The non-smooth contact dynamics method

Michel Jean

The non-smooth contact dynamics method

M. Jean

Laboratoire de Mécanique et d'Acoustique, CNRS-ESM2 Technopôle de Château-Gombert, 13451 Marseille Cedex 20, France

The main features of the Non-Smooth Contact Dynamics method are presented in this paper, the use of the dynamical equation, the non-smooth modelling of unilateral contact and Coulomb's law, fully implicit algorithms to solve the dynamical frictional contact problem for systems with numerous contacting points, in particular large collections of rigid or deformable bodies. Emphasis is put on contact between deformable bodies. Illustrating numerical simulation examples are given for granular materials, deep drawing and buildings made of stone blocks.

1. Introduction

1.1. Dynamics

The Non-Smooth Contact Dynamics method for short Contact Dynamics, has been initiated and developed by the author and J.J. Moreau (see a paper by this last author in this volume [33]) during the last decade [20,21,11,19]. It is used for quite a number of applications but has proven to be particularly successful when collections of rigid or deformable bodies are packed together and subject to loading deformations. Advances have been made in the understanding of different discrete formulations of frictional contact. Also, much attention has been paid to the way the method converges and how to appreciate the quality of a numerical simulation, which is a difficult issue when dealing with large collections of bodies. The author has developed this method within a Fortran software LMGC, the architecture of which follows closely the mechanical setting developed hereafter, so that it may be applied as well to rigid or deformable body collections with minor options.

Models of granular materials such as Schneebeli materials (assembly of piled up rolls) have been studied during these last years with the Contact Dynamics method, yielding new results. In these divided materials, the commonly accepted ideas about distributions of stresses or strains in continuous media are not true any more. Chains of forces act between grains and strong forces support shear stress while weak forces contribute to internal pressure [27]. Localisation, shear bands, dilatancy phenomena may also be exhibited. As an application of the method, models of monuments made of rigid blocks have also been studied by J.J. Moreau. Rigid bodies are governed by the dynamical equation and the frictional contact equations. Even in supposedly quasi-static situations such as stress–strain experiments, episodes of shocks or quick rearrangements might happen (and are actually experimentally observed) that dynamics is able to describe where quasi-static algorithms fail.

When oscillations or vibrations are to be studied, for instance monuments under seismic actions or wind actions, the rigid body models are not relevant any more and elasticity of the bodies must be taken into account. Here, the physical situation is dynamical and quasi-static algorithms are irrelevant. Recent investigations on monuments made of deformable stone blocks are made with the Contact Dynamics method.

There are also quasi-static situations where dynamics may help. A typical situation is deep drawing numerical

simulation where several hard difficulties are met: large displacements and deformations, frictional unilateral contact, elasto-plasticity, plastic localisations, wrinkles, and so on. In most algorithms, at some place, within some loop, iterations of the Newton–Raphson method are performed, using a tangent matrix, some stiffness matrix. It may happen to be ill-conditioned, specially when plastic localization or wrinkles appear. In these circumstances solving the linear system is very tedious or even impossible. When dynamics are taken into consideration, the dynamical tangent matrix is well conditioned and iterations are quick. As a drawback, small time steps must be used.

1.2. Unilateral contact and dry friction

One of the main features of unilateral contact is impenetrability, which means that particles candidates for contact must not cross the boundaries of antagonist bodies. This is expressed by writing that the distance between contacting bodies or the gap g is non-negative. Also, it is assumed that contacting bodies are not attracting each other, i.e. the reaction force R is non-negative, and this reaction force vanishes when the contact is not active. The graph of this relation, see Fig. 2, commonly known as the *Signorini condition* is not a mapping graph, in the sense that neither g is a function of R, nor R is a function of g. This graph is 'infinitely steep'. As will be seen later on, such graphs may be used when implicit methods are adopted. Another way of modelling contact is to assume that the reaction force is a function of the gap violation, where violation means penetration of some reference surfaces, in the neighbourhood of the contacting bodies' boundaries. For instance, it may be assumed that the reaction force is proportional to the gap violation, referring to some elastic properties of asperities lying on contacting bodies rough surfaces. Hertz' law may also be assumed or even some law between g and weighted values of the density of reaction force on some neighbouring support of the contacting point [16]. In these circumstances, the graph of the relation between g and R appears as a steep mapping graph. Any algorithm devoted to strong non linear differentiable problems may thus be used. From a mathematical point of view, replacing the infinitely steep Signorini graph by some steep mapping graph may be considered mathematically as a *regularizing* (Moreau–Yoshida) method or from a numerical analysis point of view as a *penalty* method. The slope of the elastic response of such rough surfaces is difficult to measure with accuracy and reproducibility. Furthermore, as far as small scale phenomena are considered, tribological responses of rough surfaces are generally much more complicated than mere elasticity. To decide if, for numerical structural analysis purposes, some steep mapping graph law is a better model than the Signorini condition is thus a matter of personal acceptance and convenience according to the numerical or mathematical technique mastered by the user.

The friction under consideration in appplications such as deep drawing or contact between grains or stone blocks, is of the dry friction kind. A friction law relates the sliding velocity U_T and the reaction force R. Dry friction means that some friction force R_T, the tangential component of R, may be exerted without any sliding, and that some threshold has to be overcome to allow sliding. In *Coulomb's law*, the threshold is proportional to the normal component R_N and the friction force remains opposite to the sliding velocity with the value of the threshold when sliding is occurring. Here again, the graph relating U_T and R_T is not a mapping graph, see Fig. 3. The infinitely steep graph may be replaced by some steep mapping graph which means that in the circumstances where no sliding is to occur, some very small sliding velocity is actually produced. In the case where small tangential displacements are expected, the sliding velocity being proportional to the tangential displacement, the dry friction law is viewed by some authors as a relation between the tangential displacement and R_T. The alternative steep part of the graph may then be viewed as modelling elasticity effects of some asperities. Here again, tribological phenomena are much more complicated and choosing a friction model is a matter of personal acceptance and convenience. As far as regularized or non classical frictional contact laws are concerned basic ideas in the line of variational methods for strongly nonlinear systems may be found in [16,23,18].

In the Contact Dynamics method, the Signorini condition and Coulomb law are adopted. An equivalent form of the Signorini condition, involving the gap velocity is also introduced [19]. If local elasticity between contacting bodies is a relevant phenomenon, for instance elasticity of mortar joints between stone blocks, it may easily be taken into account in the algorithm, with a minor change of variable. Similarly, the dynamical friction coefficient or some model of adhesion may be introduced. The basic relations are still the Signorini condition and Coulomb law.

1.3. Fully implicit algorithms

Explicit algorithms are usually favored in dynamics because they are easy to implement, the number of floating point operations is low and high accuracy may be obtained. For instance, crash problems are successfully treated with explicit algorithms. It is known that equilibrium under Coulomb friction is difficult to reach with an explicit algorithm computing the frictional force from the sliding velocity at the previous time step. Oscillatory artefacts are generated which are to be killed by appropriate numerical devices. In the applications viewed in this paper, large collections of contacting bodies, the status of contacting points, no contact or sticking or sliding, are subjected to many changes. In these circumstances, though being more complex to implement, fully implicit algorithms behave nicely.

2. The Signorini condition, the Coulomb law

A candidate to contact is a unique pair of a candidate object O, and antagonist object O', with a candidate particle $P \in O$ and an antagonist particle $P' \in O'$. The particles P and P' are proximal such that $\boldsymbol{P'P}$ defines a unit normal vector $\boldsymbol{N}$ directed from the antagonist object toward the allowed region for the candidate particle. The vector $\boldsymbol{N}$ is equipped with two other vectors to form an orthonormal frame, so called local frame, see Fig. 1. Normal components of vectors in the local frame are denoted with the subscript N and tangential components orthogonal to $\boldsymbol{N}$ with the subscript T. The following mechanical items are used in order to write a frictional contact law: the components of the relative velocity of P with respect to $O' : U = (U_T, U_N)$, the components of the reaction force exerted by O' on $O : R = (R_T, R_N)$, the gap: $g = \overline{P'P}$.

When a contact is occurring the velocity or more exactly the relative velocity of the impacting particle shows a discontinuity. The classical mathematical tool to describe such velocities is the set of functions defined on some interval of time I, $t \to U(t)$, with bounded variation. A bounded variation function possesses at every time t a left limit U^-, the velocity just before the impact, if any, and a right limit, U^+, the velocity just after the impact. The proper mathematical tool to describe impulses is the set of differential measures on I. When writing $R(t)$ it is meant that $R(t)$ is some density of an impulsive reaction force with respect to some positive measure on I. For instance, if t is the time of some impact, $R(t)$ is the density of impulse with respect to a Dirac measure at time t. When smooth contact is occurring, the function R is generally a density with respect to the Lebesgue measure, and U is continuous. It is not within the scope of this paper to go into mathematical details but it is necessary to introduce the concept of right and left velocity.

2.1. Unilateral conditions

The following relations have been commented on in the Introduction:
impenetrability: $g \geq 0$,
no attraction is acting between objects: $R_N \geq 0$,

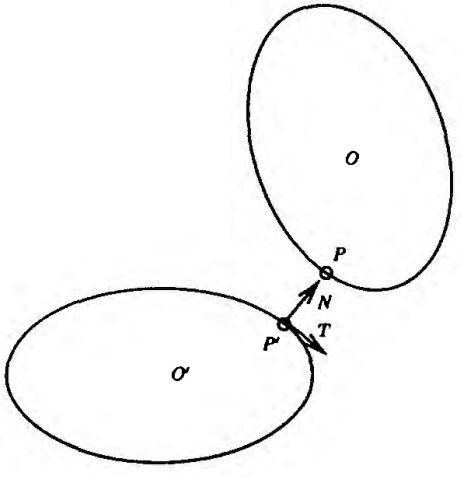

Fig. 1. Local frame.

the reaction force vanishes when the objects are not contacting: $g > 0 \Rightarrow R_N = 0$.

This set of relations may be summarized in the following equivalent complementary equation, the so-called *Signorini condition*:

$$g \geq 0 \qquad R_N \geq 0 \qquad gR_N = 0 \,, \tag{1}$$

or equivalently,

$$R_N = \mathrm{proj}_{\boldsymbol{R}^+}(R_N - \rho g) \,, \tag{2}$$

where ρ is some strictly positive number, Eq. (2) being actually verified for any strictly positive ρ as soon as it is true for some strictly positive number. The graph of this relation is displayed in Fig. 2. This a monotone multi-mapping graph. The above relations are familiar in the context of Convex Analysis. The following relation,

at some initial time t_0, $g(t_0) \geq 0$;

$$\text{for all } t \in I \quad g(t) \leq 0 \quad \Rightarrow \quad U_N^+(t) \geq 0 \,. \tag{3}$$

implies the impenetrability condition,

$$\text{for all } t \in I \quad g(t) \geq 0 \,.$$

Conversely, if the motion is smooth enough, i.e. U exists, impenetrability implies 3 for all $t \in I$. The relation 3 together with

$$R_N(t) \geq 0 \,; \qquad R_N(t) > 0 \quad \Rightarrow \quad U_N^+(t) = 0 \,, \tag{4}$$

yields the following so-called *velocity Signorini condition* [19],

at some initial time t_0, $g(t_0) \geq 0$;

$$\text{for all } t \in I, \quad \text{if } \; g(t) \leq 0 \; \text{ then} \tag{5}$$

$$U_N^+(t) \geq 0 \quad R_N(t) \geq 0 \quad U_N^+(t)\, R_N(t) = 0 \,.$$

In the following, the relation (5) is to be understood together with the relation $g(t) > 0 \Rightarrow R_N(t) = 0$, which shall not be written down. Provided that some regularity conditions be satisfied, for instance U exists and RN is a continuous function of time, (5) and (1) are equivalent [19]. The relation 4 means that as soon as some strictly contacting pressure is exerted, the right velocity is lying in the tangent plane to the contact point. It is a reasonable contacting behaviour introduced here to facilitate the unilateral formulation using the velocity as a primary variable.

