eqns (53) and (60)

$$u^c = K_{cc}^{-1} C_1^T P + K_{cc}^{-1} C_2^T \bar{f}(t) + K_{cc}^{-1} f_c^u(t) \tag{61}$$

where the notations $K_{cc} = D^{cT} E D^c$ and $f_c^u(t) = -D^{cT} E D^0 \bar{u}(t)$ have been used. Premultiplying eqn (61) with C_1 and using eqn (54) we get

$$w = C_1 K_{cc}^{-1} C_1^T P + C_1 K_{cc}^{-1} C_2^T \bar{f}(t) + C_1 K_{cc}^{-1} f_c^u(t). \tag{62}$$

This is the inverse relation of (57) and it is obtainable when D^c is one-to-one, i.e. when the structure is not a mechanism in its non-contacting state. Also, note that C_1 is surjective and, therefore, the inverse of $C_1 K_{cc}^{-1} C_1^T$ exists.

Finally, it should be pointed out that a relation, having the same form as eqn (57), can be obtained under the mild restriction that there exists a state of equilibrium for the structure; this is shown by using a force method.

5. VARIATIONAL FORMULATIONS

In this section we will combine the contact boundary conditions of Section 3 with the relations obtained in Section 4, in order to obtain some different variational formulations of contact problems.

Relations (62) and (57) can be written as

$$w = \beta P - V \tag{63}$$

$$P = \kappa w - R \tag{64}$$

where $\kappa(\beta)$ is a "stiffness" ("flexibility") transformation and V and R are "force" vectors. Depending on the contact boundary conditions under consideration P and w will be indexed by both N and T in the sequel.

The contact boundary conditions on Section 3 were assumed to hold pointwisely on the contact surface. Therefore, in order to combine them with eqn (63) or eqn (64) they should be extended to relations on the contact boundary spaces $\mathscr{W}$ and $\mathscr{P}$. Due to the finite dimensional setting used in this paper, the contact boundary conditions can be extended by means of finite sums. It is not immediately clear that this can be done unambiguously; the mathematical question marks are, however, removed by Andersson[28].

5.1. *Static laws*

When considering boundary conditions in the form of static laws the space of contact boundary displacements $\mathscr{W}$ is composed of ν disjoint subspaces $\overset{i}{X}_r$, $i = 1, \ldots, \nu$; i.e. the elements of $\mathscr{W}$ and its dual space are defined by adding the elements of the subspaces, so that, if $\overset{i}{w}_N \in \overset{i}{X}_r$, then $w_N = \sum_{i=1}^{\nu} \overset{i}{w}_N \in \mathscr{W}$, and if $\overset{i}{P} \in \overset{i}{X}'_r$, then $P = \sum_{i=1}^{\nu} \overset{i}{P} \in \mathscr{P}$.

Convex or non-convex superpotentials $\overset{i}{J}_N$, as described in Section 3.1 are defined on $\overset{i}{X}'_r$. If these local superpotentials $\overset{i}{J}_N$ are assumed to be proper we can, by addition, construct a global superpotential J_N, defined on $\mathscr{P}$; i.e.

$$J_N(-P_N) = \sum_{i=1}^{\nu} \overset{i}{J}_N(-\overset{i}{P}_N). \tag{65}$$

The set of statically admissible contact force states is introduced

$$\chi_N = \{P_N \in \mathscr{P} : J_N(-P_N) < \infty\}.$$

According to Andersson[28], if $P_N \in \chi_N$ and $\overset{i}{J}_N$ are lsc and proper, then

$$w_N \in \partial J_N(-P_N) \tag{66}$$

if and only if

$$\overset{i}{w}_N \in \partial \overset{i}{J}_N(-\overset{i}{P}_N), \qquad i = 1,\ldots,\nu.$$

From relation (5) it follows that, for $P_N \in \chi_N$, relation (66) is equivalent to

$$J_N^0(-P_N, -P_N' + P_N) \geq \langle w_N, -P_N' + P_N \rangle_{\mathscr{P}} \qquad \forall\, P_N' \in \mathscr{P}. \tag{67}$$

Indexing P and w, in eqn (63), by N, and inserting into relation (67) we obtain

$$J_N^0(-P_N, -P_N' + P_N) + A(P_N\, P_N' - P_N) \geq \langle V, P_N' - P_N \rangle_{\mathscr{P}} \qquad \forall\, P_N' \in \mathscr{P} \tag{68}$$

where the bilinear form $A(P_N, P_N') = \langle \beta P_N, P_N' \rangle_{\mathscr{P}}$ has been introduced.

We can now formulate the following problem:

(1) Find $P_N \in \chi_N$ such that relation (68) is satisfied.

This problem constitutes a variational formulation of the problem of contact between linear elastic bodies, when contact boundary conditions are represented by static laws. Relations such as (68) have been called hemi-variational inequalities[16].

Note that the requirement $P_N' \in \mathscr{P}$ in relation (68) could be substituted for $P_N' \in \chi_N$ since directions $-P_N' + P_N$ that are unique to the first case correspond to $J_N^0(-P_N, -P_N' + P_N) = \infty$.

