Indetermination due to dry friction in multibody dynamics
Jean Jacques Moreau

To cite this version:
Jean Jacques Moreau. Indetermination due to dry friction in multibody dynamics. European Congress on Computational Methods in Applied Sciences and Engineering, Jul 2004, Jyväskylä, Finland. hal-01825251

HAL Id: hal-01825251
https://hal.archives-ouvertes.fr/hal-01825251
Submitted on 28 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Indetermination due to dry friction
in multibody dynamics

Jean Jacques MOREAU
Laboratoire de Mécanique et Génie Civil
case 048, Université Montpellier II, 34095 Montpellier Cedex 5, France
e-mail: moreau@lmgc.univ-montp2.fr

Abstract

Coulomb friction commonly generates problems with multiple solutions. After examining an elementary example, one proposes to include the values of contact forces in the description of each state of the investigated system. This turns out to be what popular time-stepping numerical techniques implicitly do. Specifically, the Contact Dynamics computational policy, whose principle is recalled, offers this as an option with the additional advantage of improving computation speed. Thanks to this option, the evolution of the considered systems is predicted with negligible ambiguity. Alternatively, Contact Dynamics techniques may be used to explore the plurality of the solutions to problems stated in the traditional way. In this line, two examples of the assessment of contact forces in granular assemblies are presented, with the set of solutions graphically displayed.

1 INTRODUCTION

1.1 Features of dry friction

Tribology teaches us the complexity of the phenomena which may take place when two bodies are in contact, should sliding occur or not. Anyway, the law of Coulomb provides an irreplaceable framework for the first approach to dry friction. In many industrial situations, the only quantitative information provided by technical documents about friction-based devices is communicated in terms of Coulomb friction coefficients, with possible distinction between sticking and sliding values.

That the motions calculated from such a description of dry friction frequently are mathematically nonsmooth only reflects everyday observation: in the absence of lubrication, mechanisms may exhibit chattering motions, while emitting creaking noise, or jam.

Discussions have taken place for a long time about the fact that, in the dynamics of systems involving Coulomb friction, the initial value problem, i.e. the prediction of the motion consequent to an instant at which the positions and the velocities of the system elements are given, may have several solution or no solution at all. In the eyes of a 19th. century scientist like Painlevé, fascinated by the then recent theory of differential equations and Laplace’s “determinism”, this was an unforgivable defect of the law of friction, even inducing him to reject more generally the very concept of a contact force. The disquieting expression of “Painlevé’s paradox” caused great harm to mechanical science since engineers who had, from another part, some reasons to be unsatisfied with the moderate quantitative precision of the Coulomb law, tended to view this as a hint to some underlying logical defect. The irrelevance of Painlevé’s standpoint was already pointed at in his time [19].
In the author's view, a physical model is nothing but a *format*, in the sense of Computing Science, in which one decides to record the available information concerning a certain physical situation, to treat this information and finally to communicate conclusions. Fundamentally, the information one may collect about some investigated situation is always incomplete, so the model through which this information is coded cannot be expected to generate exhaustive predictions. Also, as the situation evolves, the validity limits of the model may come to be overrun.

1.2 An elementary example

Figure 1 shows, in a vertical plane, the following system. Two fixed walls slightly converge upward. A rigid rod, subject to gravity, is inserted between them, with Coulomb friction at the two contact points and abandoned at zero velocity. If the configuration angles have suitable values, compared with the friction angles at contact points, the problem of finding the consequent motion visibly has two solutions: either the rod loses contact and falls freely or it remains wedged.

The quadrilateral area filled with light hachure is the intersection of the Coulomb cones drawn from the two contact points. The triangular part above $AB$ intersects the vertical of the center of gravity along a line segment. Any position of $I$ on this segment corresponds to values of the contact forces $R_A$ and $R_B$ compatible with Coulomb condition for no-sliding and equilibrating the bar weight. In systems affected with dry friction, one should be ready to meet such uncertainties.

Intuitively, the outcome depends of the “degree of wedging” of the rod. The natural measure of such a degree precisely consists in the values of the contact forces and it will be seen in the sequel that the computational methods which include these values in the description of each state of the investigated systems allow for practically unambiguous predictions.