When deformable bodies are contacting, unilateral contact may be expressed using the Signorini condition (1) or the almost equivalent velocity Signorini condition (5). When rigid bodies are contacting, the Signorini condition does not give enough information. A *shock law* has to be adopted. During an impact, complex local deformability phenomena are occurring at a time and space scale which are ignored either because they are

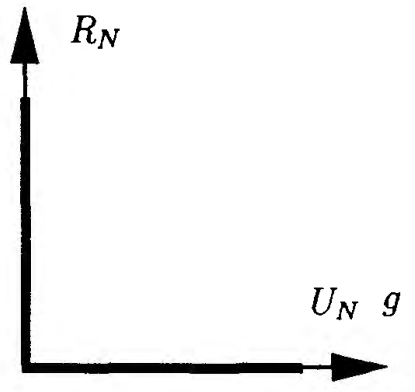

Fig. 2. The Signorini graph.

beyond the scientist's knowledge, or because he or she deliberately chooses to ignore them, neglecting the contact duration. Experiments or mathematical analysis on simple spring mass models show that in some circumstances, the velocity after the impact, the velocity before the impact and the impulse are related by some simple relation, namely a shock law, whatever are the local phenomena occurring during the duration of the impact. For instance, the Newton restitution law, $U_N^+(t) = -eU_N^-(t)$ happens to be true between billiard balls. It is easy to verify that the velocity Signorini condition (5) implies the Newton restitution law with $e = 0$, i.e. the *inelastic shock law*. Since this paper is mainly devoted to deformable bodies or inelastic rigid bodies, shock laws will not be discussed any further (see [19,6]).

Some authors [8,13,5], assuming right and left limits for the second derivative $\ddot{q}$, propose relations to be satisfied by these limits, derived from unilateral conditions. They are indeed numerically useful when second-order numerical schemes for the dynamical equation are used and when the solution behaves as assumed. When a large number of simultaneous impacts are occurring, specially between rigid bodies, the solution is not expected to have such a regularity. In the contact Dynamics method, any use of $\ddot{q}$ is avoided.

2.2. Coulomb's law

The basic features of the Coulomb dry friction are: the friction force lies in Coulomb's cone: $\|R_T\| \leq \mu R_N$, μ friction coefficient, if the sliding velocity is different from zero, the friction force is opposed to the sliding velocity with magnitude μR_N: $U_T^+ \neq 0 \Rightarrow R_T = \mu R_N U_T^+ / \|U_T^+\|$. These two conditions may be summarized under the form of a maximum dissipation principle:

$$R_T \in \mathrm{D}(\mu R_N) \quad \forall S \in \mathrm{D}(\mu R_N) \quad (S - R_T)\, U_T^+ \geq 0\,, \tag{6}$$

where $\mathrm{D}(R_N)$ is the section of Coulomb's cone, the disk with center 0 and radius μR_N. The above relation is equivalent to,

$$R_T = proj_{\mathrm{D}(\mu R_N)}(R_T - \rho U_T^+)\,, \tag{7}$$

where ρ is some strictly positive number. Notice that the right limit U_T^+ of the sliding velocity has been used in the formulation, which is a choice motivated by numerical experiments. The above formulation accounts only for friction. A simple way to introduce *tangential shocks* proposed by Moreau, is to replace U_T^+ by some linear combination of U_T^+ and U_T^-. Distinction between static and dynamic friction coefficient may be introduced and also cohesive friction laws, which results actually in mere changes of variables. The graph of Coulomb's law is displayed in Fig. 3. This is a monotone multi-mapping graph.

2.3. Kinematic relations

Since the paper is devoted to numerical methods, the systems under consideration are either collections of rigid bodies or discrete models of deformable bodies. Thus, the configuration of the medium is given by some

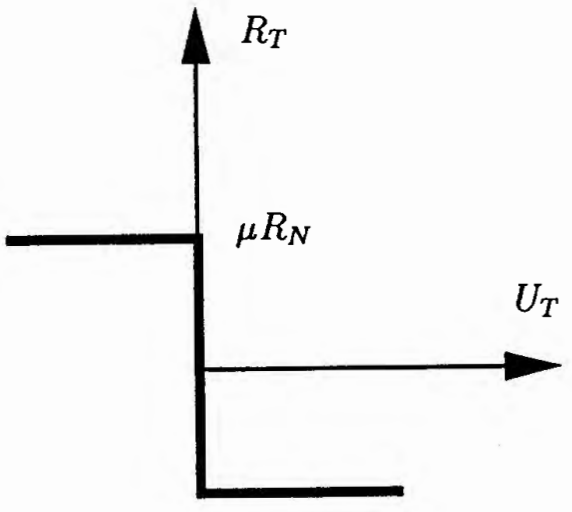

Fig. 3. Coulomb's graph.

variable q, for instance, the assembled vector of coordinates of the centers of gravity and rotational components of rigid bodies, or the assembled vector of the node displacements in a finite element description of deformable bodies. The time derivative is denoted $\dot{q}$ a function of time with bounded variations. Superscripts $\alpha\beta$ are used to denote some candidates to contact, $\alpha, \beta \in 1, \ldots, \chi$. It comes from classical kinematic analysis that the relative velocity at some contact α is given by a relation,

$$U^{\alpha} = H^{*\alpha}(q)\, \dot{q} \ , \tag{8}$$

and from duality considerations (conservation of the power expressed with local variables or generalized variables), the representative r^{α} of the local reaction force R^{α} for the generalized variables system satisfy the relation,

$$r^{\alpha} = H^{\alpha}(q)\, R^{\alpha} \ . \tag{9}$$

The mappings $H^{\alpha}(q)$, $H^{*\alpha}(q)$, are linear and $H^{*\alpha}(q)$ is the transposed mapping of $H^{\alpha}(q)$. For instance, in the case of a node candidate to contact with a line between two other nodes, the mapping H involves changes of variables from the local frame to the general frame and interpolation between velocities of the three nodes. The following relation expressing that the normal component of the relative velocity is the time derivative of the gap function is a key relation as far as discrete forms of unilateral conditions are to be defined,

$$\dot{g}^{\alpha} = U^{\alpha}_{\mathrm{N}} \ . \tag{10}$$

2.4. Discretizing continuous media

Discrete systems are considered in this paper. When a continuous medium is considered, described by some displacement function $u(t,x), x \in \Omega$ the domain occupied by the medium, similar kinematic formulas may be written at any contact point ξ. They involve: $\dot{u}(t, x)$ instead of $\dot{q}$; $U(t, \xi)$ the relative velocity at the point ξ; the density of impulse and density of reaction force $\sigma(t, \xi)$ instead of R^{α}; the representative of $\sigma(t, \xi)$ instead of r^{α}. Similarly, unilateral conditions and frictional laws are written using the above variables. According to the kind of finite element used to discretize the continuous medium, the discrete forms of the Signorini condition or Coulomb law are more or less complicated involving weight functions. When q is the assembled vector of nodal displacements, an usual approximation is to 'concentrate frictional contact forces on nodes', these nodes being considered as material particles. It may be proven that when linear interpolation is used between nodes, the consistent approximation, in the sense of virtual power formulation, results in concentrating frictional contact forces on nodes [10]. A very complete study has been done in this line using the virtual power method by Laursen and Simo [28]. It is not within the scope of this paper to give more details.

3. The dynamical equation

For smooth motions the dynamical equation is

$$M\ddot{q} = F(q, \dot{q}, t) + r \ ,$$

where $F(q, \dot{q}, t)$ represents internal and external forces. Since shocks are expected, the derivatives in the above equation are to be understood in the sense of distributions. It is more convenient to write this equation as a measure differential equation,

$$M\, \mathrm{d}\dot{q} = F(q, \dot{q}, t)\, \mathrm{d}t + r\, \mathrm{d}\nu \ , \tag{11}$$

where $\mathrm{d}t$ is the Lebesgue measure, $\mathrm{d}\dot{q}$ is a differential measure representing the acceleration measure, $\mathrm{d}\nu$ is a non-negative real measure relative to which $\mathrm{d}\dot{q}$ happens to possess a density function, and r is the representative of local density of impulses exerted when contact occurs. It may be written in the following equivalent form: for all t, τ, in $]0, T]$,

$$\begin{cases} M(\dot{q}(t) - \dot{q}(\tau)) = \int_{\tau}^{t} F(t, q, \dot{q})\, \mathrm{d}s + \int_{]\tau,t]} r\, \mathrm{d}\nu\,, \\ q(t) = q(\tau) + \int_{\tau}^{t} \dot{q}\, \mathrm{d}s\,, \end{cases} \tag{12}$$

where $\mathrm{d}s$ represents the Lebesgue measure.

4. Time discretization

When time discretization is performed, an elementary subinterval $]t_i, t_{i+1}]$ of length h is considered. The main idea developed in the process of time discretization is that discrete variables are not necessarily to be defined at some special time belonging to this interval, since anyway the impact times are usually unknown or costly to approximate or even difficult to separate when simultaneous contacts are occurring. Discrete frictional contact relations are thus defined 'over' the interval of time $]t_i, t_{i+1}]$.

