A generalized Newton method for contact problems with friction

Alain Curnier, Pierre Alart

A generalized Newton method for contact problems with friction

A. CURNIER et P. ALART

École Polytechnique Fédérale de Lausanne, Département de Mécanique,
Laboratoire de Mécanique Appliquée, 1015 Lausanne, Suisse

RESUME

Cet article passe en revue les méthodes numériques utilisées depuis quelques années dans le programme TACT pour résoudre des problèmes de contact avec frottement non-associé de Coulomb. Ces méthodes comprennent : une méthode de pénalité pour imposer les conditions de contact et d'adhérence, une méthode de projection implicite pour intégrer la loi de glissement, la méthode des éléments finis pour effectuer la discrétisation spatiale et une méthode de Newton généralisée pour résoudre les nonlinéarités dues au contact et au frottement.

Des progrès récents améliorant la robustesse de l'algorithme global de contact avec frottement sont discutés. En particulier, une condition nécessaire et suffisante sur le coefficient de frottement garantissant l'unicité de la solution du contact plat est énoncée et un facteur d'amortissement garantissant la convergence de l'algorithme vers cette solution est introduit dans le cas bidimensionnel. Le problème du poinçon plat sert à illustrer à la fois la précision et l'efficacité de la méthode.

ABSTRACT

This article reviews the numerical methods used for a few years in the program TACT to solve contact problems with non-associated Coulomb's friction. These methods include : a penalty method to enforce the contact and adherence conditions respectively, an implicit projection method to integrate the slip rule, the finite element method for the spatial discretization and a generalized Newton method to overcome the contact and friction non linearities.

Recent advances improving the robustness of the resulting frictional contact algorithm are reported. In particular, a necessary and sufficient condition on the friction coefficient for the solution to flat contacts to be unique is stated and a damping factor is introduced to guarantee the algorithm convergence to this solution in the two-dimensional case. The flat punch problem is used to illustrate both the accuracy and efficiency of the method.

I. CONTINUUM AND CONTACT MECHANICS BACKGROUND

The type of problems adressed in this article falls in the class of quasi-static contact problems* between deformable solids with interface friction. The continuum and contact formulations of such problems are summarized below as an introduction.

I.1 Solid mechanics summary

Solid mechanics problems are conveniently formulated in the material Lagrangean description which can be summarized as follows [1]. The position after deformation, of a material particle identified by its location X in an undeformed configuration, is given by the placement $x = x(X)$. The displacement vector of the particle X is defined by $u(X) = x(X) - X$. The deformation of an infinitesimal fiber is captured by the (unsymmetric) deformation gradient defined through $dx = FdX$. The corresponding transformations for an oriented surface element and a volume element are $da = JF^{-T}dA$ and $dv = JdV$, where $J = \det F$. Excluding body forces for clarity, the "contact" force $dQ(X)$ exercised on a particle X by its neighbors, is regarded as the resultant of a nominal stress vector** $p(X)$ acting on an oriented material surface i.e. $dQ = p(X)dA$. The state of stress at a particle is recorded by the (unsymmetric) nominal stress tensor** $P(X)$ defined by the fundamental formula $p = P.N$ where $N(X)$ refers to the unit normal to the reference surface dA . Conservation of mass and moment equilibrium are built in the material description and the only principle of mechanics which remains to be satisfied is the equation of static equilibrium $\mathrm{Div}\, P = 0$, in V , where $\mathrm{Div}\, P = \mathrm{tr}(\partial P/\partial X)$ denotes the material divergence of P . This field equation must be completed by proper boundary conditions (BC) for the problem to be well-posed e.g. $u(X) = \bar{u}$ and $p(X) = \bar{p}$ on complementary subsets of the undeformed configuration boundary A . A weak form of the equation of equilibrium, more suitable for discretisation, is the **principle of virtual work** :

$$\int_V \mathrm{tr}(G^T P)\, dV = \int_A w^T \bar{p}\, dA \quad , \quad \forall\, w \tag{1}$$

where $w = w(X)$ is an arbitrary test function, best interpreted as a virtual displacement, with material gradient $G = \partial w/\partial X$ and $\mathrm{tr}(G^T P) = G \cdot P = G_{iI}\, P_{iI}$. The same BC as above must be prescribed, except for the force BC $p(X) = \bar{p}$, which has been inserted in (1) since $w(X)=0$ wherever $u(X) = \bar{u}$. Principle (1) govern the equilibrium of all deformable bodies, regardless of their constitutive materials. A constitutive law is needed to complete the formulation. For instance, an elastic-plastic law based on Drucker-Prager criterion [2] represents an instructive model to keep in mind for understanding Coulomb's friction formulation to come.

I.2 Contact mechanics summary

A dry static contact occurs when two bodies, gradually pressed together, coalesce over a certain portion of their boundaries. By hypothesis, the two bodies can come in contact, deform each other and then separate, but they cannot penetrate each other across the interface. Along the contact surface, they may stick, slip or rub against each other. The formulation of a

---

* Readers interested in dynamic impact problems are referred to [3,6,7] for corresponding formulations.

** The nominal stress vector and tensor are related to the spatial (Cauchy) stress vector $t[x(X)]$ and tensor $\sigma[x(X)]$ by $pdA = tda$ and $P = J\, F^{-1}\, \sigma\, F^{-T}$.