A deeper insight into the situation may be expected from the use of a *richer model* than above. To this end, let us take into account a certain *longitudinal elasticity* of the rod. This involves the introduction of a “small” variable $s$, the total elongation counted from some reference state. As long as contacts are effective, the value of $s$ is geometrically connected with the displacements of points $A$ and $B$ on the respective walls, still treated as indeformable. A law of longitudinal elasticity of the rod should be given. Combined with the longitudinal compo-
nents of the distributed forces, in particular gravity forces, that the rod undergoes, this leads
to determining the stress distribution along $AB$ through a differential equation for which the
longitudinal components of $R_A$ and $R_B$ provide the endpoint data. For simplicity, assume $AB$
horizontal and all the distributed forces vertical. Then this elastic analysis simply results in
a monotone relationship between the variable $s$ and the (opposite to each other) longitudinal
components of $R_A$ and $R_B$.

This model allows one to understand the scenario of rod insertion. If, after the first contact,
increasing upward forces are applied, the evolution assumed quasi-static consists in points $A$ and
$B$ sliding upward. At every time, vectors $R_A$ and $R_B$ have directions imposed by Coulomb law
in sliding regime. Thanks to the quasi-equilibrium equations of the rod, these contact forces
are then known. Since the elastic relationship connects their longitudinal components to the
positions of $A$ and $B$, elasticity intervenes secondarily to connect the displacements of these
points to the actual value of the resultant of weight and pushing forces.

If, at some instant, one decides to make the vertical pushing forces decrease and eventually
take downward direction, a new episode of motion begins, still assumed quasi-static. Points
$A$ and $B$ remain fixed so that the elastic relationship, \textit{whatever it is}, makes the longitudinal
components of $R_A$ and $R_B$ constant. At each time, the values of these vectors result from the
quasi-equilibrium equations and the no-slide regime persists so long as both vectors lie in the
respective Coulomb cones. This allows one to determine the value that the resultant of the
downward applied forces should eventually reach to produce detachment.

This model has the merit of making understand the processes of insertion and detachment
but, to the lay observer it conveys no tractable information in the absence of a measuring device
for microdisplacements, commonly not at hand. Since the proper elasticity law of the rod turns
out to be immaterial, one may expects that some model simply ignoring this elasticity could
be able to reproduce the above scenario. In fact, the Contact Dynamics computation method,
described in the sequel, has been found to yield numerically a response of contact forces to the
pushing history which agrees with that precedes.

2 NUMERICAL TECHNIQUES

2.1 Computation in (unilateral) multibody dynamics

The numerical dynamics of collections of bodies treated as perfectly indeformable, subject to
the constraints of non-interpenetrability, with friction taken into account in the event of contact,
currently is an active domain of research. Applications include the dynamics of machines, in
particular robots, the dynamics of masonry works submitted to transient actions (earthquakes,
gusts of wind or impacts), animated computer graphics and numerical simulation in granular
mechanics. In all these domains, assuming the perfect indeformability of each part of the system
leads to efficient numerical procedures which, in many circumstances, satisfy the needs. Possibly
also, some variables are added in order to also account for a certain deformability of these parts
[15], without essentially changing the computational strategies.

The techniques used in unilateral multibody dynamics may be classified into the three fol-
lowing categories.

2.2 Event-driven methods

The methods so qualified (abbr. ED) are practical mainly when the investigated time-interval
equals the union of not too many subintervals, a priori unknown, over which the \textit{status} of the
various contacts remain unchanged, i.e. no collision which would create new contacts occurs, no
contact either gets loose, nor any critical situation needing a change in the analytical expression
of Coulomb law is met [14]. On each of these subintervals, the same numerical techniques, resting on high performance integrators, as in the investigation of systems with classical bilateral, possibly frictional, constraints may be used. As integration proceeds, the program has only to watch the evolution of some indicators. In particular, the contact forces will be calculated. If, after a certain instant, some of the values computed for these forces are found to have directions incompatible with the unilaterality of the non-interpenetrability constraints (neglecting adhesive, i.e. gluing effects), the program decides that the motion has to be calculated otherwise. But one should keep in mind that the contacts which get loose after the critical instant are not necessarily those for which an unfeasible contact force has just been evaluated [8]. A popular approach to such discussions consists of reducing them to complementarity problems, similar to what is commonly met in constrained optimization.