If J_N is a convex superpotential, $J_N(-P_N') - J_N(-P_N)$ replaces $J_N^0(-P_N, -P_N' + P_N)$ and relation (67) becomes a variational inequality. When J_N represents a perfect unilateral constraint, i.e. the static law of Example 1 in Section 3 with $Q_N \to \infty$, the resulting variational inequality is a finite dimensional case of what has been called the reciprocal formulation of the Signorini problem[29].

Furthermore, if J_N is convex, lsc and proper, there exists a conjugated function J_N^c of J_N with the property that

$$-P_N \in \partial J_N^c(w_N). \tag{69}$$

Introduce the set of kinematically admissible contact surface displacements

$$\mathscr{Y}_N = \{ w_N \in \mathscr{W} : J_N^c(w_N) < \infty \}.$$

The dual problem of problem (1) can now be formulated by means of relation (69). Using relation (5), in the same way as above, we obtain from eqn (64) and relation (69)

$$J_N^{co}(w_N, w_N' - w_N) + \bar{A}(w_N, w_N' - w_N) \geq \langle R, w_N' - w_N \rangle_{\mathscr{W}} \qquad \forall\, w_N' \in \mathscr{W} \tag{70}$$

where $\bar{A}(w_N, w_N') = \langle \kappa w_N, w_N' \rangle_{\mathscr{W}}$.

The dual problem of problem (1) is:

(2) Find $w_N \in \mathscr{Y}_N$ such that relation (70) is satisfied.

If, regardless of the existence of J_N, there exists a superpotential $\bar{J}_N$ such that $-P_N \in \partial \bar{J}_N(w_N)$, then it is possible to obtain an inequality formally identical to relation (70) but with J_N^c replaced by $\bar{J}_N$. Finally, note also that problems (1) and (2) can be shown to follow from so-called substationarity principles[16]. These principles are dual minimization problems in the convex case.

5.2. Elasto-resistance laws

Let us now turn to boundary conditions in the form of elasto-resistance laws. Let the space of contact boundary displacements $\mathscr{W}$ be divided into the disjoint subspaces $\mathscr{W}_N$ and $\mathscr{W}_T$ of normal and tangential contact displacements, respectively. Assume that $\mathscr{W}_N$ is composed of disjoint spaces $\overset{i}{X}_r$, $i = 1,\ldots,\mu$ and similarly, $\mathscr{W}_T$ of disjoint spaces $\overset{i}{x}_s$, $i = 1,\ldots,\mu$; thus, if $\overset{i}{w}_N \in \overset{i}{X}_r$, then $w_N = \sum_{i=1}^{\mu} \overset{i}{w}_N \in \mathscr{W}_N$, and if $\overset{i}{w}_T \in \overset{i}{X}_s$, then $w_T = \sum_{i=1}^{\mu} \overset{i}{w}_T \in \mathscr{W}_T$. The dual space $\mathscr{P}$ of contact forces is decomposed in a similar way as $\mathscr{W}$. Further, introduce the space of internal parameters $\mathscr{A}$. It is assumed to consist of μ disjoint spaces $\overset{i}{X}_\iota$, $i = 1,\ldots,\mu$, such that if $\overset{i}{\alpha} \in \overset{i}{X}_\iota$, then $\alpha = \sum_{i=1}^{\mu} \overset{i}{\alpha} \in \mathscr{A}$. The dual space $\mathscr{A}'$ of $\mathscr{A}$ is the space of internal forces and it is decomposed similarly to $\mathscr{A}$.

Now assume superpotentials $\overset{i}{J}_T$, as described in Section 3.2, to be defined on the subspaces $\overset{i}{X}'_r \times \overset{i}{X}'_s \times \overset{i}{X}'_\iota$ of $\mathscr{P} \times \mathscr{A}'$. From these local superpotentials we obtain a global one by addition

$$J_T(-P_T, A; P_N) = \sum_{i=1}^{\mu} \overset{i}{J}_T(-\overset{i}{P}_T, \overset{i}{A}; \overset{i}{P}_N) \tag{71}$$

where $P_T \in \mathscr{P}_T$, $P_N \in \mathscr{P}_N$ and $A \in \mathscr{A}'$. Introduce the set of statically admissible contact and internal force states

$$\chi_T = \{(P, A) \in \mathscr{P} \times \mathscr{A}' : J_T(-P_T, A; P_N) < \infty\}.$$

To deduce a global description of the elasto-resistance law, we define reversible and non-reversible contact boundary displacements, w_T^A, $w_T^S \in \mathscr{W}_T$, as $w_T^A = \sum_{i=1}^{\mu} \overset{i}{w}{}_T^A$, and as $w_T^S = \sum_{i=1}^{\mu} \overset{i}{w}{}_T^S$, so that

$$w_T = w_T^A + w_T^S. \tag{72}$$

Similarly to the case of static laws, according to Ref. [8] we have: if $(P, A) \in \chi_T$ and $\overset{i}{J}_T$ are lsc and proper on $\overset{i}{X}'_s \times \overset{i}{X}_\iota$ for a prescribed $P_N = \sum_{i=1}^{\mu} \overset{i}{P}_N$, then

$$(\dot{w}_T^S, -\dot{\alpha}) \in \partial J_T(-P, A; P_N) \tag{73}$$

is equivalent to similar local subdifferential relations, defined for each subspace $\overset{i}{X}'_r \times \overset{i}{X}'_s \times \overset{i}{X}'_\iota$.