Accordingly, normal contact is characterized by two complementary unilateral constraints : the kinematic condition of "impenetrability" and the static condition of "intensility"

$$d_n \geq 0 \quad , \quad p_n \leq 0 \quad , \quad p_n \, d_n = 0 \tag{6}$$

Thus, either X^1 and $X^2(X^1)$ are separated ($d_n>0$, $p_n=0$) or they are in contact ($d_n=0$, $p_n<0$) the later alternative showing that the relationship $p_n(d) = p_n(d_n)$ is a multivalued function. Using the formalism of convex analysis, this law can be compacted in a single statement [8,9]

$$p_n \in \partial \Psi_{R^+}(d_n) \tag{7}$$

where Ψ_{R^+} is the indicator function of the positive half-line and $\partial\Psi$ its generalized gradient* [10]. Inclusion (7) shows that p_n derives from a non-differentiable potential. If the intensility condition (6b) is general, the impenetrability condition (6a) is restricted to small curvatures and fairly straight approach trajectories.

b. **Isotropic rigid-adherence perfect-friction law.** Adherence is associated to sticking resistance whereas friction is reserved for sliding resistance. "Rigid" adherence neglects reversible microslips (due to elastic deformations of asperities) whereas "perfect" friction excludes wearing-in mechanisms. More specifically, the tangential friction law considered here is based on Coulomb's criterion to delimit adherence from friction and a non-associated slip rule for governing the slip velocity [11,12,13,14,15,16]

$$\dot{d}_t = \alpha \frac{\partial Y(p)}{\partial p_t} = \alpha \frac{p_t}{\|p_t\|} \quad , \quad \alpha \geq 0 \quad , \quad Y(p) = \|p_t\| + \mu \, p_n \leq 0 \quad , \quad \alpha \, Y = 0 \tag{8}$$

where μ is the (constant) coefficient of friction, $\|p_t\| = |p_t|$ in 2D and $= \sqrt{p_{t_1}^2 + p_{t_2}^2}$ in 3D and α is a positive multiplier introduced to express the colinearity of the slip velocity with the friction force. Thus, either X^1 and $X^2(X^1)$ adhere to each other ($\dot{d}_t=0$, $\|p_t\|<-\mu p_n$) or they slip on one another ($\dot{d}_t\neq 0$, $\|p_t\|=-\mu p_n$) ; the former alternative showing that the relationship $p_t[d] = p_t[d,\dot{d}_t]$ is again a multivalued function. If Coulomb's cone section is denoted by $C(p_n)$, then the friction criterion and the slip rule (8) can be combined in a single inclusion [8,9]

$$\dot{d}_t \in \partial \Psi_{C(p_n)}(p_t) \quad , \quad C(p_n) = \{ \, p_t \, / \, \|p_t\| + \mu \, p_n \leq 0 \, \} \tag{9}$$

In this non-smooth analysis formalism, the lack of normality (i.e. the absence of a potential) is reflected by the dependence of the convex set on the normal stress. Provided the identification of the slave particle $X_0^2(X^1)$ which first came in contact with X^1 is correct, the initial condition for the slip rule is merely $d_{to} = 0$.

---

* In essence, the generalized gradient is the convex hull of all the "adjacent" gradients : $\partial f(x) = co \, \{\lim \nabla f(x_i), \, x_i \to x\}$ e.g. $\partial f(x) = [f'_-(x), \, f'_+(x)]$ if x and f are scalar. The generalized gradient reduces to the classical gradient wherever f(x) is smooth and to the subgradient [8] whenever f(x) is convex.

II. CONTACT AND FRICTION LAW TREATMENT

The multivalued character of the rigid laws (6/7-8/9) and the rate nature of the slip rule (8/9) require specific treatments.

II.1 Adherent contact penalization method

Perhaps the simplest method to model the "rigid" aspects of the contact and adherence laws (6-9) is to allow for a slight penetration proportional to the compression in the normal direction and a microslip proportional to the shear in the tangential direction. To this end, the tangential contact distance d_t is partitioned into the sum of a (reversible) adhesive part d_t^a, proportional to the shear p_t, and an (irreversible) slip part d_t^s, governed by the slip rule (9) [14]. The adherent contact law resulting from this treatment (after elimination of $d_t^s = d_t - d_t^s$) is

$$p = \frac{1}{\varepsilon}\left[d_n\, n + (d_t - d_t^s)\right] \qquad \text{if} \quad d_n < 0 \tag{10}$$

where $1/\varepsilon$ is a **penalty** coefficient, taken large in comparison to the stiffnesses of the contacting solids, to keep the penetration and microslip infinitesimal. The penalization method replaces the exact inclusions (7) and (9) by approximate (but continuous) functions without introducing any additional variable. An exact treatment of the rigid laws (6-9) requires the introduction of one extra variable (namely a dual Lagrange multiplier) in addition to the displacement and one extra equation (typically the complementarity condition) in addition to the equilibrium one. This technique has been widely used for the normal contact law [5,6] but more seldomly applied to the tangential friction law.

II.2. Friction projection method

Explicit forms of the friction law (8/9) depend on the algorithm used to integrate the slip rule in time. The predictor-corrector algorithm described here is an adaptation of the radial return algorithm used in plasticity [17], also called the catching-up algorithm [8] or implicit projection algorithm.

In short, given a new contact distance d_t and an old slip d_{to}^s, the new final shear stress p_t is obtained by projecting a trial adhesive stress $p_{to} = \frac{1}{\varepsilon}(d_t - d_{to}^s)$ on the final friction criterion $Y(p_t) < 0$ and the new slip d_t^s updated accordingly. In case of **adherence**, (i.e. if p_{to} satisfies the criterion $Y(p_{to})<0$), the projection degenerates into a mere identity $p_t = p_{to}$. In case of **slip** (i.e. if p_{to} violates the criterion $Y(p_{to})>0$) the projection reduces to a similarity, due to the particularly simple form of the isotropic criterion $C(p_n)$: a segment in 2D and a disk in 3D*

$$p_t = \frac{-\mu p_n}{\|p_{to}\|}\, p_{to} = -\frac{1}{\varepsilon}\,\mu d_n \frac{d_t - d_{to}^s}{\|d_t - d_{to}^s\|} \qquad \text{if} \quad \|d_t - d_{to}^s\| + \mu d_n > 0 \tag{11}$$

To complete this description, it is emphasized that the slip history d_t^s is updated at convergence only, to avoid premature adherence during iterations.

---

* In case of an anisotropic rugosity, modeled for instance by an elliptic criterion, no closed form solution for the implicit projection seems available.

III. SPATIAL DISCRETISATION BY THE FINITE ELEMENT METHOD

The FEM replaces the exact problem continuous in space by an approximate discrete problem, more amenable to computations.