4.1. Discrete form of the dynamical equation

Integrating both sides of the dynamical equation yields

$$\begin{cases} M(\dot{q}(t_{i+1}) - \dot{q}(t_i)) = \int_{t_i}^{t_{i+1}} F(t, q, \dot{q})\, \mathrm{d}s + \int_{]t_i,t_{i+1}]} r\, \mathrm{d}\nu\,, \\ q(t_{i+1}) = q(t_i) + \int_{t_i}^{t_{i+1}} \dot{q}\, \mathrm{d}s\,. \end{cases} \tag{13}$$

The *mean value impulse* denoted $r(i+1)$,

$$r(i+1) = \frac{1}{h} \int_{]t_i,t_{i+1}]} r\, \mathrm{d}\nu\,, \tag{14}$$

emerges as a primary unknown. A numerical scheme is defined by a choice of approximate expressions for the two other integrals in (13). Since discontinuous velocities are expected, the use of higher order approximations is unnecessary and even troublesome. First-order schemes are well enough when many shocks and contacts are expected. These numerical schemes involve $\dot{q}(i)$, $q(i)$, $\dot{q}(i+1)$, $q(i+1)$, the approximations of $\dot{q}(t_i)$, $q(t_i)$, $\dot{q}(t_{i+1})$, $q(t_{i+1})$, respectively. In order to avoid technical details in this section, numerical schemes for the dynamical equation will be presented in a section further on. To introduce the Contact Dynamics method, it is just suitable to write the numerical scheme as a linear relation between the approximate values $\dot{q}(i+1)$ and $r(i+1)$ together with a linear relation allowing the construction of $q(i+1)$ from $q(i)$, $\dot{q}(i)$, $\dot{q}(i+1)$, here a simple interpolation formula,

$$\begin{cases} \dot{q}(i+1) = v_{\mathrm{free}} + whr(i+1)\,, \\ q(i+1) = q(i) + \theta h\dot{q}(i+1) + (1-\theta)h\dot{q}(i)\,, \end{cases} \tag{15}$$

where θ is some real parameter, w is the inverse of some mass matrix, v_{free} is a term representing the value of $\dot{q}(i+1)$ when frictional contact is not taken into account, i.e. when $hr(i+1)$ is set to vanish. It is constructed with values of internal forces, external forces, known from $q(i)$, $\dot{q}(i)$. To summarize, the primary unknowns suggested by the discrete form of the dynamical equation are the approximation of the velocity $\dot{q}(i+1)$, and the impulse $hr(i+1)$.

4.2. Discrete kinematic formulas

According to the above relation (8), the following approximate relative velocities $U^{\alpha}(i+1)$, $U^{\alpha}(i)$, are defined as

$$U^{\alpha}(i+1) = H^{*\alpha}(\bar{q})\,\dot{q}(i+1)\,, \qquad U^{\alpha}(i) = H^{*\alpha}(\bar{q})\,\dot{q}(i)\,, \tag{16}$$

where $\bar{q}$ is some auxiliary intermediate value of q. Similarly, the following impulses are introduced

$$R^{\alpha}(i+1) = \frac{1}{h}\int_{]t_i,t_{i+1}[} R^{\alpha}\,\mathrm{d}\nu\,, \qquad r^{\alpha}(i+1) = H^{\alpha}(\bar{q})\,R^{\alpha}(i+1)\,. \tag{17}$$

The construction formula for $q(i+1)$ together with the kinematic relation (10) suggests a predictive formula for the approximate gap,

$$g^{\alpha}(i+1) = g^{\alpha}(i) + \theta h U_N^{\alpha}(i+1) + (1-\theta)h U_N^{\alpha}(i)\,. \tag{18}$$

Actually, the convenient approximate gaps are

$$\bar{g}^{\alpha}(i+1) = g^{\alpha}(i+1) + (1-\theta)h U_N^{\alpha}(i+1)\,, \tag{19}$$

$$\bar{g}^{\alpha}(i) = g^{\alpha}(i) + (1-\theta)h U_N^{\alpha}(i)\,, \tag{20}$$

with the *predictive relation* which will be commented on in Section 6.

$$\bar{g}^{\alpha}(i+1) = \bar{g}^{\alpha}(i) + h U_N^{\alpha}(i+1)\,, \tag{21}$$

4.3. Discrete frictional contact relations

The proposed discrete forms of the Signorini condition (1) and of the velocity Signorini condition (5) are commented on in Section 6,

$$\bar{g}(i+1)\geq 0 \qquad R_N(i+1)\geq 0 \qquad \bar{g}(i+1)R_N(i+1)=0\,, \tag{22}$$

if some contact is forecast within the interval $[i, i+1]$ then

$$U_N(i+1)\geq 0 \qquad R_N(i+1)\geq 0 \qquad U_N(i+1)R_N(i+1)=0\,. \tag{23}$$

The relation 'if some contact is forecast within the interval $[i, i+1]$' means that some test is being performed when beginning or while computing the unknowns U, R, for the current time step. Using the predictive formula (21) the Signorini condition may be written as well,

$$\begin{aligned} &\frac{\bar{g}^{\alpha}(i)}{h} + U_N^{\alpha}(i+1)\geq 0\,,\\ &R_N^{\alpha}(i+1)\geq 0\,,\\ &\left(\frac{\bar{g}^{\alpha}(i)}{h} + U_N^{\alpha}(i+1)\right)R_N^{\alpha}(i+1)=0\,. \end{aligned} \tag{24}$$

As far as inelastic shocks are concerned, the following discretization formula, proposed by the author, so called *quasi-inelastic shock* owes both to (22) and (24),

$$\begin{aligned} &\frac{\bar{g}^{\alpha}(i)^{\mathrm{pos}}}{h} + U_N^{\alpha}(i+1)\geq 0\,,\\ &R_N^{\alpha}(i+1)\geq 0\,,\\ &\left(\frac{\bar{g}^{\alpha}(i)^{\mathrm{pos}}}{h} + U_N^{\alpha}(i+1)\right) R_N^{\alpha}(i+1)=0\,. \end{aligned} \tag{25}$$

where $\bar{g}^{\alpha}(i)^{\mathrm{pos}} = \max(0, \bar{g}^{\alpha}(i))$, is the positive part of $\bar{g}^{\alpha}(i)$. With such a discrete formulation, if a shock is to occur, it will be numerically accomplished within two time steps; at the first step, taking into account the gap, some normal velocity is found so as to establish the contact, i.e. a null gap; at the second step, the gap being null, the inelastic shock law is activated. When the gap is negative, the inelastic shock law is activated only. Coulomb's law, being positively homogeneous, is compatible with the quasi-inelastic shock law, in the sense that computing a shock within two time steps gives the same cumulated impulse as if it were computed in a single step. It may be easily proven, when the local mass matrix, relating impulsive reactions and relative

velocities at the contacting point, is spherical, i.e. is a diagonal matrix when expressed in the local frame. Then, tangential and normal responses are uncoupled. It is the case when contactors are spherical grains or material points. In the general case, it may be wrong. It is also known that the concept of shock law when the local mass matrix is not spherical might not be relevant and yield energy violations. Thus, the quasi-inelastic shock law might be considered as much relevant or irrelevant as any other shock law. The quasi-inelastic shock discretization formula is easy to handle since forecasting possible contacts is avoided. It has proven to behave nicely for highly dense granular material simulations.

The proposed discrete form for Coulomb's law is:

$$R_T^\alpha(i+1) \in \mathrm{D}(\mu R_N^\alpha(i+1)) \,, \quad \forall\, S \in \mathrm{D}(\mu R_N^\alpha(i+1)) \quad (S - R_T^\alpha(i+1))U_T^\alpha(i+1) \geq 0 \,, \tag{26}$$

A different way of discretizing Coulomb's cone has been proposed by Klarbring [15], used also by Stewart and Trinkle [32] for rigid body contacts. Coulomb's cone is approximated by an intersection of hyperplanes, so that selecting the active hyperplane allows to write Coulomb's friction law as a complementary condition.

Any of these unilateral conditions, (22), (24), (25), together with Coulomb's law (26) may be shortly referred to as

$$\mathrm{SignCoul}(i, U^\alpha(i+1), R^\alpha(i+1)) \,. \tag{27}$$

The index i stands for a data: this is $\bar{g}^\alpha(i)$, evaluation of the gap at the step before the current time step when the Signorini condition is chosen; this is the forecast status when the velocity Signorini condition is chosen. Notice that the mapping $(U, R) \to \mathrm{SignCoul}(i, U, R)$ is a continuous piecewise affine mapping in the 2 dimensional case, and a continuous raywise mapping in the 3 dimensional case, see relations (2), (7).

It has already been mentioned that when one of the contactors is deformable the Signorini condition (22), is to be used, while when both contactors are rigid the velocity Signorini condition (24) or the quasi-inelastic shock law (25) are to be used. The Signorini condition (22) has the advantage to make the unilateral contact selfcorrective, in the sense, that if some penetration is occurring, the condition prescribes the gap to vanish at the next time step. So a penetration results in some correcting impulse causing some slight positive relative velocity, while inelastic shock demands null relative velocity. The Signorini condition is used by Stewart and Trinkle, [32] for contact between rigid bodies. It happens in large collections of rigid bodies (typically 1000 grains), that if, at some time step, at some contacting point, some non negligible penetration is occurring, then the correcting impulse due to the Signorini condition might not be negligible and acts as an artefact source of energy well up to destroy the realistic dynamical behaviour. The Signorini condition is thus not recommended with large collections of rigid bodies. More comments about computational errors will be found in Section 8.

5. Solving the basic frictional contact problem

Using the kinematic relations, a linear equation relating relative velocities and mean values of the impulses may be derived from the discrete form of the linearized equation (15),

$$U^\alpha(i+1) = U^\alpha_{\mathrm{free}} + \sum_\beta W^{\alpha\beta} h R^\beta(i+1) \,, \tag{28}$$

$$W^{\alpha\beta} = H^{*\alpha}(\bar{q}) w H^\beta(\bar{q}) \,, \tag{29}$$

$$U^\alpha_{\mathrm{free}} = H^{*\alpha}(\bar{q}) v_{\mathrm{free}} \,. \tag{30}$$

This equation has to be written together with frictional contact relations.

5.1. The basic frictional problem for the candidate α

The index $(i+1)$ is omitted here for simplicity's sake. The basic frictional problem for the candidate α is

$$\begin{cases} U^\alpha = U^\alpha_{\mathrm{locfree}} + W^{\alpha\alpha} h R \,, \\ \mathrm{SignCoul}(i, U^\alpha, R^\alpha) \,. \end{cases} \tag{31}$$

with

$$U^{\alpha}_{\text{locfree}} = U^{\alpha}_{\text{free}} + \sum_{\beta \neq \alpha} W^{\alpha\beta} hR^{\beta} \ , \tag{32}$$

The set of equations (31) is thus composed of a linearized form of the dynamical equation giving the relative velocity of the candidate α as an affine function of its impulse, together with the frictional contact relations. The data are

- the approximations $\tilde{q}$, $\dot{q}(i)$,
- the gap $\bar{g}^{\alpha}(i)$ at the step before the current time step obtained by geometric computations from the configuration $\tilde{q}$ while updating the local frames,
- provisional values of the impulses on other candidates R^{β}, $\beta \neq \alpha$, giving the term $U^{\alpha}_{\text{locfree}}$.

The unknowns are

- the relative velocity U^{α} and the impulse R^{α}.

It is thus a problem with four unknowns in the 2 dimensional case, and six unknowns in the 3 dimensional case. These unknowns are the components of U^{α}, R^{α}.