Global forms of eqns (26) and (27) are also needed. Thus, from the local contact stiffness transformations $\overset{i}{Q}_T$ we obtain a global one, Q_T, as

$$P_T = \sum_{i=1}^{\mu} \overset{i}{P}_T = \sum_{i=1}^{\mu} -\overset{i}{Q}_T \overset{i}{w}{}_T^A = -Q_T w_T^A. \tag{74}$$

Similarly, a global form of eqn (27) is defined by

$$A = \sum_{i=1}^{\mu} \overset{i}{A} = \sum_{i=1}^{\mu} \overset{i}{H} \overset{i}{\dot{\alpha}} = H\alpha. \tag{75}$$

Relations (72)–(75) describe a global elasto-resistance law. To combine this law with eqn (63) we note that relation (73) is, according to relation (5), equivalent to

$$J_{\mathrm{T}}^{0}((-P_{\mathrm{T}}, A), (-P'_{\mathrm{T}}, A') - (-P_{\mathrm{T}}, A); P_{\mathrm{N}}) \geqslant \langle \dot{w}_{\mathrm{T}}^{\mathrm{S}}, -P'_{\mathrm{T}} + P_{\mathrm{T}} \rangle_{\mathscr{P}_{\mathrm{T}}} + \langle -\dot{\alpha}, A' - A \rangle_{\mathscr{A}'} \quad \forall (\mathscr{P}'_{\mathscr{T}}, \mathscr{A}') \in \mathscr{P}_{\mathrm{T}} \times \mathscr{A}'. \tag{76}$$

Furthermore, eqn (63) can be time differentiated and written in a weak form

$$\langle \dot{w}_{\mathrm{N}}, P'_{\mathrm{N}} - P_{\mathrm{N}} \rangle_{\mathscr{P}_{\mathrm{N}}} + \langle \dot{w}_{\mathrm{T}}, P'_{\mathrm{T}} - P_{\mathrm{T}} \rangle_{\mathscr{P}_{\mathrm{T}}} = \langle \beta \dot{P}, P' - P \rangle_{\mathscr{P}} - \langle \dot{V}, P' - P \rangle_{\mathscr{P}} \quad \forall P' \in \mathscr{P}. \tag{77}$$

Combining relations (72) through (77) we obtain

$$J_{\mathrm{T}}^{0}((-P_{\mathrm{T}}, A), (-P'_{\mathrm{T}}, A') - (-P_{\mathrm{T}}, A); P_{\mathrm{N}}) + B(\dot{P}, P' - P) + C(\dot{A}, A' - A) \geqslant \langle \dot{w}_{\mathrm{N}}, P'_{\mathrm{N}} - P_{\mathrm{N}} \rangle_{\mathscr{P}_{\mathrm{N}}} + \langle \dot{V}, P' - P \rangle_{\mathscr{P}} \quad \forall (P', A') \in \mathscr{P} \times \mathscr{A}' \tag{78}$$

where we have used the following definitions of bilinear forms

$$B(P, P') = \langle \beta P + Q_{\mathrm{T}}^{-1} P_{\mathrm{T}}, P' \rangle_{\mathscr{P}} \quad \forall P, P' \in \mathscr{P}$$

$$C(A, A') = \langle H^{-1} A, A' \rangle_{\mathscr{A}'} \quad \forall A, A' \in \mathscr{A}'.$$

Inequality (78) reveals the fact that eqn (63) and the elasto-resistance law do not immediately make up a proper problem formulation: the rate of normal displacement $\dot{w}_{\mathrm{N}}$ must be eliminated from inequality (78). To that end, in some physical problems it is plausible to assume that $\dot{w}_{\mathrm{N}}$ is prescribed for each time $t \in [0, T]$ under consideration; in other cases a linear or non-linear mapping from $\mathscr{P}$ to $\mathscr{W}_{\mathrm{N}}$ can be assumed such that $\dot{w}_{\mathrm{N}}$ can be eliminated from inequality (78). Assuming one or the other of these two possibilities we have that, if V: $[0, T] \to \mathscr{W}$ is a known mapping and if proper initial conditions are prescribed for $P \in \mathscr{P}$ and $A \in \mathscr{A}'$ at $t = 0$, then the following problem can be formulated:

(3) Find (P, A): $[0, T] \to \chi_{\mathrm{T}}$ such that inequality (78) is satisfied for each time $t \in [0, T]$.

This is a variational formulation of the problem of contact between linear elastic bodies, when contact boundary conditions in the form of elasto-resistance laws are valid.