III.1 Solid discretisation summary

Within the solids, the basic idea is to approximate the virtual and real displacement functions occuring in (1) by means of finite expansions* in the form $u_h^i(X) = \Phi^M(X)U^M$ (M = 1, NB.NODES) where U^M are discrete displacement values at the nodes M and $\Phi^M(X)$ are piecewise polynomial basis functions, equal to unity at node M and to zero at all the other nodes. Substitution of these expansions in (1) results in a **system of nonlinear equations** [18]

$$\mathbf{K(U) = Q} \tag{12}$$

where **U** denotes the displacement vector, **K(U)** the internal force vector and **Q** the external force vector. Of course, BC are also discretized. Expressions for **K(U)** and **Q** in case of axisymmetry (from which the 2D and 3D cases are easily recovered) can be found in [19].

III.2 Contact discretisation

The spatial discretisation of the contact term (4) is less classical. Consistent with the asymmetric definition of the continuous contact surface A^1, the discrete contact surface is defined as the set of nodes I located on the mesh boundary of body-1. The definition of a discrete contact depends then on the meshes and the kinematics of the two bodies. If a one-to-one correspondance between the boundary nodes of the two bodies can be established and maintained throughout the contact duration (adherent or small slip contact), then a node-on-node geometry is quite adequate. Otherwise a node-on-facet contact must be used to account for initial mismatching as well as subsequent sliding (moderate slip contact).

With the **node-on-node** geometry, the slave node J(I) on mesh-2 which is the closest to I can be assigned in advance and the discrete contact distance is defined as the nodal distance

$$D^I = x^{2J} - x^{1I} = X^{2J} - X^{1I} + U^{2J} - U^{1I} \qquad \text{(J(I) known)} \tag{13}$$

This definition supposes that nodes I and J come in contact exactly one on top of the other which is exceptionally the case. Slight deviations from this situation may be conveniently accounted for by initializing the slip history to $d^s_{to} = D^I_t$ (instead of 0) at impact time.

With the **node-on-facet** geometry, the slave particle location $X^{2I} \equiv X^2(X^{1I})$ on the slave facet can be expressed in terms of the facet corner coordinates X^{2J} (J=1,nb. corners) by means of the interpolation functions used for body-2 : $X^{2I}_h = \Phi^{2J}(X^{2I})X^{2J}$. The discrete contact distance becomes

$$D^I = x^{2I}_h - x^{1I} = \Phi^{2J}(X^{2I})(X^{2J} + U^{2J}) - X^{1I} - U^{1I} \qquad \text{(J(I) known, } X^{2I} \text{ unknown)} \tag{14}$$

---

* Summation on repeated indexes is assumed throughout.

In this case, the identification of the slave particle X^{2I} (i.e. the solution of the contact equation $D^I = 0$) is the crucial step for a correct evaluation of the contact distance. Since this operation depends on the interpolation functions used, it is not described here. Particular treatments may be found in [4,7,20].

Consistent with an isoparametric approximation of the geometry and displacement of body-1, an approximate contact distance is then defined in both cases as $d_h(X^1) = \Phi^{1I}(X^1)\, D^I$. Because this approximate contact distance is continuous and piecewise differentiable, the contact stress may be assumed discontinuous (e.g. piecewise constant) over each master facet and the integral in (4) can be approximated by a discrete sum over the master contact nodes* [5,21,22]

$$\int_{A_1} p(d_h) \cdot \overset{v}{d}_h \, dA = \sum_I F^I(D^I)\, \overset{v}{D}^I \qquad (15)$$

In (15), $F^I(D^I) = p(D^I)\, A^I$ where A^I can be interpreted as a tributary area of master node I, are contact forces concentrated at the master nodes and $\overset{v}{D}^I$ is the variation of D^I. For the node-on-node contact, $\overset{v}{D}^I = W^{2J} - W^{1I}$ produces two equal and opposite nodal forces at nodes I and J (a discrete version of the principle of action and reaction). For the node-on-facet arrangement, the contact distance variation $\overset{v}{D}^I$ is more complicated to derive (due to the dependence of X^{2J} on X^{1I}) and the reaction distribution on the facet nodes also [4,20]. The nodal contact distance and force are resolved into normal and tangential components as in (5). For the node-on-node contact, the contact normal is assumed to be known a priori and to remain fixed throughout the process i.e. $n = n^1[x^1(X^1)] = n^1(X^{1I})$. For the node-on-facet contact, the facet normal $n = -\, t_1 \times t_2 = -\, n^2[x^2(X^{2I})]$ is used for this purpose. The continuous frictional-contact law (10-11) is directly transformed into a discrete law upon replacing d by D and p by F.

From a **programming standpoint**, it is convenient to look at discrete contacts formed by a master node I and a slave node or a slave facet J as a node-on-node or a node-on-facet element respectively (in spite of their unconventional node pattern). In this perspective, if $F(U)$ denotes the global vector obtained by assembling the contact element forces $\mp F^I$, **the global equilibrium of two discrete bodies in frictional contact can be summarized by**

$$G(U) \equiv K(U) + F(U) = Q \qquad \text{equilibrium} \qquad (a)$$

$$F(U) = \underset{I}{A}\, F^I[D^I(U)] \quad \left(= \underset{I}{A} \begin{pmatrix} -F[D] \\ +F[D] \end{pmatrix}\right) \qquad \text{assembly} \qquad (b)$$

(16)

	$D_n > 0$	$D_n < 0$	
		$\lvert D_t - D^s_{to}\rvert + \mu D_n \leq 0$	$\lvert D_t - D^s_{to}\rvert + \mu D_n > 0$
$F[D] =$	0	$\frac{1}{\varepsilon}(D - D^s_{to})$	$\frac{1}{\varepsilon}\left(D_n n - \mu D_n \frac{D_t - D^s_{to}}{\lvert D_t - D^s_{to}\rvert}\right)$ (c)
	gap	adherent contact	slip contact

* The same result can be obtained by assuming a continuous piecewise differentiable contact stress distribution and using special quadrature rules to lump the contact forces at the nodes such as the trapezoidal rule for piecewise linears and Simpson's rule for piecewise quadratics.