5.2. Solving the 2D basic frictional contact problem for the candidate α

As already mentioned, the frictional contact relation (27) may be written as a set of two equations involving the proj mapping, see (2), (7), i.e. involving piecewise continuous affine functions in the 2-dimensional case. The solution of problem (31) is obtained by seeking for the intersection of piecewise hyperplanes, or as well, of a hyperplane with the Signorini Coulomb graphs. The solution may be straightforwardly computed. It is known and has been mentioned in the literature by Alart and Klarbring [14]. The subscript α is omitted for simplicity's sake. One sets

$$W = \begin{pmatrix} W_{TT} & W_{TN} \\ W_{NT} & W_{NN} \end{pmatrix},$$

$$U_{\mathrm{f}} = U_{\text{locfree}} \quad , \quad J_{\mathrm{f}} = W^{-1} U_{\mathrm{f}} \quad , \quad J_{NN} = -\frac{U_{\mathrm{f}}}{W_{NN}} \quad , \quad a_{NT} = -\frac{W_{NT}}{W_{NN}} \ .$$

Assuming $-1 < \mu a_{NT} < 1$, (33)

the 2D basic solution is,

if $U_{\mathrm{f}N} > 0$ then: *no contact*

$hR_T = 0$, $hR_N = 0$,

if $U_{\mathrm{f}N} \leq 0$ and $J_{\mathrm{f}T} + \mu J_{\mathrm{f}N} > 0$ then: *forward frictional contact*

$$hR_T = -\mu hR_N \ , \qquad hR_N = \frac{J_{NN}}{1 + \mu a_{NT}} \ ;$$

if $U_{\mathrm{f}N} \leq 0$ and $J_{\mathrm{f}T} - \mu J_{\mathrm{f}N} < 0$ then: *backward frictional contact* (34)

$$hR_T = \mu hR_N \ , \qquad hR_N = \frac{J_{NN}}{1 - \mu a_{NT}} \ ;$$

if $U_{\mathrm{f}N} \leq 0$ and $J_{\mathrm{f}T} + \mu J_{\mathrm{f}N} \leq 0$ and $J_{\mathrm{f}T} - \mu J_{\mathrm{f}N} \geq 0$ then: *sticking*

$hR_T = -f_T$, $hR_N = -f_N$.

Condition (33) is mentioned under different forms in the literature. When it is not satisfied, the solution is not unique. When contact between deformable bodies is under consideration, the masses being concentrated on nodes, if the time step is small enough, the extra diagonal terms of W are overtaken by the diagonal ones so that

the condition (33) is always satisfied. When contact between rigid bodies is under consideration, the extra diagonal terms might be large enough so that (33) is not satisfied. In these circumstances reference is made to Painlevé paradox. The fact is that mere rigid bodies models are not fine enough to ensure uniqueness.

5.3. Solving the 3D basic frictional contact problem for the candidate α

In the 3 dimensional case, when the matrix W has spherical symmetry properties, which is the case for instance when contact is occurring between balls, formulas as above may be exhibited. In the general case it is not possible to write down explicit formulas. Adopting the forms (2), (7), the basic 3 dimensional system may be written

$$\begin{cases} U^\alpha - U^\alpha_{\text{locfree}} - W^{\alpha\alpha} hR = 0 \,, \\ R^\alpha_N - \text{proj}_{R^+}(R^\alpha_N - \rho U^\alpha_N) = 0 \,, \\ R^\alpha_T - \text{proj}_{D(\mu R^\alpha_N)}(R^\alpha_T - \rho U^\alpha_T) = 0 \,, \end{cases}$$

i.e. some system of the kind,

$$\Phi(U^\alpha, R^\alpha) = 0 \,,$$

where Φ is some continuous raywise mapping. This system of six unknowns is solved using a generalized Newton method. The solution is obtained with a few iterations, usually 1 or 2 or 3.

Using a hyperplane approximation of Coulomb's cone, the 2D or 3D basic α problem is a complementary system which may be solved by Lemke's techniques [32].

5.4. Solving the frictional problem for the collection of candidates

The algorithm is the following: for a candidate α, assume provisional values of, U^β, R^β, known from the current iteration for $\beta < \alpha$ and known from the previous iteration for $\beta > \alpha$; applying (34) or the 3D basic α solution compute U^α, R^α; proceed to the next candidate; run over the list of candidates until satisfactory convergence.

do iter = 1, itermax
  do $\alpha = 1$, χ
    compute $U^\alpha_{\text{locfree}}$ from stored R^β, $\beta \neq \alpha$
    compute U^α, R^α, using (34) or 3D basic solution
    store U^α, R^α,
  end do
if satisfactory, exit
end do

This loop will be referred to as the *Signorini Coulomb loop*. It is quite similar to a nonlinear block Gauss Seidel algorithm, finding U^α, R^α as the result of an elementary nonlinear four or six unknowns system, [9]. Further on it will be seen how this loop may be nested or not within other loops. Referring to substructuration techniques, it may be said that the above problem has been concentrated on candidates for contact.

The generalized Newton method is used by Alart and Curnier [22,1] to solve the whole system with unknown the vector q, U, R. In doing so, a large hollow nonsymmetrix matrix has to be stored and a huge linear system has to be solved. In the case of large collections of bodies, storage problems would occur using this method. Complementarity techniques, mathematical programming, Lemke's techniques are used by Klarbring [15] and Chabrand et al [26]. More comments will be found in Section 9.2.

The question of the convergence of the Contact Dynamics method is not adressed in this paper. The method has proven in applications to behave nicely as soon as the time step is small enough and the friction coefficient is lying in the range allowed by (34). It is mathematically proven in [9] that the convergence is ensured for the dynamical frictional problem of a discrete system of particles with some elastoplastic internal forces (such as a

finite element model of elastoplastic shell) provided the time step is small enough. The proof is based on a theorem by Alart in [3], giving sufficient conditions for a piecewise linear mapping to be a homeomorphism. Regarding existence theorems of frictional contact problems equations, there are many contributions in the literature for regularized problems. There are fewer results about the non-smooth Signorini Coulomb problem. In [24], Jean and Pratt give sufficient conditions to ensure the existence of a solution without loss of contact for a rigid body unilateral Coulomb frictional problem. Pioneering results for the unilateral problem of a single particle with inelastic shock have been given by Monteiro Marques. Generalisations of this problem by the same author are to be found in [17]. Foundations for mathematical developments in the line of differential inclusions may be found in Moreau [19]. In a recent paper by Stewart [30], the question of existence of solutions to rigid dynamics and Painlevé paradox is revisited with new results.

6. More about time discretization

6.1. Discrete kinematic formulas

The approximate relative velocities, $U^\alpha(i+1)$, $U^\alpha(i)$, have been defined as

$$U^\alpha(i+1) = H^{*\alpha}(\bar{q})\dot{q}(i+1)\,,$$

$$U^\alpha(i) = H^{*\alpha}(\bar{q})\dot{q}(i)\,,$$

where $\bar{q}$ is some auxiliary intermediate value of q. Possible choices are $q(i)$, $q(i) + h/2\dot{q}(i)$, or some element $q(k+1)$ of an approximating sequence of the unknown $q(i+1)$, etc. In fact when changes of q within a time step are small with respect to curvatures of contacting surfaces, the normal vector changes are also small and choices of $\bar{q}$ are not very important. This gives the opportunity to save time when computing local frames and gaps, which is usually a costly operation. The following impulses have been introduced

$$R^\alpha(i+1) = \frac{1}{h}\int_{]t_i,t_{i+1}[} R^\alpha \,\mathrm{d}\nu\,,$$

$$r^\alpha(i+1) = H^\alpha(\bar{q})R^\alpha(i+1)\,,$$

so that $r^\alpha(i+1)$ appears as the image of the mean value impulse R^α by $H^\alpha(\bar{q})$. This is not the mean value of $t \to r^\alpha(t) = H^\alpha(q(t))\,R^\alpha(t)$, except if $t \to H^\alpha(q(t))$ is a constant mapping. It is so when one of the contactors is a rigid plane.

6.2. Discretizing the frictional contact relations

The contact index α is omitted in this section for simplicity's sake. As already mentioned the mean value of the impulse appears as a primary unknown:

$$R(i+1) = \frac{1}{h}\int_{]t_i,t_{i+1}[} R \,\mathrm{d}\nu\,.$$

The approximation $U(i+1)$ is used as a primary discrete unknown to play the part of U^+ in relations (1), (5), (26), the approximation $U(i)$ playing the part of U^-. The approximation of the gap $\bar{g}$ to be taken into account within the time interval is a crucial choice. The idea of *consistency of the gap approximation with unilateral condition* may be a guide for this choice. Consistency means that the following relation is satisfied.

$$\bar{g}(i) = 0 \quad \text{and} \quad \bar{g}(i+1) = 0 \quad \Rightarrow \quad U_N(i+1) = 0\,. \tag{35}$$

where $\bar{g}(i)$ is some evaluation of the gap related to $]t_{i-1}, t_i]$ and $\bar{g}(i+1)$ is some evaluation of the gap related to $]t_i, t_{i+1}]$. This means that when some contact situation is supposed to occur within two consecutive intervals, the right normal relative velocity should vanish. A possible classical choice is $\bar{g} = g$. In this case, taking into account the kinematic formula (18), the consistency relation is satisfied only if $\theta = 1$. Another choice is

$$\bar{g}(i+1) = g(i+1) + (1-\theta)\, hU_N(i+1)\,, \qquad \bar{g}(i) = g(i) + (1-\theta)\, hU_N(i)\,.$$

The kinematic formula 4.6 yields the predictive formula,

$$\bar{g}(i+1) = \bar{g}(i) + hU_N(i+1)\,,$$

so that with this definition of the approximate gap, the consistency condition is satisfied. It may be shown on numerical examples that, when consistency is not satisfied, some oscillatory artefacts with step period are generated, due to the fact that there is some inadequacy between the direction of the velocity and the value of the gap which should be zero in contact, see [12,25]. It may be seen that $\bar{g}(i)$ is approximately the gap corresponding to the configuration $\bar{q}(i) = q(i) + (1-\theta)h\dot{q}(i)$, and $\bar{g}(i+1)$ is approximately the gap corresponding to the configuration $\bar{q}(i) = q(i+1) + (1-\theta)h\dot{q}(i+1)$. It is known in dynamical numerical schemes that there might be some advantage to compute $\dot{q}$ and q at different times of the time step. As far as unilateral contact is concerned, the intermediate values, $\bar{q}(i)$, $\bar{q}(i+1)$, are a possible choice. In particular $\bar{q}(i)$ is a good choice for $\bar{q}$ in the kinematic formulas. The above choices, sensible for mechanical reasons, have been validated by many numerical elementary experiments. To summarize, (1), (5), (26), may be written,

$$\bar{g}(i+1) \geq 0 \qquad R_N(i+1) \geq 0 \qquad \bar{g}(i+1)\, R_N(i+1) = 0\,,$$

if some contact is forecast within the interval $[i, i+1]$ then

$$U_N(i+1) \geq 0 \qquad R_N(i+1) \geq 0 \qquad U_N(i+1)\, R_N(i+1) = 0\,.$$

The relation 'if some contact is forecast within the interval $[i, i+1]$' means that some test is being performed when beginning or while performing computations of the unknowns, U, R, for the current time step. It may be some test of the kind,

a contact is forecast if

$\bar{g}(i) \geq 0$, or ,

$\bar{g}(i) + hU_N(i) \geq 0$, or ,

any prediction of the gap $\bar{g}$ is negative.

The choice of the contact prediction test is not a crucial one. What actually is crucial is the consistency between gap approximation and unilateral condition. The quasi-inelastic shock law may be used between rigid bodies avoiding forcasting.