Inequality (78) resembles a quasi-variational inequality, since P_{N} is present in J_{T}^{0}. However, since J_{T} can be a non-convex superpotential, a proper name for inequality (78) seems to be a quasi-hemi-variational inequality.

No problem exists which could be regarded as dual to problem (3), at least not in the usual sense. However, under certain additional assumptions, now to be given, a formulation of the present problem in terms of contact boundary displacements can be presented. If J_{T} is convex, when restricted to $\mathscr{P}_{\mathrm{T}} \times \mathscr{A}'$, a conjugated function $J_{\mathrm{T}}^{\mathrm{c}}$ exists with the property that

$$(-P_{\mathrm{T}}, A) \in \partial J_{\mathrm{T}}^{\mathrm{c}}(\dot{w}_{\mathrm{T}}^{\mathrm{S}}, -\dot{\alpha}; P_{\mathrm{N}}). \tag{79}$$

Furthermore, require that $\dot{w}_{\mathrm{T}}^{\mathrm{S}} = \dot{w}_{\mathrm{T}}$, i.e. $Q_{\mathrm{T}} \to \infty$, and that there exists a linear or non-linear transformation Z: $\mathscr{W} \to \mathscr{P}_{\mathrm{N}}$. A superpotential $J_{\mathrm{T}}^{\mathrm{w}}$ can then be defined as

$$J_T^w(\dot{w}_T, -\dot{\alpha}; w) = J_T^c(\dot{w}_T, -\dot{\alpha}; Zw). \tag{80}$$

Obviously relation (79) holds with J_T^c replaced by J_T^w.

Introduce the set of kinematically admissible contact displacement and internal parameter rates

$$\mathscr{Y}_T(w) = \{(\dot{w}, \dot{\alpha}) \in \mathscr{W} \times \mathscr{A} : J_T^w(\dot{w}_T, -\dot{\alpha}; w) < \infty\}.$$

Moreover, the following weak form of eqn (64) is useful:

$$\langle P_T, w'_T - \dot{w}_T \rangle_{\mathscr{W}_T} + \langle P_N, w'_N - \dot{w}_N \rangle_{\mathscr{W}_N} = \langle \kappa w, w' - \dot{w} \rangle_{\mathscr{W}} - \langle R, w' - \dot{w} \rangle_{\mathscr{W}} \qquad \forall w' \in \mathscr{W}. \tag{81}$$

When eqn (80) is valid it follows from relation (5) that relation (79) is equivalent to

$$J_T^{w^0}((\dot{w}_T, -\dot{\alpha}), (w'_T, -\alpha') - (\dot{w}_T, -\dot{\alpha}); w) \geqslant \langle -P_T, w'_T - \dot{w}_T \rangle_{\mathscr{W}_T} + \langle A, -\alpha' + \dot{\alpha} \rangle_{\mathscr{A}} \qquad \forall (w'_T, \alpha') \in \mathscr{W}_T \times \mathscr{A}. \tag{82}$$

Combining relations (81) and (82) we obtain

$$J_T^{w^0}((\dot{w}_T, -\dot{\alpha}), (w'_T, -\alpha') - (\dot{w}_T, -\dot{\alpha}); w) + \bar{B}(w, w' - \dot{w}) + \bar{C}(\alpha, \alpha' - \dot{\alpha}) \geqslant \langle P_N, w'_N - \dot{w}_N \rangle_{\mathscr{W}_N} + \langle R, w' - \dot{w} \rangle_{\mathscr{W}} \qquad \forall (w', \alpha') \in \mathscr{W} \times \mathscr{A} \tag{83}$$

where the following bilinear forms have been introduced:

$$\bar{B}(w, w') = \langle \kappa w, w' \rangle_{\mathscr{W}} \qquad \forall w, w' \in \mathscr{W}$$

$$\bar{C}(\alpha, \alpha') = \langle H\alpha, \alpha' \rangle_{\mathscr{A}} \qquad \forall \alpha, \alpha' \in \mathscr{A}.$$

Similarly to the case of problem (3), P_N has to be eliminated from relation (83) in order for this inequality to make sense as a formulation of our problem. However, in this case, the transformation Z, which is already assumed to exist, can be used to eliminate P_N from relation (83). Therefore, assuming the transformation $R: [0, T] \to \mathscr{P}$ to be known together with proper initial conditions for $w \in \mathscr{W}$ and $\alpha \in \mathscr{A}$, the following problem can be formulated.

(4) Find (w, α): $[0, T] \to \mathscr{W} \times \mathscr{A}$ such that, for each time $t \in [0, T]$

(i) $(\dot{w}, \dot{\alpha}) \in \mathscr{Y}_T(w)$

(ii) relation (83) is satisfied.

Convex or non-convex superpotentials such as J_T^w can of course exist independently of the existence of J_T, i.e. relation (80) does not need to hold. In that case, relation (83) can be made to make sense as a formulation of the problem by assuming P_N: $[0, T] \to \mathscr{P}$ to be a known mapping.

Oden and Martins[25] have recently formulated a dynamic analogy of problem (4).