IV. A GENERALIZED NEWTON METHOD FOR NON-SMOOTH OPERATORS

Based on successive linearizations, the Newton method replaces the discrete nonlinear problem (16) by an iterative sequence of linear problems, directly solvable by standard methods of linear algebra. Formally, the algorithm may be summarized as follows [24]

$$\begin{cases} U^0 = \bar{U}\ ,\ k = 0 & \text{(a)} \\ [E(U^k) + J(U^k)] \cdot \Delta U = Q - K(U^k) - F(U^k) & \text{(b)} \\ U^{k+1} = U^k + \Delta U & \text{(c)} \end{cases} \qquad (17)$$

where ΔU is the displacement increment, k the iteration index and $E(U) = dK/dU$ is called the **tangent stiffness matrix** and $J(U) = dF/dU$ the **tangent contact matrix.** Stricly speaking, these two jacobian matrices are not defined everywhere since the internal and contact force vectors K and F are only piecewise differentiable. At singularities, they should be replaced by elements of the **generalized jacobian** defined in the next section, where Newton's method is also properly generalized (thus the title of this article). However, the probability for an iterate U^k to fall right at a singularity is close to null in finite arithmetic and if by exception this situation arises, then anyone of the "adjacent" jacobians turns out to be adequate, so that (17) is acceptable for all practical purposes.

The expression of the tangent stiffness matrix in case of axisymmetry (from which the 2D and 3D expressions are easily derived) can be found in [19]. For a plastic law, it involves a tangent elasto-plastic tensor which can be found in [17]. Just like the contact force F in (16b), the global jacobian J can be obtained by assembly of local contact jacobians J^I

$$J(U) = \mathop{\mathbf{A}}_{I}\ J^I[D^I(U)] \qquad \left(= \mathop{\mathbf{A}}_{I} \begin{bmatrix} J & \bar{J} \\ \bar{J} & \end{bmatrix}\right) \qquad (18)$$

The element matrix pattern indicated in parentheses in (18) corresponds to the node-on-node configuration. For this node-on-node case with fixed normal, the discrete contact jacobian $J = dF/dD$ can be derived from (16c) to be

$J =$	Gap	adherent contact	slip contact*	
Intrinsic formulas 2D and 3D	0	$\frac{1}{\varepsilon}\,\mathbf{1}$	$\frac{1}{\varepsilon}[n \otimes n - \mu\, t \otimes n + \rho(1 - n \otimes n - t \otimes t)]$	(19)
In local coordinates 2D : (t,n)	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\frac{1}{\varepsilon}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	forward $\frac{1}{\varepsilon}\begin{bmatrix} 0 & -\mu \\ 0 & 1 \end{bmatrix}$	backward $\frac{1}{\varepsilon}\begin{bmatrix} 0 & \mu \\ 0 & 1 \end{bmatrix}$ (20)
In local coordinates 3D : (t_1,t_2,n)	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\frac{1}{\varepsilon}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\frac{1}{\varepsilon}\begin{bmatrix} \rho s^2 & -\rho sc & -\mu c \\ -\rho sc & \rho c^2 & -\mu s \\ 0 & 0 & 1 \end{bmatrix}$	(21)

where $\mathbf{1} = \delta_{ij}$ is the identity matrix, $t = (D_t - D^s_{to})/\|D_t - D^s_{to}\|$ is the trial slip increment unit vector, $\rho = -\mu D_n/\|D_t - D^s_{to}\| \in (0,1]$ is the projection scaling factor, $\otimes$ denotes the diadic product $tn = tn^T = t_i n_j$ and $s = \sin\theta = t.t_1$ and $c = \cos\theta = t.t_2$.

---

* The slip contact jacobian is derived as follows :

$$\frac{dF}{dD} = \frac{1}{\varepsilon}\left[n \otimes \frac{dD_n}{dD} - \mu\, t \otimes \frac{dD_n}{dD} - \mu\, D_n \frac{dt}{dD}\right] \quad \text{where} \quad \frac{dD_n}{dD} = n \text{ and } \frac{dt}{dD} = \frac{1}{\|D_t - D^s_{to}\|}[1 - nn - tt]$$

b. 3D truss example [30]. Turn now to the 3D truss shown in Fig. 1b. The stiffness matrix may be partitioned into

$$E = \begin{bmatrix} K & k \\ k & k_n \end{bmatrix}$$ where K is a 2x2 tangential stiffness submatrix, k_n the normal stiffness and k a coupling vector of $\mathbb{R}^2$.

The contact generalized jacobian extremes evaluated at $D_0 = 0$ are now the three matrices listed in (21), the last being variable. The uniqueness condition $\det(E+J_B) > 0$, $\forall s,c$, implies the restriction

$$\mu < \lambda^K_{min}/\|k\| \quad (= \cos\phi/\sqrt{2}\sin\phi \text{ if } \phi = \omega \text{ in Fig. 1b}) \tag{24}$$

where λ^K_{min} denotes the smallest eigenvalue of K . This 3D condition degenerates well in the 2D case.

V.2. Algorithm convergence

A rigorous and straightforward extension of Newton's method (17) to solve Lipschitzian equations such as (16) consists in replacing (17b) by

$$H^k \cdot \Delta U = Q - G(U^k) , \quad H^k \in \partial G(U^k) \tag{25}$$

where $\partial G(U^k)$ is the generalized jacobian of G at U^k . Newton's method is well known for its fast rate of local convergence but also for its small radius of convergence. For instance, the method may cycle between forward and backward slip (when the solution is stick) in an as simple problem as the 2D truss in Fig. 1a [30]. Ensuring its global convergence usually requires either drastic conditions (G of class C^1, monotony, convexity) or special damping techniques difficult to implement [24]. However, in the present context a simple damping method and an even simpler control procedure can be successfully applied.