7. Some numerical schemes for the dynamical equation

Two typical classes of problems are presented here which will allow to discuss special features when dealing with frictional contact: the case of small perturbations where the system is governed by a constant coefficient linear dynamical equation; the case of static or dynamic large deformations.

7.1. Small perturbations

The theta method scheme is presented here; see also other methods in this line [25,12],

$$\begin{cases} \displaystyle\int_{t_i}^{t_i+1} F(q(s), \dot{q}(s), s)\ \mathrm{d}s \approx h\theta F(q(t_{i+1}), \dot{q}(t_{i+1}), t_{i+1}) + h(1-\theta)F(q(t_i), \dot{q}(t_i), t_i)\,, \\ q(i+1) = q(i) + h\theta\dot{q}(i+1) + h(1-\theta)\dot{q}(i)\,. \end{cases}$$

The value $\theta = 1$, yields the implicit Euler method. The theta method is unconditionally stable for $\theta \geq \frac{1}{2}$ as far as constant linear systems are concerned, i.e. when

$$F(q, \dot{q}, t) = -V\dot{q} - Kq + P(t)\,,$$

where V is the damping matrix and K is the stiffness matrix. Setting $\dot{q}(i)$, $q(i)$, $\dot{q}(i+1)$, $q(i+1)$, approximations of $\dot{q}(t_i)$, $q(t_i)$, $\dot{q}(t_{i+1})$, $q(t_{i+1})$, respectively, the theta method scheme is

$$M(\dot{q}(i+1) - \dot{q}(i)) = h\theta(-V\dot{q}(i+1) - Kq(i+1) + P(t_{i+1})) + h(1-\theta)(-V\dot{q}(i) - Kq(i) + P(t_i)) + hr(i+1) .$$

For consistency reasons it may be useful to introduce the auxiliary displacements,

$$\bar{q}(i+1) = q(i+1) + h(1-\theta)\dot{q}(i+1) , \qquad \bar{q}(i) = q(i) + h(1-\theta)\dot{q}(i) .$$

When $\theta = 1$, the variable $\bar{q}$ is identically q. The theta-method is written as

$$\begin{cases} \dot{q}(i+1) - \dot{q}(i) = w(-hV\dot{q}(i) - hK(\bar{q}(i) + (2\theta-1)h\dot{q}(i)) + hP(i+1) + hr(i+1)) , \\ \bar{q}(i+1) - \bar{q}(i) = h\dot{q}(i+1) , \\ \text{where} \\ w = (M + h\theta V + h^2\theta^2 K)^{-1} , \\ P(i+1) = \theta P(t_{i+1}) + (1-\theta)P(t_i) . \end{cases} \tag{36}$$

It is assumed that the matrix $M + h\theta V + h^2\theta^2 K$ is invertible, an assumption which will be discussed later on.

7.2. Finite deformations

For simplicity's sake F is assumed to have the form:

$$F(q, \dot{q}) + P(t) .$$

The implicit Euler method is

$$\begin{cases} M(\dot{q}(i+1) - \dot{q}(i)) = hF(q(i+1), \dot{q}(i+1)) + P(t_{i+1}) + hr(i+1) , \\ q(i+1) - q(i) = h\dot{q}(i+1) . \end{cases}$$

The unknowns $q(i+1)$, $\dot{q}(i+1)$, $hr(i+1)$, are sought as limits of sequences $q(i+1, k+1)$, $\dot{q}(i+1, k+1)$, $hr(i+1, k+1)$, (for simplicity's sake the index $i+1$ shall be omitted when referring to these approximating sequences of $q(i+1)$, $\dot{q}(i+1)$):

$$\begin{cases} M(\dot{q}(k+1) - \dot{q}(i)) = hF(q(k+1), \dot{q}(k+1)) + hP(t_{i+1}) + hr(k+1) , \\ q(k+1) - q(i) = h\dot{q}(k+1) . \end{cases}$$

Using a first-order expansion of F with respect to q, $\dot{q}$, as done in the Newton–Raphson method:

$$F(q(k+1), \dot{q}(k+1)) = F(q(k), \dot{q}(k)) - K(k)(q(k+1) - q(k)) - V(k)(\dot{q}(k+1) - \dot{q}(k)) ,$$

where

$$K(k) = -\frac{\partial F}{\partial q}(q(k), \dot{q}(k)) ,$$

$$V(k) = -\frac{\partial F}{\partial \dot{q}}(q(k), \dot{q}(k)) .$$

The numerical scheme is

$$\begin{cases} \dot{q}(k+1) - \dot{q}(k) = w(k)(-M(\dot{q}(k) - \dot{q}(i)) + hF(q(k), \dot{q}(k)) + hP(t_{i+1}) + hr(k+1)) , \\ q(k+1) - q(k) = h\dot{q}(k+1) , \\ \text{where} \\ w(k) = (M + hV(k) + h^2K(k))^{-1} \end{cases} \tag{37}$$

The term $-M(\dot{q}(k) - \dot{q}(i))$ is an approximation of the acceleration impulse.

7.3. The linearized dynamical equation

The numerical schemes (36) and (37) have some similarities:

(1) In the linear case, initializing with values $\dot{q}(i)$, $q(i) + h\dot{q}(i)$, when performing a single iteration, the same approximate values $\dot{q}(i+1)$, $q(i+1)$, yielded by the theta method (36) with $\theta = 1$ are found.

(2) If $\theta = 1$, $M = 0$, $V = 0$, and if K is positive definite, the theta method degenerates into a standard linear resolution for the quasi-static case.

(3) If $M = 0$, $V(k) = 0$, and if $K(k)$ is a positive definite matrix, the Euler–Newton–Raphson scheme (37) degenerates into the standard classical Newton–Raphson method for the quasi-static case.

(4) When given boundary conditions do not prevent rigid motions (translations or rotations), several cases have to be considered. In any case the mass matrix M is positive definite. $V = 0$, $K = 0$, is the case of rigid bodies collections. If V, K, are not positive definite (they are nevertheless usually positive), the matrix $M + h\theta V + h^2\theta^2 K$ is positive definite for sufficiently small h and thus has an inverse.

(5) The 'correcting matrix' w in the Newton–Raphson scheme (37) may be replaced by any matrix $\bar{w}$ positive definite, for instance

$$\bar{w}(k) = M^{-1} , \tag{38}$$

$$\bar{w}(k) = M^{-1} - M^{-1}K(k)M^{-1} , \tag{39}$$

or other choices (see [2,12]).

(6) When the correcting matrix $\bar{w}(k) = M^{-1}$ is used, initializing with values $\dot{q}(i)$, $q(i) + h\dot{q}(i)$, performing a single iteration, the method degenerates into the classical Euler explicit scheme.

The use of these different possible schemes in the context of frictional problems will be discussed later on.

To summarize, the Euler–Newton–Raphson (37) may be viewed at some sub-iteration $k+1$ within a time step as a linear relation between the approximate time derivative $\dot{q}(k+1)$ and the mean value of the impulse $hr(k+1)$:

$$\dot{q}(k+1) = v_{\text{free}}(k) + W(k)hr(k+1) ,$$

where

$$v_{\text{free}}(k) = \dot{q}(k) + W(k)(-M(\dot{q}(k) - \dot{q}(i)) + hF(q(k), \dot{q}(k)) + hP(t_{i+1}) .$$

In the same way, the theta scheme (36) may be viewed at some time step $i+1$ as a linear relation between the approximate time derivative $\dot{q}(i+1)$ and the mean value of the impulse $hr(i+1)$ yielding a relation similar to the one above.

8. Controlling the error

In this section, the question of accuracy is addressed in the context of large dense collections of contacting bodies. The question of computational error or accuracy is difficult to handle whatever the considered problem. Indeed, the exact solution is usually unknown and the only information which might be exhibited is some distance to it. Upper and lower bounds of such a distance may be obtained using the mathematical convergence properties of the algorithm, which means that: (a) a unique solution has been proved to exist; (b) the convergence of the algorithm is established. Except in some quite peculiar cases, these circumstances are not realized. The choice of the distance is important. For instance, max distance imposes each variable to lie in the

prescribed accuracy interval, while the quadratic distance allows the well behaved variables to compensate other badly behaved variables. Asking high accuracy might be very time consuming in favor of a slight improvement of the results, while asking not asking enough accuracy might drastically deteriorate the results. What should be the accuracy range to be prescribed to obtain good enough results? Still at last what is meant by good enough results? The main results under consideration when studying a large collection of contactors are, resulting pressures, average stress tensors, localization of stresses and strains, distribution of reaction forces. These are macroscopic or statistical results. The experience and the common sense have taught the scientist not to expect much reproducibility of local data such as velocity, reaction force at some contact, at some given time t. Even the trajectory of some grain might wander under slight changes in the computational circumstances such as time step changes. Reproducibility would imply strong asymptotic stability properties. Nevertheless, it is expected that macroscopic and statistic properties are not sensitive to the computational circumstances. Here again, sensitivity is a personal matter of acceptance. The fact is that even if mathematical considerations are of some help, the numerical simulation has to be considered as a physical experiment and the criterion to qualify results are of the same kind that those used by physicists.

Another feature of frictional contact problems is that, except in the case where a unique solution exists, the obtained approximate solution yielded by some iterative scheme might be different according to the way iterations are monitored. In a dynamical system, the position and velocity at the end of a time step may be computed knowing positions and velocities at the beginning of the time step. When frictional contact is considered, the solution depends on the history of loading, more precisely on the frictional forces just before the time step. For instance, looking for some equilibrium state under a given load, the solution might be different acording to the fact that this state has to be reached while loading or unloading. In a similar way, when iterations are performed, some wandering provisional values of reaction forces may be obtained acting as some initial data, to be numerically decreased while mechanical loading is currently processed or to be numerically increased while mechanical unloading is processed. The results, though fulfilling the error criterion might not be satisfactory. Incidentally, as quoted above, the solution depending on the history of loading, more precisely on the frictional forces just before the time step, starting iterations with those values is not only a matter of saving time but is actually seeking for an approximate solution among those favored by the history.

It happens in large collections of contactors that, when the accuracy is sufficient, if at some step some contact is badly computed, it is most often re-made in the next few steps. It means that the behaviour is correct as far as mean time values are under consideration. The software LMGC uses a certain number of non classical numerical devices to control the accuracy, which are also used to accelerate the computation. It is not within the scope of this paper to give such technical details.

9. Applications

9.1. Granular materials

A Couette granular flow is presented here, as an example of collection of rigid bodies numerically simulated by the Contact Dynamics method. A sample of 2400 polydisperse rigid disks or rolls, so-called Schneebeli material, is kept within two drums. The outer drum is a membrane subject to a constant pressure 75 kPa, while the inner drum is rotating with a constant speed, 0.1 turn/s. The friction coefficient between grains is 0.5 and 0.75 between grains and drums. The numerical simulation shows chains of forces, Fig. 4, and some features observable, in Fig. 5, from the displacements of grains after a quarter of turn. The sample behaves quasi rigidly, except near the boundary of the inner drum, where a shear layer has developed inducing dilatancy. More details will be found in a forthcoming paper by Vardoulakis and Zervos. Other examples for bi-dimensional strain-stress experiments have been computed with 4000, 16 000 grains samples.