5.3. *Incremental laws*

This section will be concluded by a treatment of incremental laws. Consider the same decomposition of $\mathscr{W}$ as above and introduce global superpotentials $\hat{J}_N(\dot{P}_N)$ and $\hat{J}_T(\dot{P}_T; \dot{P}_N)$ in the following way:

$$\hat{J}_N(-\dot{P}_N) = \sum_{i=1}^{\mu} \overset{i}{\hat{J}}_N(-\overset{i}{\dot{P}}_N) \tag{84}$$

$$\hat{J}_T(-\dot{P}_T; \dot{P}_N) = \sum_{i=1}^{\mu} \overset{i}{\hat{J}}_T(-\overset{i}{\dot{P}}_T; \overset{i}{\dot{P}}_N) \tag{85}$$

where $\overset{i}{\hat{J}}_N$ and $\overset{i}{\hat{J}}_T$ are local superpotentials, as described in Section 3.3, defined on $\overset{i}{X}_r$ and $\overset{i}{X}_r \times \overset{i}{X}_s$, respectively.

The set

$$\hat{\chi} = \{\dot{P} \in \mathscr{P}: \hat{J}_N(\dot{P}_N) < \infty, \hat{J}_T(\dot{P}_T; \dot{P}_N) < \infty\}$$

of statically admissible contact force rate states is introduced. If $\dot{P} \in \hat{\chi}$ and $\overset{i}{\hat{J}}_N$ are lsc and proper, and $\overset{i}{\hat{J}}_T$ are lsc and proper on $\overset{i}{X}_s$ for a prescribed $\dot{P}_N = \sum_{i=1}^{\mu} \overset{i}{\dot{P}}_N$, then

$$\dot{w}_N \in \partial \hat{J}_N(-\dot{P}_N) \tag{86}$$

and

$$\dot{w}_T \in \partial \hat{J}_T(-\dot{P}_T; \dot{P}_N) \tag{87}$$

are equivalent to similar local subdifferential relations, defined on the subspaces $\overset{i}{X}'_r$ and $\overset{i}{X}'_r \times \overset{i}{X}'_s$.

Note that the weak form of eqn (63), described by (77), is still valid when, instead of the variations $P' - P$, the variations $P' - \dot{P}$ are taken. Combining such a relation with relations (86) and (87) and taking relation (5) into account we obtain

$$\hat{J}^0_N(-\dot{P}_N, -P'_N + \dot{P}_N) + \hat{J}^0_T(-\dot{P}_T, -\dot{P}_T + \dot{P}_T; \dot{P}_N) + \hat{B}(\dot{P}, P' - \dot{P}) \geqslant \langle \dot{V}, P' - \dot{P} \rangle_{\mathscr{P}} \qquad \forall P' \in \mathscr{P} \tag{88}$$

where $\hat{B}(P, P') = \langle \beta P, P' \rangle_{\mathscr{P}}$.

Relation (88) is a quasi-hemi-variational inequality and the following problem can be formulated:

(5) Find $\dot{P} \in \hat{\chi}$ such that relation (88) is satisfied.

As was the case when elasto-resistance laws were studied, no problem exists which could be regarded as dual to problem (5). To achieve a formulation in terms of displacement variables, one should look for an incremental superpotential $\bar{J}_T(\dot{w}_T; \dot{w}_N)$, such that $-\dot{P}_T \in \partial \bar{J}_T(\dot{w}_T; \dot{w}_N)$.

6. SOLUTION METHODS AND LCP FORMULATIONS

Due to the novelty of the concept of hemi-variational inequalities, only heuristic approaches to their numerical solution are available. In Ref. [19] it is suggested that one should introduce a "smoothing" of the non-differentiable superpotentials. This transforms the hemi-variational inequality into a variational equality and a non-linear algebraic equation solver can be used for solution. In cases of the evolutionary problems (3) and (4), this idea has to be combined with a time discretization, and an iterative procedure to determine the implicit variable P_N; i.e. the ideas used to solve contact problems with convex superpotentials in Ref. [9] could be generalized.

In this paper, the above method will not be outlined; instead alternative formulations of the present problems will be introduced, which lends themselves naturally to numerical treatment. These formulations are recognizable as what is known in mathematical programming theory as linear complementarity problems (LCP), and they are possible to obtain if the subdifferential relations are piecewise linear in such a way that they can be expressed using non-negative multipliers. This was the case in Examples 1, 3 and 5 of Section 3, and further examples are shown in Fig. 4. Coulomb's friction law can be described by non-negative multipliers if the surface $\{P_N \in R, P_T \in R^2, \phi = |P_T| - \mu|P_N| = 0, P_N < 0\}$ is approximated according to Fig. 5. The LCP obtained using this approximation is investigated in Ref. [23].

The LCP concept has been widely used to model plastic material laws and to investigate the load response of elasto-plastic structures; indeed, a rather extensive analogy exists between these problems and the present ones[23]. In this paper we will consider two examples of LCP formulations, related to frictional systems.