a. 2D systematic damping method. In the 2D case where the contact operator is conewise linear, it is possible to adapt a damping technique, originally developed by [32] and [33] for piecewise linear operators, which guarantees global convergence (provided the solution is unique). In essence, when progressing from U^k towards U^{k+1} in the direction $H^{k-1}[Q-G(U^k)]$, this method consists in stopping on the first encountered hyperplane delimiting the cone containing U^k . The next iteration is carried out using the jacobian matrix associated with the next cone on the other side of the hyperplane. More specifically, if the cone C containing U^k is indexed with the iteration counter k to simplify notations, then Newton's update step (17c) is amended into

$$U^{k+1} = U^k + \alpha\, \Delta U \quad \in \quad C^k \cap C^{k+1} \tag{26}$$

where α is a damping factor determined by solving the inclusion in (26) as follows. Because the global contact operator $F(U)$ is the assembly of N local contact contributions $F[D]$, each global cone C^k is the cartesian product of N local cones (chosen among 4^N combinations)

the operator F is :

- **conewise linear** in 2D (linear on 4^N convex cones with apex $-D_0$ where $D_0 = X^2 - X^1 - D^s_{to}$ and N denotes the number of master contact nodes),
- **raywise linear** in 3D (linear on 2^N convex cones and on an infinite number (∞) of half-lines with origin $-D_0$).

Due to these additional properties, the following specialization of theorem 1 to flat contacts (i.e. for which the normal n is constant along the contact surface A^1) is proved in [30].

Theorem 2 : A necessary and sufficient condition for the discrete frictional flat contact problem $E.U + F(U) = Q$ to have a unique solution for any Q, is that the **extreme** matrices of its generalized jacobian evaluated at the origin be non singular :

$$\forall J_i \in \overset{\circ\circ}{\partial F}[-D_0] \ , \quad \det(E + J_i) > 0 \tag{22}$$

In (22), $\overset{\circ\circ}{\partial F}[-D_0]$ represents the extreme matrices obtainable by assembly of the contact element jacobian matrices (19), in all possible combinations. They are in finite number (4^N) in 2D and an infinity ($2^{N+\infty}$) in 3D. Necessity of (22) is obvious from linear algebra but sufficiency is not trivial, as indicated by several counter examples [30]. When applied to a specific problem, conditions (22) provides the sharpest possible bound on Coulomb's friction coefficient for a unique solution as illustrated by the following two elementary examples.

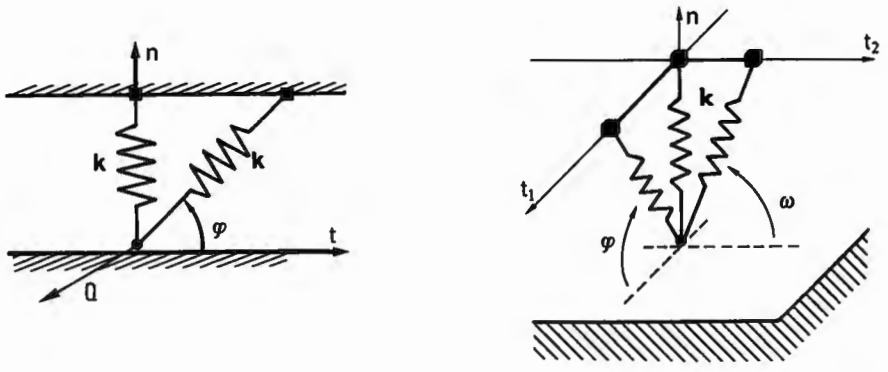


Fig. 1 2D and 3D truss examples

a. 2D truss example [15]. Consider the 2D truss depicted in Fig. 1a, with its only free node in grazing contact with a rigid plane. Let

$$E = \begin{bmatrix} k_t & k_{tn} \\ k_{tn} & k_n \end{bmatrix} \quad \left(= k \begin{bmatrix} \cos^2\phi & \sin\phi\cos\phi \\ \sin\phi\cos\phi & \sin^2\phi + 1 \end{bmatrix}\right)$$

denote the stiffness matrix of the truss condensed at the free node.

The generalized jacobian extremes of the frictional contact operator F evaluated at $D_0 = X^2 - X^1 - D^s_{to} = 0$ are the four matrices listed in (20) and the generalized jacobian is their convex hull : $\overset{\circ\circ}{\partial F}(0) = \{(20)\}$ and $\partial F(0) = \text{co}\ \overset{\circ\circ}{\partial F}(0)$. Enforcement of the conditions $\det(E + J_i) > 0$ produces for i=3 (forward slip) the condition

$$\mu < k_t / |k_{tn}| \quad (= \cos\phi/\sin\phi) \tag{23}$$

A direct analysis of the truss, shows that (23) is indeed the exact uniqueness condition [15].

b. 3D truss example [30]. Turn now to the 3D truss shown in Fig. 1b. The stiffness matrix may be partitioned into

$$E = \begin{bmatrix} K & k \\ k & k_n \end{bmatrix}$$ where K is a 2x2 tangential stiffness submatrix, k_n the normal stiffness and k a coupling vector of $\mathbb{R}^2$.

The contact generalized jacobian extremes evaluated at $D_0 = 0$ are now the three matrices listed in (21), the last being variable. The uniqueness condition $\det(E+J_s) > 0$, $\forall s,c$, implies the restriction

$$\mu < \lambda^K_{min}/\|k\| \quad (= \cos\phi/\sqrt{2}\sin\phi \text{ if } \phi = \omega \text{ in Fig. 1b}) \tag{24}$$

where λ^K_{min} denotes the smallest eigenvalue of K . This 3D condition degenerates well in the 2D case.