In the case of an assembly of rigid grains, the variable q is the vector the components of which are the coordinates of centers of gravity and of rotation vectors. In this case the internal forces, in the sense of continuous medium, are not to be written. The implicit Euler scheme ($\theta = 1$) and the quasi-inelastic shock law are used. The dynamical equation (36) is written,

$$\begin{cases} \dot{q}(i+1) - \dot{q}(i) = w(hP(i+1) + hr(i+1)) \,, \quad w = M^{-1} \,. \\ q(i+1) - q(i) = h\dot{q}(i+1) \,, \end{cases}$$

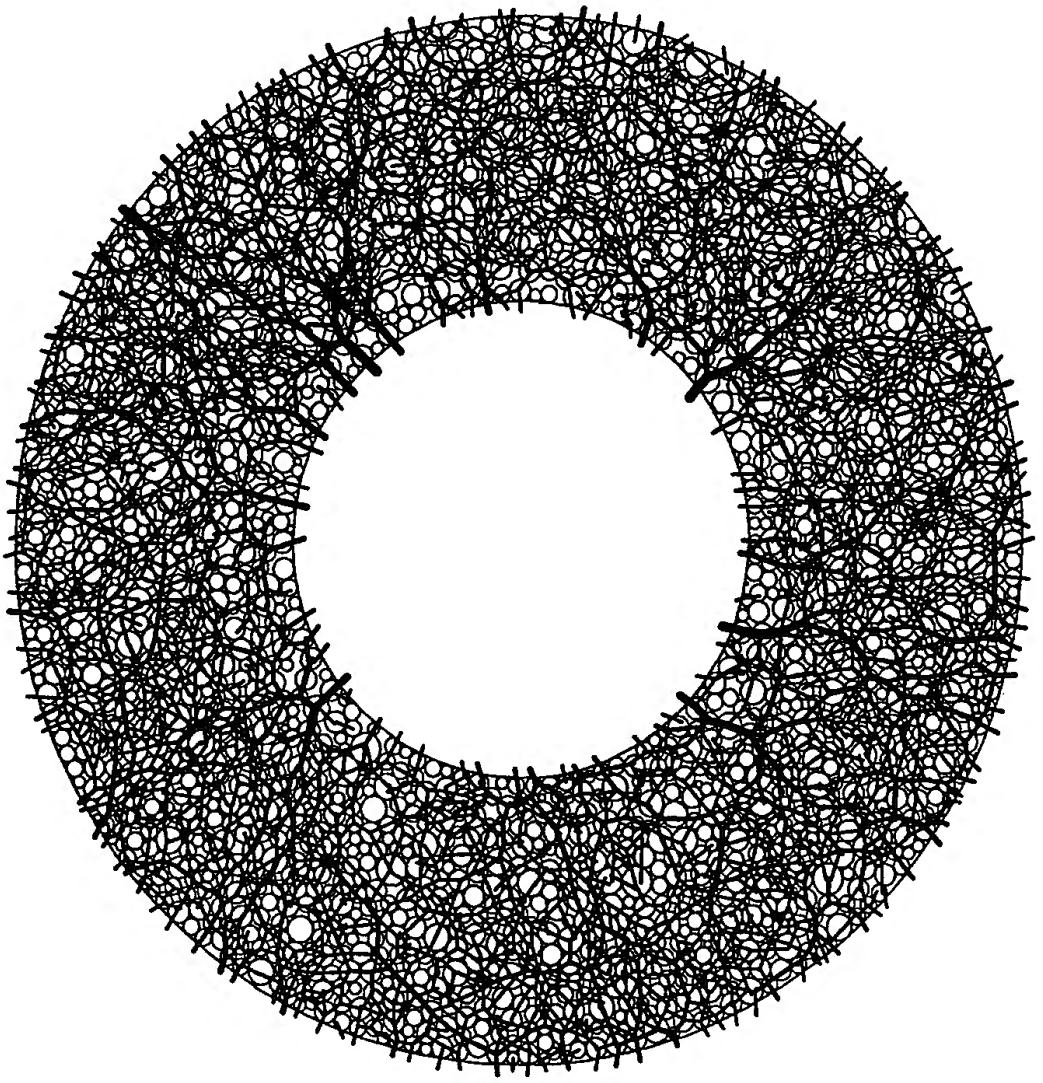

Fig. 4. Chain forces.

where w is a diagonal matrix. Due to coupling between grains, the matrix W is a matrix with size, $2\chi \times 2\chi$, or $3\chi \times 3\chi$ where χ is the number of candidates for contact. The band width depends on the connectivity between candidates. Except for a very small collection of grains, computing and storing such a matrix is avoided. Nevertheless, performing elementary operations such that, $v^\alpha = whr^\alpha$, $U^\alpha = H^{*\alpha}v^\alpha$, $r^\alpha = H^\alpha R^\alpha$, require a small number of floating point operations. Thus, solving the basic frictional contact problem for a candidate α is a quick elementary process. Once the impulses have been obtained after a sufficient number of iterations of the Signorini Coulomb loop, the values $\dot{q}(i+1)$, $q(i+1)$, are straightforwardly obtained through the dynamical equation. The number of iterations at each time step depends on the prescribed accuracy and on the mechanical situation. Kinematic constraints between rigid grains might prevent them moving while fully satisfying the frictional contact laws. From a mathematical point of view the existence of a solution is not guaranteed in all circumstances. Nevertheless, it may be satisfactory to ensure frictional contact relations up to a certain accuracy, allowing thus some ease between grains. From numerical experiments in the 2 dimensional case, it is found that a severe accuracy of 0.15% needs an average number of iterations, lying between $\sqrt{\chi}$ and χ. It is to be understood that the whole list of candidates for contact is being processed once during an iteration. The compaction of the sample took 10 h CPU time and 21 h CPU time for 1/4 turn shearing on IBM SP2. This time may be reduced by asking less accuracy and by omitting floating point instructions for any shape grains which are useless when grains are disks or spherical.

9.2. Deep drawing simulation

A typical example of large deformations and large displacements frictional problem is presented here as an application. When deep drawing, a metal sheet is being held with dry friction in between blank holders and punched within the matrix. The modelling of frictional forces is very important since these forces are monitoring

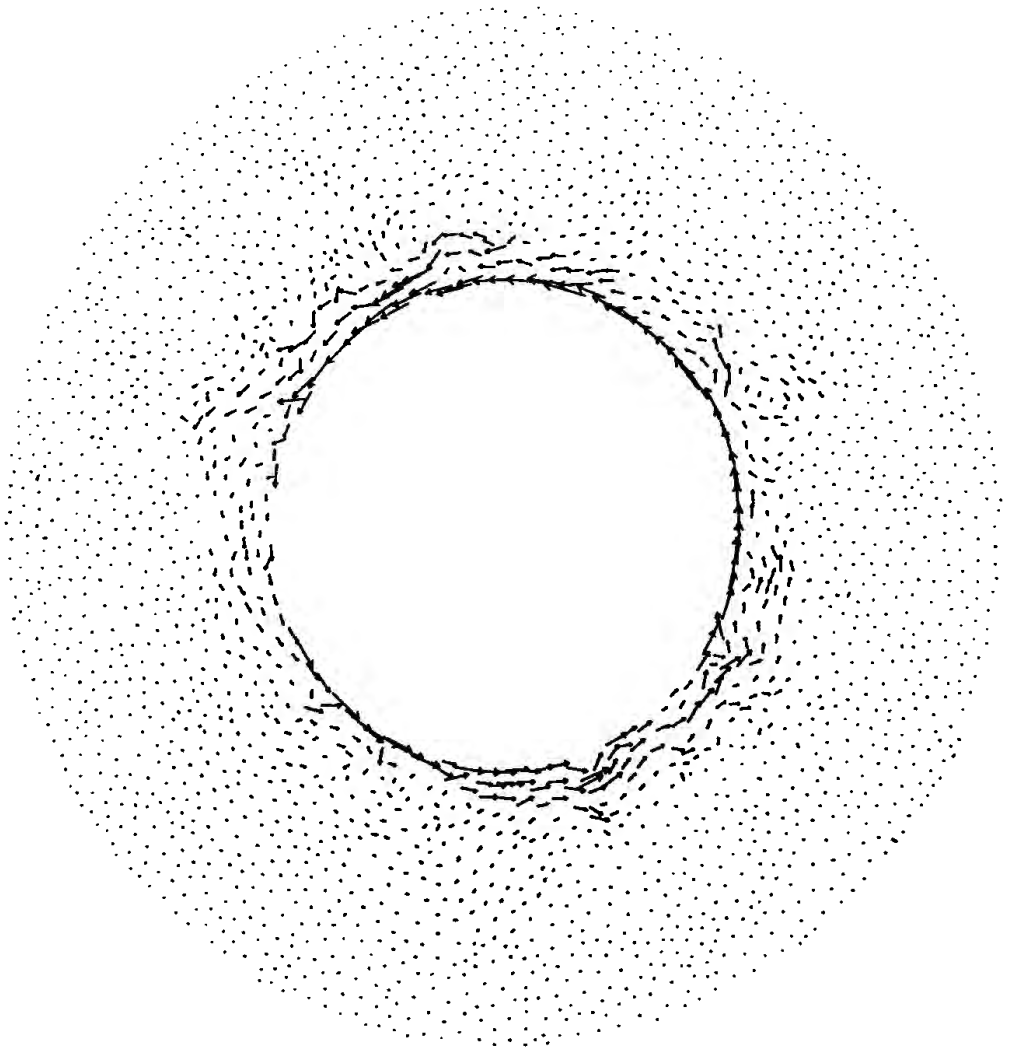

Fig. 5. Velocity field.

the deep drawing process. In this application, with 1800 elements, most of the nodes are candidates for contact. Some nodes are sliding while others are sticking. The Contact Dynamics method has been implemented in Simem3, a software developed by Renault Automotive Industry and his partners. Details are to be found in [31,2]. The elasto-plastic shell finite element model has been developed by Gélin et al. [7]. Coulomb's dry friction law is commonly accepted with friction coefficient in the range 0.15, 0.25. In the quasi-static approach, a scheme such that (37) is used, with $M = 0$, $V = 0$, neglecting acceleration terms. In this case, the scheme is exactly the classical Newton–Raphson scheme. It happens at some stage that due to plastic localisations and wrinkles the tangent matrix $K(k)$ is ill conditioned and that the Newton–Raphson algorithm fails to converge. When dynamics are taken into consideration, with mass concentrated on nodes, with sufficiently small time step h, the correcting matrix $M + h^2K(k)$ is invertible and the conditionning is improved so that the computation may