Other cases of changing status arise from Coulomb law. The most critical one is that of an episode during which a certain contact has been found not sliding. If after some instant, the continuation of such a motion would require a contact force not belonging to the Coulomb cone, the transition to an episode of different status needs to be computed. This transition may sometimes consist in contact breaking but, more usually, in the inception of sliding. The integrator has to start with a new assessment of contact forces and acceleration variables. The crucial observation is that the inceptive direction of the sliding velocity \( \mathbf{v} \) (direction involved in Coulomb law) is that of the time-derivative \( \dot{\mathbf{v}} \) of this vector, since initially \( \mathbf{v} = 0 \). Now \( \dot{\mathbf{v}} \) is analytically connected with the acceleration variables. The mathematical problem to be solved is similar to the “Problem(P)” met at each step of a Contact Dynamics computation, as it is sketched in the sequel [18]. Uniqueness of solution should not be expected in general.

Difficulties of different nature occur for calculating the new velocity in case the critical instant is that of a collision. Contact forces take very large values during a very short time-interval and only poor phenomenological information is usually available about such physical circumstances. Even in the simplest case, that of the collision of two otherwise free members of the system, the traditional coefficient of restitution is known to depend not only on the materials these bodies are made of, but also on their shapes and relative orientations at the collision locus [27]; only the case of spherical beads appears relatively comfortable [10]. Furthermore, if some of the colliding bodies belong to clusters of already contacting ones, percussional reactions should be expected at all existing contact points. Newton’s restitution concept should at least be extended so as to handle all these contacts collectively [23].

Though rigid body collisions currently make an active domain of research [4][5][28][9], computation has to rely on pragmatic rules whose validity should be checked in each type of application. Anyway, a theoretical framework may be found in [11][12] which secures the thermodynamic correctness of the empirical information one may introduce into calculation.

2.3 Smoothing methods

“Nonsmoothness” is the salient feature of the problems in view. In fact, after the set of the possible positions of the investigated system has been parametrized through an element \( q \) of \( \mathbb{R}^n \), the geometric restriction that the non-interpenetrability constraints impose on \( q \) are expressed by a set of inequalities. Hence, instead of running in a smooth submanifold as in traditional analytical dynamics, the point \( q \) is confined in a region of \( \mathbb{R}^n \) whose boundary consists of a lot of pieces of hypersurfaces (millions or billions of them in current applications to granular materials) joined along multidimensional edges: this is nonsmoothness in space. Furthermore, possible collisions are expected to induce velocity jumps: this is nonsmoothness in time. To end, the contact forces associated with the non-interpenetrability constraints are governed by highly irregular laws. These forces vanish as soon as the corresponding contacts break while, if contact holds, the commonly stipulated mechanical conditions do not express them as functions of \( q \). If,
in addition, Coulomb friction is taken into account this introduces some irregular relationships between contact forces and the sliding velocities. All this may be called *nonsmoothness in law*. In such a state of the affairs, a natural move is to replace, approximately, the nonsmooth governing relationships by some *regularized* ones, a technique also used in the mathematical study of these problems [24]. First the non-interpenetrability constraints will be replaced by some stiff repulsion laws which take effect as soon as two members of the system come close to each other. This automatically handles the possible collisions, as far as one considers them as “elastic”, while the dissipativity of collisions may be accounted for by adding some damping actions or also by using different repulsion laws in the episodes of approach and of separation. Similarly, frictional contact may be somewhat regularized through the introduction of local elastic microdeformation and of viscosity-like effects. The dynamics of the approximate system is then governed by differential equations with sufficient regularity to be handled through standard numerical techniques. The drawback is that the need of precision requires the use of very stiff approximating laws. Hence the time-stepping procedures applied, usually of the *explicit* type, must resort to very small step-length and possibly also have to enforce numerical stability by introducing artificial damping or artificially increasing inertia. When treating dynamical applications, the effect of such an artificial alteration of the mechanical data may blur the picture. Significant simulations of loose (collisional) flows of granular materials have been obtained in that way, but when compact collections of bodies are concerned (pieces of masonry or dense granulates) the method is mainly applied to *quasi-static evolutions* in which only a succession of equilibrium states is looked for. Dynamical computation is then used only as a way of attaining each of these equilibrium states, a strategy referred to as “Dynamic Relaxation”.