V.2. Algorithm convergence

A rigorous and straightforward extension of Newton's method (17) to solve Lipschitzian equations such as (16) consists in replacing (17b) by

$$H^k \cdot \Delta U = Q - G(U^k) , \quad H^k \in \partial G(U^k) \tag{25}$$

where $\partial G(U^k)$ is the generalized jacobian of G at U^k . Newton's method is well known for its fast rate of local convergence but also for its small radius of convergence. For instance, the method may cycle between forward and backward slip (when the solution is stick) in an as simple problem as the 2D truss in Fig. 1a [30]. Ensuring its global convergence usually requires either drastic conditions (G of class C^1, monotony, convexity) or special damping techniques difficult to implement [24]. However, in the present context a simple damping method and an even simpler control procedure can be successfully applied.

a. 2D systematic damping method. In the 2D case where the contact operator is conewise linear, it is possible to adapt a damping technique, originally developed by [32] and [33] for piecewise linear operators, which guarantees global convergence (provided the solution is unique). In essence, when progressing from U^k towards U^{k+1} in the direction $H^{k-1}[Q-G(U^k)]$, this method consists in stopping on the first encountered hyperplane delimiting the cone containing U^k . The next iteration is carried out using the jacobian matrix associated with the next cone on the other side of the hyperplane. More specifically, if the cone C containing U^k is indexed with the iteration counter k to simplify notations, then Newton's update step (17c) is amended into

$$U^{k+1} = U^k + \alpha\, \Delta U \quad \in \quad C^k \cap C^{k+1} \tag{26}$$

where α is a damping factor determined by solving the inclusion in (26) as follows. Because the global contact operator $F(U)$ is the assembly of N local contact contributions $F[D]$, each global cone C^k is the cartesian product of N local cones (chosen among 4^N combinations)

Consequently the global damping factor α is the minimum of N local factors α^I relative to each discrete contact

$$\alpha = \min_{I=1,N} \alpha^I \quad ; \quad \alpha^I = \min \{ \{ -\frac{D_n^k}{\Delta D_n} ; -\frac{\mu D_n^k + D_t^k - D_{to}^s}{\mu \Delta D_n + \Delta D_t} ; -\frac{\mu D_n^k - D_t^k + D_{to}^s}{\mu \Delta D_n - \Delta D_t} ; 1 \} \cap (0,1] \} \quad (27)$$

Convergence occurs in a finite number ($<4^N$) of iterations. In essence, the proof [30] relies on the fact that, while U^k wanders along a broken line in displacement space, $G(U^k)$ progresses, in force space, along the straight segment $[G(U^P), Q]$ which crosses the convex image of each C^k at most once. From a finite element standpoint, the procedure consists in **changing the status of only one contact element per iteration,** which is a common strategy in optimization [15]. An advantage of the present technique is that it can be turned on only when cycling is detected. The technique can be applied to the 3D problem provided Coulomb's cone is replaced by a prism with a finite number of facets as in [15].

b. Slip reversal control technique. Another technique which has proved very reliable in practice (to the extent that it is prefered to the above damping technique even in 2D) consists in enforcing adherent contact wherever and whenever a slip reversal is detected between two successive iterations. More specifically (16) and (19) are modified by enforcing

$$F[D] = \frac{1}{\varepsilon}(D - D_{to}^s) \qquad J = \frac{1}{\varepsilon} 1 \qquad \text{if} \quad D_n^k < 0 \quad \text{and} \quad \begin{array}{l} \| D_t^k - D_{to}^s \| + \mu D_n^k \leq 0 \\ \text{or} \quad (D_t^{k+1} - D_{to}^s) \cdot (D_t^k - D_{to}^s) < 0 \end{array} \quad (28)$$

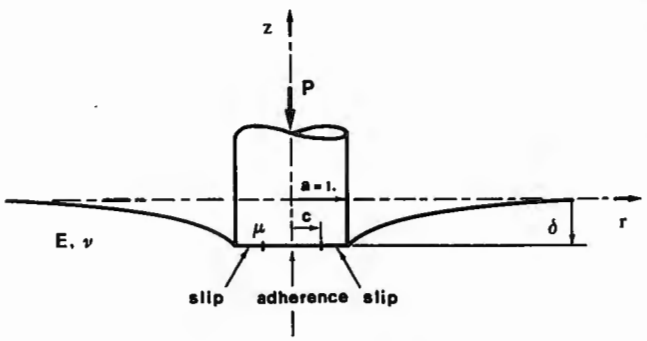
It is in order to recall that the slip history D_t^s is updated at convergence only (to prevent erroneous adherence during iterations) and to add that the adherent contact status (28) is enforced at the first iteration of each new load increment if $D_n^0 < 0$, (to detect unloading as early as possible). Unlike (27), procedure (28) may change the status of several contact elements per iteration, presumably progressing faster towards the solution. However convergence has not been established for this ad hoc procedure. It is applicable to 2D and 3D problems and easier to implement than (27).

VI. THE FLAT PUNCH BENCHMARK PROBLEM

The indentation of a linear elastic half space by a rigid cylindrical flat punch with Coulomb's friction along the interface represents a good problem to test a tangential friction algorithm. Indeed, if the normal contact problem poses little difficulty (edge singularity excepted) since the contact area remains constant, the tangential friction problem is delicate because the partition of the contact area into stick and slip bands varies rapidly with Coulomb's coefficient of friction and Poisson's contraction ratio and changes drastically upon unloading. Moreover, the flat punch problem is one of the rare contact problems involving friction for which a (semi-) analytical solution is known [34] and several numerical solutions are available [35,36].

VI.1. Definition of the problem

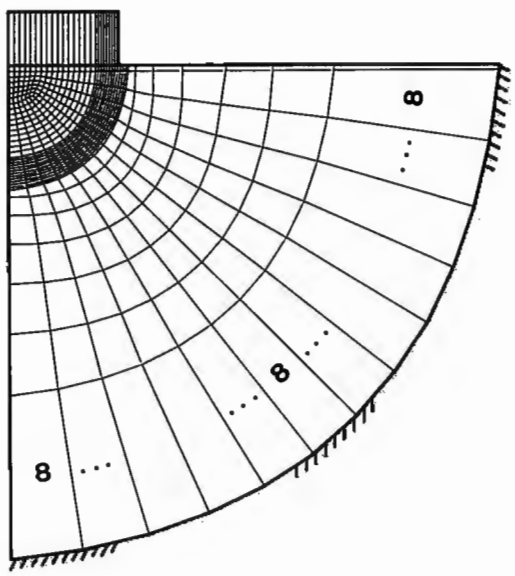
a. Analytical data. A rigid axisymmetric flat punch of unit radius is pressed against a linearly elastic half space as shown in Fig. 2. The mechanical and material data are specified besides.