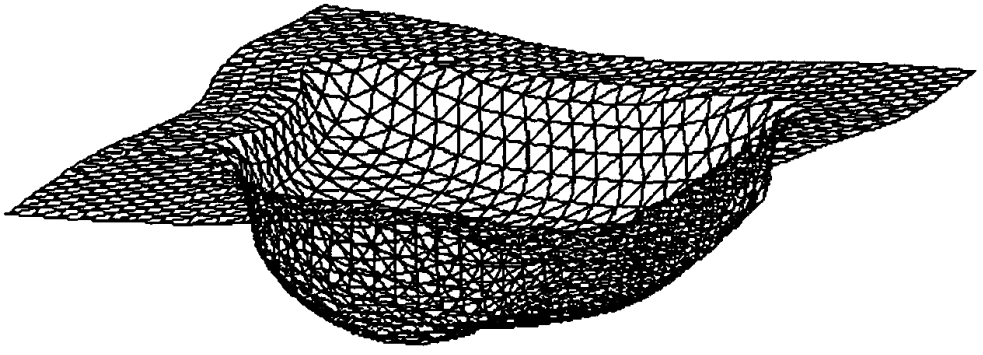

Fig. 6. Twingo cup.

be processed successfully. Using a fully implicit algorithm as suggested in (37) means that reaction forces are computed in the so-called Signorini Coulomb loop, nested within a so called Newton–Raphson k-indexed loop where internal and external forces and the tangent matrix $K(k)$ are actualized:

time step loop
for each time step i do

Newton–Raphson loop
for each k compute $M + h^2K(k)$, internal and external forces, $v_{\text{free}}(k)$,

Signorini Coulomb loop
for each candidate α compute U^α, R^α,
Signorini Coulomb loop end

compute $\dot{q}(k)$, $q(k)$,
Newton–Raphson loop end

time step loop end

Roughly put as above, the algorithm, though converging nicely is much time consuming. Indeed, computing $v_{\text{free}}(k)$ needs solving a linear system with matrix $M + h^2K(k)$, and computing $\dot{q}(k)$, $q(k)$, needs solving again the linear system with modified second member. Furthermore, the inner Signorini Coulomb iteration is also time consuming since the matrix $W(k)$ has to be computed first and stored; it is the price to pay for concentrating the problem on nodes candidates for contact. It may be worthwhile in some applications. Improvements are:

(1) In small perturbation problems, $K(k)$ is approximated by some K, w and W are computed and stored once for all.
(2) If the time step h is sufficiently small, the mass matrix M dominates $h^2K(k)$ and it is not necessary to update the tangent matrix at any subiteration k using the evaluation $M + h^2K(1)$ at the first iteration $k = 1$. Furthermore, the matrix $w(1) = (M + h^2K(1))^{-1}$ being sufficiently close to a diagonal matrix, so is W and the number of iterations in the Signorini Coulomb loop may be drastically reduced.
(3) The time step h might be sufficiently small so as $w = M^{-1}$ to be a good enough approximation of $M + h^2K(k)$. In these circumstances the values U^α, R^α, are obtained performing a single iteration (nodes are disconnected), and only a few Newton–Raphson iterations are necessary. This is one among several schemes used for the above deep drawing application.
(4) It might happen that a single Newton–Raphson iteration is sufficient. Then, the algorithm may be identified to an explicit scheme, though, the way reaction impulses are computed is purely implicit.
(5) The implicit approach makes springback effects easy to compute.

9.3. Buildings made of blocks

Buildings made of blocks are the object of growing interest in ancient architecture as well as in civil engineering. Much attention has been paid to motions, global deformations and possible appearance of cracks, i.e. openings of joints. The distribution of stresses within and between blocks is more questionable. It is expected, and actually numerically observed, that the results depend very much on the modelling of blocks, on the way loading is applied and on the kind of algorithm used for the numerical simulation. This particular feature is mainly due to Coulomb's frictional contact which allows an infinity of equilibrium solutions. Physical situations are alike. The selected state depends on the history of loading. The choice of an algorithm and the way it is monitored is also, in some way, part of the history of loading.

Software, as Udec or Tridec, initiated by Cundall [4,29], in the same line as Trubal for granular materials, may compute the motion of such buildings, under severe loadings. In these softwares, blocks are rigid, and contact forces are exerted through springs, dampers and 'dry-dampers' in between blocks at their vertices. Moreau, using Contact Dynamics, has computed buildings made of polygonal rigid blocks, with Coulomb's law

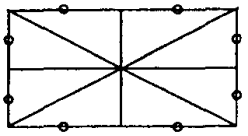

Fig. 7. Elementary block 8 T3 and candidates for contact.

and the inelastic shock law expressed at the vertices [21]. When dynamical effects such that oscillations of slender structures under gust of winds or ground waves are to be taken into account, blocks must be considered to be deformable as well as mortar joints if any. Is the distribution of stresses much influenced by the elastic deformability in quasi-static situations? For instance, consider a wall made of rigid rectangular blocks, put together without filling in with mortar, built on a perfectly rigid horizontal plane, under gravity load. The stresses between head joints may vanish, while, as it will be shown in the example below, Poisson effect makes elastic blocks enlarge, and press each other at head joints.

The algorithm used to compute the reaction forces in this application has two particular features. Firstly, the problem may be set assuming small perturbations, so that item 1 in Section 9.2 applies. Secondly, since the number of nodes of each block is very small, and since blocks are independent, the matrix K is a block diagonal matrix, with small band width. The inverse matrix $w = (M + h^2K)^{-1}$ is also a block diagonal matrix with small band width. It results that the elementary operations

$$v^\alpha = w\, hr^\alpha\,, \qquad U^\alpha = H^{*\alpha} v^\alpha\,, \qquad r^\alpha = H^\alpha R^\alpha\,,$$

involve a small number of floating point operations. The fully implicit algorithm 9.2 is thus reasonably time consuming and actually behaves like the main Signorini Coulomb loop, as in granular materials.

In the examples below 144 (10 cm × 5 cm) elastic blocks, $\rho = 500$ Kg/m^3, $E = 10^6$ Pa, $\nu = 0.3$, are lying on some perfectly rigid foundation. They are coarsely meshed as shown in the pictures, using eight T3 elements for each block, see Fig. 7. Some four T3 elements meshes have also been used. In the 3 dimensional case blocks have been meshed with eight H8 elements, or more coarsely with one single H8 element. Masses are concentrated on nodes. Frictional contact forces are concentrated on special mid points on lines joining nodes, see Fig. 7. The friction coefficient is 0.5 between blocks and blocks and foundation. The gravity load is applied to the unconstrained wall at some initial time, so that a gravity wave propagates. Results are stored at time T when the resulting force on the foundation equals the weight of the wall. At this time acceleration terms are small so that the obtained reaction forces are possible candidates for the static equilibrium. Indeed, using these forces together with the obtained deformed state as initial conditions and setting null velocities, one observes vanishing small oscillations. Fig. 8 shows the distribution of reaction forces at time T, the direction being shown by thick lines, thickness and length being proportional to the magnitude. Fig. 9 shows principal stresses of the moment stress tensor on each block. The moment stress tensor of a block is defined at equilibrium as,

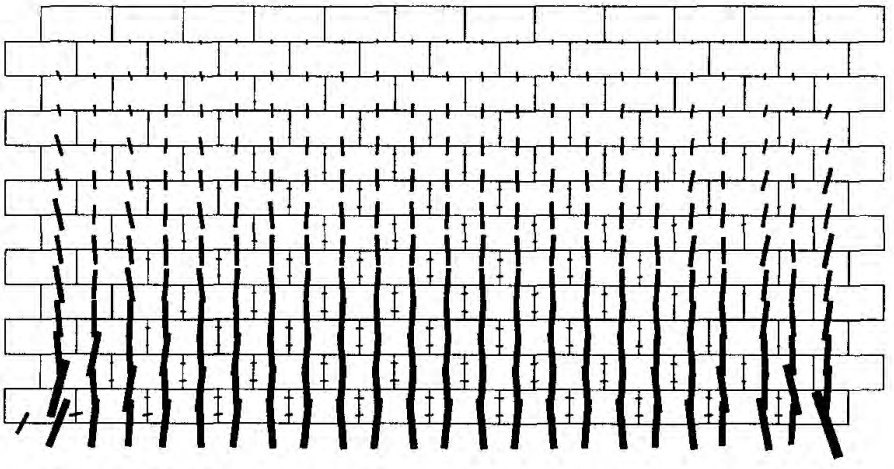

Fig. 8. Distribution of reactions between blocks.

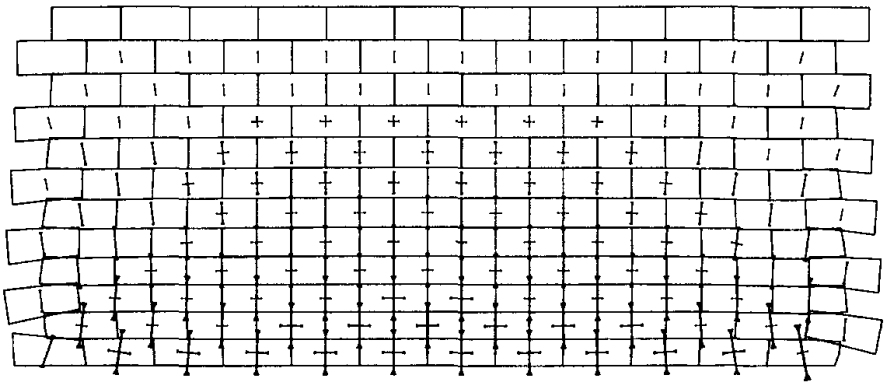

Fig. 9. Distribution of principal stresses (deformed configuration magnified 500 times).

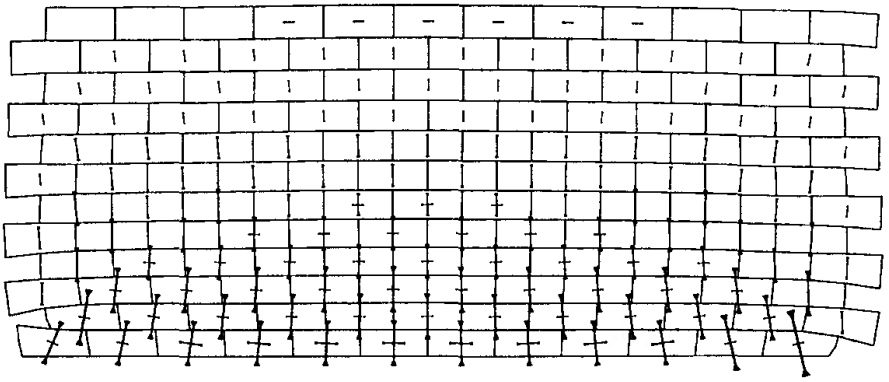

Fig. 10. Distribution of principal stresses in the single piece solid wall (deformed configuration magnified 500 times).

$M_{ij} = \Sigma_\alpha x_i^\alpha r_j^\alpha$, where the summation is performed for all reactions r^α forces exerted on the block and the vector x^α is OM^α, M^α being the contact point where r^α is exerted. This moment tensor is practically equal to the sum of the Cauchy stress tensors on the block. The deformed configuration is also represented magnified 500 times. It may be compared with the tensor field obtained for a single piece solid wall, with the same distribution of T3 elements, except that elements are connected through sharing nodes, Fig. 10. In this body, tensile stresses are appearing, emphasized in Fig. 11, not seen in the blocks wall model 9. Taking for initial state the T state with null velocities, as explained above, small oscillations are observed with two different modes in the vertical and horizontal direction. After a dozen oscillations the distribution of reactions has changed and contact status with the foundation has moved from sticking to sliding, allowing relaxation of the reaction forces between headjoints, see Figs. 12 and 13. The deformed configuration is not magnified in Fig. 13. Two top blocks are beginning to fall. The fact that under oscillations, head joints begin to open, mainly at lower block layers, has been observed on a small collection of 10 finely meshed blocks, and also on a wall made of rigid blocks on a shaking ground, simulated by Moreau [21].