The pioneering work of P. Cundall [6] was precisely based on the regularization strategy, today implemented in the majority of commercial pieces of software intended to handle non-interpenetrability. Because such computation techniques are close to those applied in molecular simulations, they are commonly referred to as “Molecular Dynamics” methods (abbr. MD), specially in the domain of granular mechanics [29].

Anyway our point is that this approach rests on the introduction of local microdisplacements, to which the normal and tangential components of contact forces are connected. At the beginning of each time-step, these quantities are involved in computation, as the result of the antecedent one.

*In this way, MD methods implicitly comply with our recommendation of treating contact forces as part of the description of each state.*

### 2.4 Contact Dynamics

This is the technique (abbr. CD) used in preparing the examples shown in this communication. It originated from [21], where the unilateral contact, possibly frictional and/or collisional, between rigid bodies received a formulation in terms of elementary convex analysis which proves suitable for elaborating computational approaches. Mathematically, the resulting evolution problems are governed in smooth cases by *differential inclusions*. General information on the latter concept may be found in [3][7], but the need of treating also non-smooth evolutions calls for its extension to that of a *measure differential inclusion* [20].

A directing idea of the Contact Dynamics approach is that the main object of computation is the *velocity function* $t \mapsto u \in \mathbb{R}^n$. Time-stepping algorithms essentially have to determine the evolution of this function; the position function $t \mapsto q$ is only to be updated at each step through adequate integration. Possibly, some members of $u$ are “pseudo-parameters” such as the components of the spin vector of a solid, instead of the time-derivatives of position parameters. From the geometrical standpoint, one should observe that, in analytical dynamics, the position $q$ ranges in some differential manifold, while for each $q$ the possible velocities are elements of
the tangent space to the manifold at this point. The latter is a vector space relatively to which such concepts as linearity, convexity, etc. make sense, while no algebraic concept of this sort is generally available in the position manifold.

The inequalities that the non-interpenetrability of bodies imposes on the $q$ parameters are handled through velocity conditions.

Contact Dynamics procedures rest on drawing the balance of momentum of the investigated mechanical system over each time-step. No estimation of the acceleration is needed and the burden of calculating mathematically the curvatures of the involved surfaces is avoided. These curvatures are simply accounted for by the fact that, from one step to the other, the normal directions to the detected contacts vary.

At each time-step, there is to solve a highly nonlinear nonsmooth problem – call it Problem (P) – whose simultaneous unknowns are the update of the velocity variables (in the time-discretization setting, the increase of these variables reflects acceleration) and the values of the contact impulses at the detected contact points. In smooth motions, the latter equal the integrals of contact forces over the time-step, so that an estimation of contact forces is obtained by dividing impulses by the step length. These elements are required to satisfy, in discretized form, the equations of the dynamics of the whole system and the laws of frictional contact, including the non-interpenetrability conditions.

Several numerical techniques are available to solve Problem (P) (see e.g. [1]). The most ancient one, an iterative procedure à la Gauss-Seidel [23], serves the best the purpose of the present study. It consists in reviewing contacts cyclically, solving at each “pass” a single contact problem for which all other contact impulses are treated as known, after what the involved velocities and contact impulses in memory are updated. The cyclic review is continued until a chosen degree of approximation is attained.

The method results in time-stepping schemes which, at least in what concerns the velocity function, are of the implicit type. Numerical stability is thus secured, allowing one to use much larger time-steps than in MD methods. The price to pay is the computational cost of Problem (P). The nonlinear Gauss-Seidel iterative procedure needs to start from an initial guess of the contact impulses. In the numerical simulation of the motion of dense granulates, one considerably accelerates convergence by introducing as initial guess the contact impulses found at the antecedent time-step, for the contacts which were already in effect.