Punch radius	a = 1.0 cm
Normal load	P = 6.283 daN/rad
Unloading	P/4 daN/rad
Elastic modulus	E = 314. daN/cm^2
Poisson's ratio	ν = 0.
Friction coeff.	μ = 0.4

Fig. 2 The flat punch problem.

b. Numerical data. The mesh, with takes advantage of the symmetry of revolution, is shown in Fig. 3.



Punch :
24 linear axisymmetric elements
nearly rigid material
Contact :
25 frictional contact elements
contact force by penalty method,
incremental Coulomb's friction
by radial return algorithm
Halfspace :
405 linear axisym. finite elements
13 linear axisym. infinite elements
Total :
512 nodes, 958 equations

Fig. 3 The finite element mesh.

VI.2. Solution discussion

a. Loading case. The closed form solution obtained in [34] is a complicated combination of slip and stick solutions. Resulting from a displacement (rather than a velocity) formulation, it is exclusively limited to monotonic loadings ($\dot{P} > 0$) . In its purely analytical form, it assumes

that the normal pressure is unaffected by the tangential shear which happens to be exact for an incompressible material only. By using an iterative numerical scheme, which in essence solves the pressure and the shear problems in alternance, [34] succeeds to compute the exact solution for compressible materials as well.

The main characteristic of the solution is the division of the contact area into an inner stick disk and an outer slip annulus. Paradoxaly, slip is directed **inward.** The radius c of the circle delimiting the stick part from the slip part depends exclusively on Coulomb's coefficient of friction at the interface and on Poisson's contraction ratio of the elastic half space : $c = c(\nu,\mu)$. The values $\nu = 0.$ and $\mu = .4$ are selected for the numerical test because they maximize the coupling between pressure and shear (which is essential in Coulomb's friction) as well as the sensitivity of c on μ . The most representative aspect of the solution is the radial distribution of the tangential shear normalized by the product of the normal pressure by the coefficient of friction as shown in Fig. 4. The outer plateau corresponds to the slip annulus.

The numerical solution is obtained with the program TACT developed in the authors laboratory [7]. The shear distribution is obtained from the contact nodal forces and is plotted over the analytical curve. The good agreement provides a first indication of the correct functioning of the friction algorithm. (Note that the solution does not depend on the number of load increments i.e. it is path independent as already discussed).

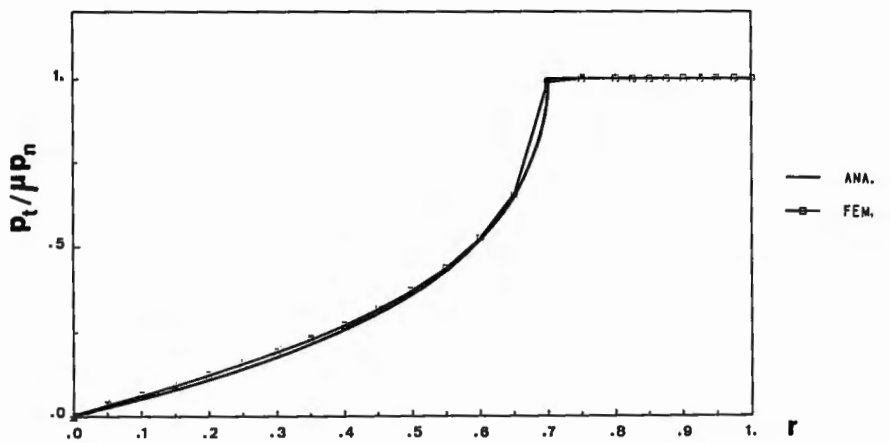


Fig. 4 Shear radial distribution (loading case).

b. Unloading case. Numerical solutions to the unloading problem have been obtained by [35, 36]. The results show that a second stick annulus develops from the punch edge as soon as unloading is initiated, pushing the inward-slip annulus towards the center. For even lighter loads, a second outward-slip annulus develops from the punch edge pushing both the first slip and the second stick annuli towards the center. This complex (stick / inward-slip / stick / outward-slip) pattern represents a severe test for friction algorithms. In particular, convergence to the exact solution may require several load decrements in order to track this path dependent process with accuracy.

The results presented here were obtained after unloading the punch down to one fourth of the original load ($P/4=\pi/2$). The shear distribution obtained after 1, 3 and 6 load decrements, using slip reversal control, are plotted in Fig. 5 to show the convergence to the presumed solution. The solution is in qualitative agreement with the ones obtained in [35,36] with different data.

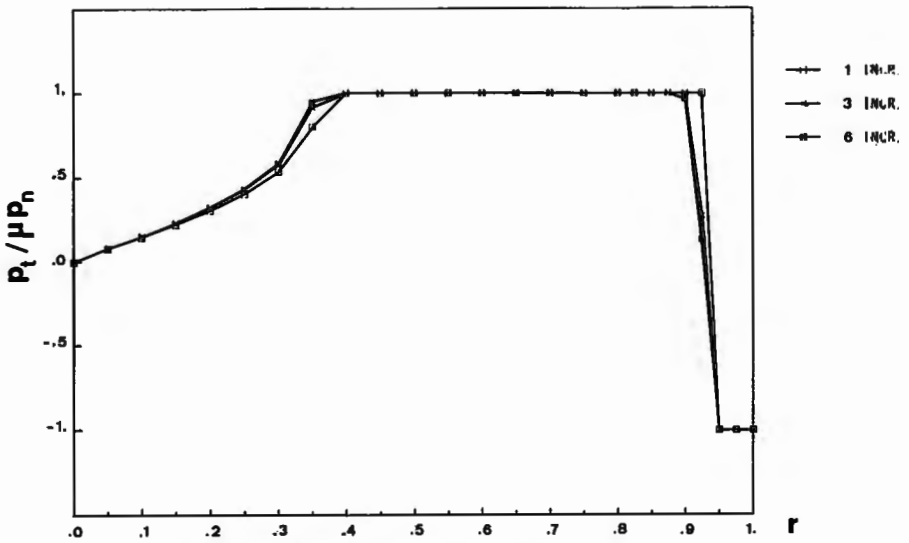


Fig. 5 Shear distribution convergence (unloading case)

VI.3. Algorithm performance assessment

Here, the robustness and efficiency of the generalized newton method without damping nor control, with systematic damping and with slip reversal control are compared on the basis of the 2D flat punch loading and unloading problem just described. The numbers of iterations (i) and seconds (s) of CPU time (on a DEC-VAX-780) are retained as comparative elements. The results are listed in Table I.