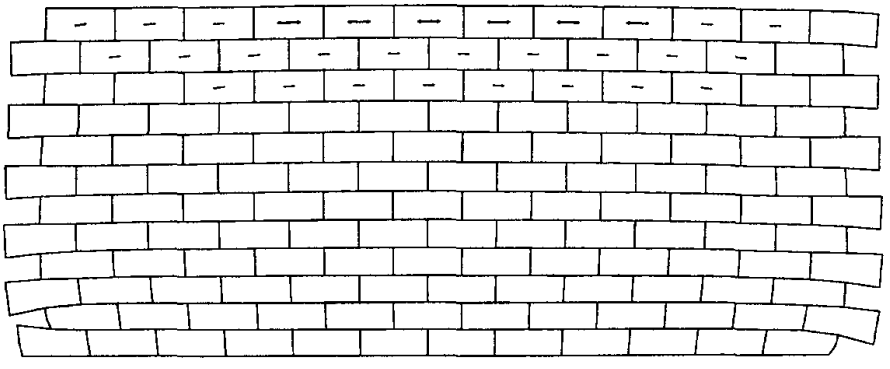

Fig. 11. Tensile stresses for the solid single piece solid wall.

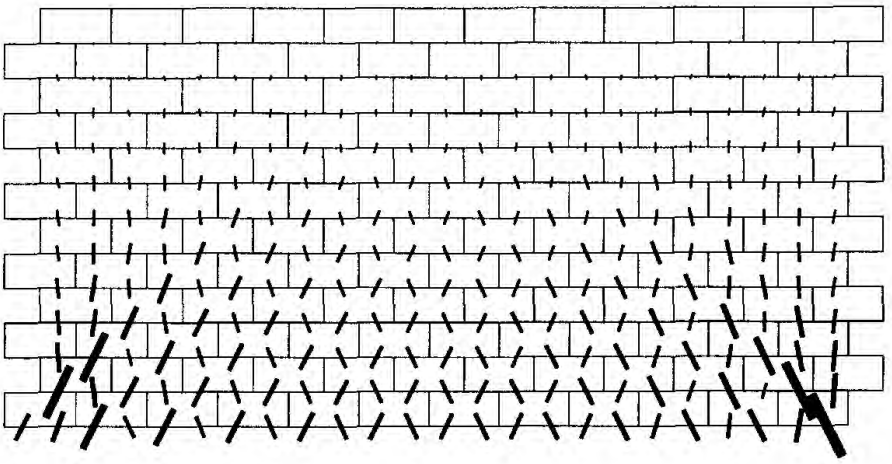

Fig. 12. Distribution of reaction forces a dozen oscillations later.

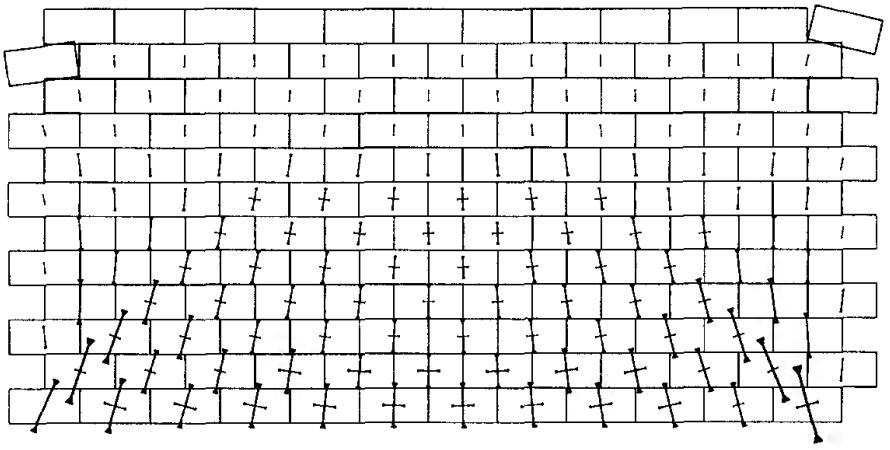

Fig. 13. Distribution of principal stresses a dozen oscillations later.

References

[1] P. Alart and A. Curnier, A generalized newton method for contact problems with friction. J. de Mécanique Théorique et Appliquée, supplément nos. 1 to 7 (1988) 67–82.

[2] F. Jourdan, M. Jean and P. Alart, An alternative method between implicit and explicit schemes devoted to frictional contact problems in deep drawing simulation, J. Mater. Process. Technol. 80–81 (1998) 257–262.

[3] P. Alart, Critère d'injectivité et de surjectivité pour certaines applications de $R \times n$ dans lui-même: application à la mécanique du contact. RAIRO, Modélisation Mathématique et Analyse Numérique, 2(27) (1993) 203–222.

[4] P. Cundall, A computer model for simulating progressive large scale movements of blocky rock systems, in: Proc. of the Symposium of the International Society of Rock Mechanics 1 (1971) 132–150.

[5] C. Glocker and F. Pfeiffer, Multibody Dynamics with Unilateral Contacts. (John Wiley & Sons, 1996).

[6] M. Frémond, Collisions of rigid bodies, in: M. Raous, M. Jean and J.J. Moreau, eds., Contact Mechanics, Proc. 2nd CMIS, (Plenum, 1995) 397–404.

[7] J.C. Boisse, J.C. Daniel and J.C. Gélin, A C^0 three nodes shell elements for non linear structure analysis, Int. J. Numer. Methods Engrg. 37 (1994).

[8] T.J.R. Hughes, The Finite Element Method: Linear Static and Dynamic Finite Element Analysis (Prentice-Hall, Inc., Englewood Cliffs, NJ 07632, 1987).

[9] F. Jourdan, P. Alart and M. Jean, A Gauss Seidel like algorithm to solve frictional contact problems, Comput. Methods in Appl. Mech. Engrg (155) (1998) 31–47.

[10] M. Jean, Unilateral contact and dry friction: time and space variables discretization, Arch. Mech., Warszawa 40(1) (1988) 677–691.

[11] M. Jean, in: A.P.S. Salvadurai and J.M. Boulon, eds., Frictional Contact in Rigid or Deformable Bodies: Numerical Simulation of Geomaterials, Elsevier Science Publisher, Amsterdam (1995) 463–486.

[12] M. Wronski and M. Jean, Some computational aspects of structural dynamics with frictional contact, in: M. Raous, M. Jean and J.J. Moreau, eds., Contact Mechanics, Proc. 2nd CMIS (Plenum, 1995) 137–144.

[13] T.J.R. Hughes, R.L. Taylor, J.L. Sackman, A. Curnier and W. Kanoknukulchai, A finite element method for a class of contact–impact problems. Comput. Methods Appl. Mech. Engrg. 8(1) (1976) 249–276.

[14] A. Klarbring, Examples of non-uniqueness and non-existence of solutions to quasi-static problems with friction and varying contact surface. RAIRO, Modélisation Mathématique et Analyse Numérique, 30(2) (1988) 1185–1198.

[15] A. Klarbring, Derivation and analysis of rate boundary-value problems of frictional contact. European J. Mech. A. (Solids), 9 (1990) 53–85.
[16] G. Duvaut and J.L. Lions, Les inéquations en Mécanique et en Physique (Dunod, Paris, 1972).
[17] M.P.D. Monteiro Marques, Differential inclusions in nonsmooth mechanical problems: Shocks and dry friction, in: Progress in Nonlinear Differential Equations and their Applications (Birkhauser, Verlag, 1993).
[18] J.T. Oden and J.A.C. Martins, Models and computational methods for dynamic friction phenomena, Comp. Methods Appl. Mech. Engrg. 52 (1985) 527–634.
[19] J.J. Moreau, Unilateral contact and dry friction in finite freedom dynamics, in: J.J. Moreau and P.D. Panagiotopoulos, Vol. 302 of International Centre for Mechanical Sciences, Courses and Lectures (Springer, Vienna, 1988).
[20] M. Jean and J.J. Moreau, Dynamics of elastic or rigid bodies with frictional contact and numerical methods, in: R. Blanc, P. Suquet and M. Raous, eds., Publications du LMA (1991) 9–29.
[21] M. Jean and J.J. Moreau, Unilaterality and dry friction in the dynamics of rigid bodies collections, in: A. Curnier, ed., Proc. of Contact Mech. Int. Symp. (1992) 31–48.
[22] A. Curnier and P. Alart, A mixed formulation for frictional contact problems prone to Newton like solution method, Comput. Methods in Appl. Mech. Engrg. 92(3) (1991) 353–375.
[23] J.T. Oden and E.B. Pires, Non local and non linear friction law and variational principles for contact problems in elasticity, J. Appl. Mech. 50 (1983) 67–76.
[24] M. Jean and E. Pratt, A system of rigid bodies with dry friction, Int. J. Engrg. Sci. (1985) 497–513.
[25] D. Vola, E. Pratt, M. Jean and M. Raous, Consistent time discretization for dynamical frictional contact problems and complementarity techniques, Revue Européenne des Eléments Finis 7(1–3) (1998) 149–162.
[26] P. Chabrand, F. Dubois and M. Raous, Programmation mathématique pour le contact avec frottement et comparaison avec d'autres méthodes, in: Actes du 2ème Colloque National en Calcul des Structures (Hermes, 1995).
[27] F. Radjai, M. Jean, J.J. Moreau and S. Roux, Force distributions in dense two-dimensional granular systems, Phys. Rev. Lett. 77 (1996) 274.
[28] T.A. Laursen and J.C. Simo, A continuum-based finite element formulation for the implicit solution of multybody, large deformation frictional contact problems, Int. J. Numer. Method in Engrg. 36 (1993) 3451–3485.
[29] P.A. Cundall and O.D.L. Stack, A discrete numerical model for granular assemblies. Geotechnique, 29(1) (1979) 47–65.
[30] D.E. Stewart, Existence of solutions to rigid body dynamics and the paradoxes of painlevé. Compte Rendu de l'Académie des Sciences, Sér. 1(325) (1997) 689–693.
[31] M. Jean, F. Jourdan and B. Tathi, Numerical dynamics for the simulation of deep drawing, in: Proc. of IDDRG 1994, Handbooks on Theory and Engineering Applications of Computational Methods, May 16–17, 1994.
[32] D. Stewart and J.C. Trinkle, An implicit time-stepping scheme for rigid body dynamics with inelastic collisions and coulomb friction. Int. J. Numer. Method in Engrg. 39 (1996) 2673–2691.
[33] J.J. Moreau, Numerical aspects of the sweeping process, Comput. Methods Appl. Mech. Engrg. 177 (1999) 329–349.