Example 1

Assume that the spaces $\mathscr{P}$ and $\mathscr{W}$ both equal the Euclidean space R^v. That is, the contact boundary consists of v discrete nodal points. Relations such as (22) are assumed to be valid at each contact nodal point. To describe this for the entire boundary simultaneously, we introduce a quadratic v by v matrix $Q_N = \operatorname{diag}[\overset{1}{Q}_N, \ldots, \overset{v}{Q}_N]$ and vectors w_N, P_N, g and λ belonging to R^v. We then have

$$w_N = g - Q_N^{-1} P_N - \lambda \tag{89}$$

$$\lambda \geqslant 0, \qquad P_N \leqslant 0, \qquad \langle P_N, \lambda \rangle_{R^v} = 0. \tag{90}$$

Relation (63) is, in this case, represented by a v by v matrix β and vectors w_N, P_N and V. Combining eqn (63) and relations (89) we obtain

$$\lambda = V + g - (\beta + Q_N^{-1}) P_N. \tag{91}$$

Relations (90) together with eqn (91) are a LCP and, therefore, several direct and iterative methods have been suggested for its solution[30]. If $\beta + Q_N^{-1}$ is symmetric and positive definite, relations (90) and (91) represent the Kuhn–Tucker conditions of a quadratic programming problem and any method applicable to such problems can therefore be used to solve the present one. ■

Example 2

Consider the elasto-resistance law in Example 3 of Section 3. Assume that ϕ_i are affine functions on the space $R^r \times R^s \times R^t$. This means that $\nabla\phi_i$, $\bar{\nabla}\phi_i$ and $\bar{\bar{\nabla}}\phi_i$ are independent of $-P_T$, A and P_N. We also assume that spaces $\mathscr{P}_N$ and $\mathscr{P}_T$ are the Euclidean spaces $R^{r\mu}$ and $R^{s\mu}$, respectively. The contact boundary consists of μ nodal points and relations such as eqns (35)–(37) are prescribed for each one. To describe this, we introduce the gradient matrices

$$G_1 = \operatorname{diag}\{[\nabla\phi_1^1, \ldots, \nabla\phi_q^1], [\nabla\phi_1^2, \ldots,], \ldots, [\nabla\phi_1^\mu, \ldots]\}$$
$$G_2 = \operatorname{diag}\{[\bar{\nabla}\phi_1^1, \ldots, \bar{\nabla}\phi_q^1], [\bar{\nabla}\phi_1^2, \ldots,], \ldots, [\bar{\nabla}\phi_1^\mu, \ldots]\}$$
$$G_3 = \operatorname{diag}\{[\bar{\bar{\nabla}}\phi_1^1, \ldots, \bar{\bar{\nabla}}\phi_q^1], [\bar{\bar{\nabla}}\phi_1^2, \ldots,], \ldots, [\bar{\bar{\nabla}}\phi_1^\mu, \ldots]\}.$$

A hardening matrix is also needed

$$H = \mathrm{diag}[\overset{i}{H}, \ldots, \overset{\mu}{H}].$$

The gradient matrices and the hardening matrix now build two (r + s)μ by $q\mu$ matrices

$$\overset{0}{N} = \begin{bmatrix} G_1 \\ 0 \end{bmatrix}, \qquad N = \begin{bmatrix} -G_1 \\ G_3 \end{bmatrix}$$

and an $q\mu$ by $q\mu$ matrix

$$\bar{H} = G_2^{\mathrm{T}} H G_2.$$

If $\dot{P} = [\dot{P}_{\mathrm{T}}^{\mathrm{T}}, \dot{P}_{\mathrm{N}}^{\mathrm{T}}]^{\mathrm{T}} \in R^{(r+s)\mu}$ and $\dot{\phi}, \dot{\lambda} \in R^{q\mu}$, relations (36) and (37) can be extended to a relation on the entire contact boundary

$$\dot{\phi} = N^{\mathrm{T}}\dot{P} - \bar{H}\dot{\lambda}. \tag{92}$$

Furthermore, a flexibility transformation is introduced as an (r + s)μ by (r + s)μ matrix

$$S = \mathrm{diag}[\overset{1}{Q}{}_{\mathrm{T}}^{-1}, \ldots, \overset{\mu}{Q}{}_{\mathrm{T}}^{-1}, 0, \ldots, 0]$$

such that, if $\dot{w} = [\dot{w}_{\mathrm{T}}^{\mathrm{T}}, \dot{w}_{\mathrm{N}}^{\mathrm{T}}]^{\mathrm{T}} \in R^{(r+s)\mu}$ and if $\dot{w}_{\mathrm{N}}$ is assumed to equal zero, then eqn (35a) can be described for the entire boundary as

$$\dot{w} = -S\dot{P} + \overset{0}{N}\dot{\lambda}. \tag{93}$$

Since ϕ_i were assumed to be affine functions eqn (92) can be integrated to yield

$$\phi = N^{\mathrm{T}}P - \bar{H}\lambda + K \tag{94}$$

where $K \in R^{q\mu}$ represents the initial (i.e. $\lambda = 0$) distance from origo to the plane $\phi = 0$, in contact force space $R^{(r+s)\mu}$. Relations (35) can now be written for μ contact nodes

$$\phi \leqslant 0, \qquad \dot{\lambda} \geqslant 0, \qquad \langle \phi, \dot{\lambda} \rangle_{R^{q\mu}} = 0. \tag{95}$$