Table I algorithm performance trends

	Newton	damping	control
LOADING P			
1 increment	7i 397s	10i 666s	7i 405s
UNLOADING P/4			
1 decrement	cycles	9i 601s	6i 350s
3 decrements	cycles	17i 1096s	16i 871s
6 decrements	cycles	25i 1580s	25i 1340s
TOTAL	--	> 19i 1267s	> 13i 755s
	--	< 35i 2246s	< 32i 1745s

As anticipated, the loading problem is too easy to assess the robustness of a friction algorithm. All three variants converge. The damping technique spends superfluous iterations on this path independent problem. The unloading problem is much more selective. The plain Newton method diverges (except for a very small initial decrement) which disqualifies it to solve frictional contact problems. Both the damping and control techniques make it converge. Their rate of convergence is comparable. Another measure which gives an idea of the overall efficiency of the algorithm is the ratio of the CPU time taken to solve the contact problem with friction over that taken for the linear elasticity problem with a fixed interface. This ratio comes out roughly proportional to the number of iteration as $\rho = 3i/2$, showing that contact elements consume as much as 50 % of the time used by the elastic elements.

VII. CONCLUSION AND RECOMMENDATIONS

In this article, a combination of numerical methods has been proposed to solve contact problems involving friction. The resulting frictional contact algorithm has proved both robust and efficient for 2D and 3D small slip problems. Several developments would be welcome however. They include

- A symmetric definition of the contact distance, applicable to large slip problems and better suited for finite element discretisation.
- More sophisticated contact and friction laws, accounting for adhesion, wear, rate dependence, anisotropy ...
- Uniqueness condition(s) for the continuous contact problem and convergence proof(s) for the 3D algorithm.
- Extension of the current contact methodology to impact problems.

REFERENCES

1 Curtin M., **An Introduction to Continuum Mechanics**, Academic Press, 1981.
2 Green E., Naghdi P., Arch. Mat. Mech. Anal., 18, 1965, p.251.
3 Curnier A., "Continuum Mechanics Simulation of the Basic Cavitation Erosion Process", LMA-DME-EPFL Archives, 1987.
4 Curnier A., Appendix C of Contact-Impact Problems, Vol.1 by Taylor &. Sackman, Nat. Tech. Inf. Serv. No DST-HS-6-01443, 1980, p.C1.
5 Simo J., Wriggers P., Taylor R.L., Comp. Math. Appl. Mech. Engng, 50, 1985, p.163.
6 Hughes T.R. J., Taylor R.L., Sackman J.L., Curnier A. and KanoknuKulchai W., Comp. Meth. Appl. Mech. Eng., 8, 1976, p.249.
7 Curnier A., TACT Methodology Summary, LMA-DME-EPFL Archives, 1983.
8 Moreau J.J., in **New Variational Techniques in Mathematical Physics**, Capriz & Stampacchia (coord.). Edizioni Cremonese, 1974.
9 Panagiotopoulos P.D., Ingenieur-Archiv vol. 44, 1975, p.421.
10 Clarke F.H., **Optimization and nonsmooth analysis**, Wiley, 1983.
11 Seguchi Y. et al, Proc. Conf. on Comp. Meth. in Nonl. Mech., Univ. of Texas, 1974, p.683.
12 Fredriksson B., Comp. & Struct, 6, 1976, p.281.
13 Michalowski R., Mroz Z., Arch. of Mech., 30-3, 1978, p.259.
14 Curnier A., Int. J. Solids Struct., 20-7, 1984, p.637.
15 Klarbring A., Thesis No 133, Linköping Univ., 1985.
16 Oden J.T., Martins J. A.C., Comp. Meth. Appl. Mech. Eng, 1985, p.
17 Simo J., Taylor R., Comp. Meth. Applied Mech., 48, 1985, p.110.
18 Oden T., **Finite Elements of nonlinear Continua**, Mc Graw Hill, 1972.
19 Rakotomanana L., thesis, LMA-DME-EPFL Archives, 1986.
20 Simo C., Wriggers P., Schweizerhof N., Taylor R., Int.J.Num. Meth. Engng, 23, 1986, p.779.
21 Campos L.T., Oden J.J., Kikuchi N., Comp. Meth. Appl. Mech. Engng, 34, 1982, p.821.
22 Johansson L., Report LiTH-IKP-EX-583, Linkoping Univ., 1986.
23 Oden J.T. and Pires E.B., Comp. & Struct. 16, 1-4, 1983, p.481.
24 Ortega R., Rheinbolt W., **Iterative Solutions of Nonlinear Equations in Several Variables**, Academic Press, New York, 1970.
25 Necas J., Jarusek J. and Haslinger J., Bollettino U.M.I. (5), 17-B, 1980, p.796.
26 Janosky V., Proc. Math. of Finite Elements and Applications, Brunel Univ., 1981.
27 Demkowicz L., Oden J.T., **Nonlinear Analysis, Theory, Meth.& Appl.** 6-10, 1982, p.1075
28 Oden J. T. and Pires E.B., J. of Appl. Mech. 50, 1983, p.67.
29 Cocu M., Int. J. Engng Sci., 22, 5, 1984, p.567.
30 Alart P., Curnier A., Document, LMA-DME-EPFL Archives, 1986.
31 Palais R.S., Trans. Amer. Math. Soc., 92, 1959, p.125.
32 Katzenelson J., Bell system Tech. J., 44, 1965.
33 Fujisawa T., Kuh E.S., SIAM J. Appl. Math., 22, 1972, p.307.
34 Spence D.A., J. of Elasticity, 5, 3-4, 1975.
35 Turner J.R., J. Inst. Maths. Applics. 24, 1979, p.439.
36 Klarbring A., Proc. ASCE/ASME Conf., 1985, p.43.