Combining a matrix form of eqn (63) with eqn (93), when integrated, yields, if the inverse of $(\beta + S)$ exists

$$P = (\beta + S)^{-1}V + (\beta + S)^{-1}\overset{0}{N}\lambda \tag{96}$$

which, when substituted into eqn (94), results in

$$\phi = K + N^{\mathrm{T}}(\beta + S)^{-1}V + (N^{\mathrm{T}}(\beta + S)^{-1}\overset{0}{N} - \bar{H})\lambda. \tag{97}$$

Assuming the map V: $[0, T] \rightarrow R^{(r+s)\mu}$ to be known together with proper initial conditions we now have the following problem:

> Find (ϕ, λ): $[0, T] \rightarrow R^{q\mu} \times R^{q\mu}$ such that eqns (97) and (95) are satisfied for each time $t \in [0, T]$.

This is a "linear complementarity problem involving derivatives". A similar problem to this arises in connection with the analysis of elasto-plastic structures, and with this application in mind, a solution algorithm has been suggested by Kaneko[31]. This algorithm is applied to the present frictional problem in Ref. [32]. ∎

It is known that contact problems with Coulomb friction need not have unique solutions[23, 33]. This fact can in our more general setting be explained to be due to the presence of P_N in the superpotential J_T. Also, in the case of static laws, non-uniqueness of solutions can occur if the superpotentials are non-convex. In the finite dimensional case, it has been shown that the theory of LCP provides a tool for investigation of these questions[23]. An example of this is provided by the following classical theorem[34], which is directly applicable to the LCP formulation in Example 1 of this section.

Theorem

The system $\omega = M\zeta + q \geqslant 0$, $\zeta \geqslant 0$, $\omega^T\zeta = 0$, where M is an n by n matrix and ω, q and ζ are n-vectors, has a unique solution (ω, ζ) for each q if and only if M is a P-matrix.†

7. CONCLUSIONS

This paper deals with the analysis of frictional systems, where the constitutive contact behaviour is allowed to be of a very general kind. The formulation of these contact boundary conditions is accomplished through the use of some recent concepts from convex and non-convex optimization. It is clear that these concepts—the subdifferential, the generalized gradient, and convex and non-convex superpotentials—are full of possibilities for a number of areas of mechanics; those based on convexity are already well established as tools of mechanical analysis[13, 14].

In order to obtain variational formulations of problems connected with frictional systems, we investigate finite dimensional relations of linear elasticity. It is pointed out that this representation is less detailed, but not less general, than the infinite dimensional one. The variational formulations obtained are of a type known as hemi-variational inequalities, which specialize to variational inequalities when the involved superpotentials are convex.

In contrast to a number of previous investigations[7, 10] we formulate the problems on the contact boundary spaces, i.e. the subspaces of the displacement and force spaces where the actual non-linearity of the problem occurs; the idea is similar to that used in Ref. [29]. This proceeding has the advantage of giving the problem a smaller dimension, and also of making the analogy with previous treatments of elasto-plastic problems clear. This analogy opens up a novel way of attacking contact problems: both numerically and theoretically. More specifically, it can be shown that several problems connected with frictional systems have the mathematical structure of a linear complementarity problem (LCP). LCPs are well known in the area of mathematical programming and results related to them have been extensively used to model, investigate and calculate elasto-plastic structural problems.

REFERENCES

1. Y. Seguchi, A. Shindo, Y. Tomita and M. Sunohara, Sliding rule of friction in plastic forming of metal. *Proc. Conf. on Comp. Meth. in Nonlinear Mech.*, University of Texas (1974).
2. B. Fredriksson, Finite element solution of surface nonlinearities in structural mechanics with special emphasis to contact and fracture mechanics problems. *Comp. Struct.* **6**, 281–290 (1976).
3. R. Michalowski and Z. Mróz, Associated and non-associated sliding rules in contact friction problems. *Arch. Mech. Stosow.* **30**(3), 259–276 (1978).
4. A. Curnier, A theory of friction. *Int. J. Solids Structures* **20**, 637–647 (1984).
5. G. Duvaut and J. L. Lions, *Inequalities in Mechanics and Physics*. Springer, Berlin (1976).
6. G. Duvaut, Problèmes mathématiques de la mécanique—Équilibre d'un solide élastique avec contact unilatéral et frottement de Coulomb. *C. R. Acad. Sci., Paris* **290A**, 263–265 (1980).

† A P-matrix is a matrix for which all principal minors are positive[34].

7. 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).
8. L. Demkowicz and J. T. Oden, On some existence and uniqueness results in contact problems with nonlocal friction. *Nonlinear Analysis, Theory, Methods Appl.* **6**, 1075–1093 (1982).
9. J. T. Oden and E. B. Pires, Numerical analysis of certain contact problems in elasticity with non-classical friction laws. *Comp. Struct.* **16**(1–4), 481–485 (1983).
10. P. D. Panagiotopoulos, A nonlinear programming approach to the unilateral contact-, and friction-boundary value problem in the theory of elasticity. *Ing.-Arch.* **44**, 421–432 (1975).
11. R. T. Rockafellar, *Convex Analysis.* Princeton University Press, Princeton, New Jersey (1970).
12. R. T. Rockafellar, *The Theory of Subgradients and its Applications to Problems of Optimization. Convex and Non-convex Functions.* Heldermann, Berlin (1981).
13. J. J. Moreau, On unilateral constraints, friction and plasticity. *New Variational Techniques in Mathematical Physics* (Edited by G. Capri and G. Stampacchia). Edizioni Cremonese, Rome (1974).
14. P. D. Panagiotopoulos, Convex analysis and unilateral static problems. *Ing.-Arch.* **45**, 55–68 (1976).
15. F. H. Clarke, Generalized gradients and applications. *Trans. Am. Math. Soc.* 205 (1975).
16. P. D. Panagiotopoulos, Nonconvex energy functions. Hemi-variational inequalities and substationary principles. *Acta Mech.* **48**, 111–130 (1983).
17. B. V. Courtney-Pratt and E. Eisner, The effect of a tangential force on the contact of metallic bodies. *Proc. R. Soc. London* **A238**, 529–550 (1957).
18. B. Halphen, Nguyen Quoc Son, Sur les matériaux standards généralisés. *J. Méc.* **14**, 40–63 (1975).
19. P. D. Panagiotopoulos and C. C. Baniotopoulos, A hemivariational inequality and substationarity approach to the interface problem: theory and prospects of applications. *Engng Analysis* **1**, 20–31 (1984).
20. G. Maier, Mathematical programming methods in structural analysis. Int. Conf. on Variat. Meth. in Engng, Southampton University (1972).
21. G. Maier, Incremental plastic analysis in the presence of large displacements and physical instabilizing effects. *Int. J. Solids Structures* **7**, 345–372 (1971).
22. D. E. Grierson, A. Franchi, O. de Donato and L. Corradi, Mathematical programming and nonlinear finite element analysis. *Comp. Meth. Appl. Mech. Engng* **17/18**, 497–518 (1979).
23. A. Klarbring, Contact problems with friction—using a finite dimensional description and the theory of linear complementarity. Linköping Institute of Technology, LIU-TEK-LIC-1984:3 (1984).
24. B. Nayroles, Opérations et opérations régularisées sur les elements méchaniques et les lois d'effort sous-différentielles. *New Variational Techniques in Mathematical Physics* (Edited by G. Capri and G. Stampacchia). Edizioni Cremonese, Rome (1974).
25. J. T. Oden and J. A. C. Martins, Models and computational methods for dynamic friction phenomena. *Comp. Meth. Appl. Mech. Engng* **52**, 527–634 (1986).
26. J. F. Besseling, The force method and its applications in plasticity problems. *Comp. Struct.* **8**, 323–330 (1978).
27. J. F. Besseling, Finite element methods. *Trends in Solid Mechanics* (Edited by J. F. Besseling and Van der Heijden), pp. 53–78. Sijthoff and Noordhoff, Delft (1979).
28. L.-E. Andersson, Relations between local and global superpotentials defined on finite dimensional linear spaces. Linköping Institute of Technology, LiTH-MAT-R-1985-01 (1985).
29. N. Kikuchi, A class of Signorini's problems by reciprocal variational inequalities. *Computational Techniques for Interface Problems* (Edited by K. C. Park and D. K. Gartling), AMD 30, pp. 135–153 (1978).
30. R. W. Cottle, Numerical methods for complementarity problems in engineering and applied science, Department of Operations Research, Stanford University, CA 94305 (1977).
31. I. Kaneko, A parametric linear complementarity problem involving derivatives. *Mathl Prog.* **15**, 146–154 (1978).
32. A. Klarbring, The influence of slip hardening and interface compliance on contact stress distributions. A mathematical programming approach. *Mechanics of Material Interfaces* (Edited by A. P. S. Salvadurai and G. Voyiadjis), pp. 43–59. Elsevier, Amsterdam (1985).
33. V. Janovský, Catastrophic features of Coulomb friction model. *Proc. Mathematics of Finite Elements and Applications*, Brunel University (1981).
34. K. G. Murty, On the number of solutions of the complementarity problem and spanning properties of complementary cones. *Linear Algebra Applic.* **5**, 65–108 (1972).
35. S. J. Kim and J. T. Oden, Generalized potentials in finite elastoplasticity. *Int. J. Engng Sci.* **22**, 1235–1257 (1984).
36. S. J. Kim and J. T. Oden, Generalized flow potentials in finite elastoplasticity—II. Examples. *Int. J. Engng Sci.* **23**, 515–530 (1985).