If this is done, CD computation happens to comply with the policy of treating contact forces as part of the description of each state of the system.

Apart from this special initialisation, the CD method presents the considerable interest of generating at will all the solutions to Problem (P), simply by changing the way the nonlinear Gauss-Seidel iterations are conducted. This is illustrated in Section 3 below.

A single computation step is needed to decide whether equilibrium in a given position is a possible motion and to obtain the possible values of contact forces.

Contact Dynamics algorithms are ready at each step to treat collisions on the same footing as persistent contacts and to handle collectively the impulsive responses of all contacts in presence. The paroxysms in contact forces improperly called “Painlevé’s paradox”[13] are also handled. But, of course, the same need of phenomenological information about collisional processes arises as in other computation methods.
3 INDETERMINATION IN THE EVALUATION OF CONTACT FORCES

3.1 Equilibrium of a granular sample under gravity

A two-dimensional computation has been conducted in order to build a collection of 28 circular grains contained in a fixed container and in equilibrium under gravity. This configuration happens to involve 56 contact points of the grains between each other or with the container. Friction coefficient was fixed at 0.3. The set of the components of contact forces makes a point \( r \) of \( \mathbb{R}^{2 \times 56} \). Coulomb law requires of this point to belong to a polyhedral cone \( C \), the product of the 56 local friction cones, angular regions in each of the 56 corresponding copies of \( \mathbb{R}^2 \), while the 84 equilibrium conditions of the 28 grains (three degrees of freedom each), with account of gravity, define an affine manifold \( A \) of dimension 28 in \( \mathbb{R}^{2 \times 56} \). For each contact, the contact force is an element of the corresponding copy of \( \mathbb{R}^2 \), namely the Cartesian projection of \( r \) onto this factor space. The set of its possible values thus equals the projection of \( C \cap A \), a polyhedral subset of \( \mathbb{R}^{2 \times 56} \).

As observed before, calculating the values of contact forces in such an equilibrium configuration is the matter of a single step of the CD algorithm. The plurality of solutions is reflected in the plurality of ways of conducting the Gauss-Seidel iterations. The latter need to start from an initial guess of the contact forces (actually appearing in computation through impulsive terms). If this initial guess is provided by random drawing from a domain estimated to contain all possible values, every solution has a chance to be reached since, as an extreme case, it could itself be the result of the random drawing. Another way of varying the Gauss-Seidel procedure is to fix the initial guess (for instance zero values at all contacts) but to change at random the ordering according to which the contacts are cyclically reviewed. If \( \kappa \) is the number of contacts, this method can yield at most \((\kappa - 1)!\) different solutions, but the following graphical display shows that this limitation does not prevent it from yielding a valid survey of the possibilities.

Figure 2 shows a detail of the piling. The calculation of contact forces is repeated as explained above and each result is represented by a dot.

Grain 1, on top, takes its bearing on two underlying ones. Its equilibrium under gravity is a standard problem, analogous to that of parag.1.2: in view of the contact points \( A \) and \( B \) lying on each side of the vertical drawn from the mass center, there exist an infinity of values of the contact forces \( R_A \) and \( R_B \) agreeing with equilibrium. In fact equilibrium conditions are preserved when opposite forces with direction \( AB \) are added to \( R_A \) and \( R_B \), so long as the latter remain in the respective friction cones. In this case, the projections of \( C \cap A \) onto the respective copies of \( \mathbb{R}^2 \) reduce to line segments visible on the figure. The case of grain 2 is similar.

Grain 3 possesses four contacts with grains which themselves are in situations of force indetermination. Contact forces in each of the points \( E \) and \( F \) are indetermined in polygonal regions whose grey rendering results from the exploration procedure described above. The computation program is repeatedly applied to the calculation of equilibrium forces, each time with different choices in the completion of the Gauss-Seidel algorithm. At each time, the values found for the contact forces in \( E \) and \( F \) are represented by dots. A surprise is to see in these grey areas which are two-dimensional projections of the multi-dimensional polyhedron \( C \cap A \), the ghosts of some edges of the latter. For reasons unexplained so far, the limit points of Gauss-Seidel iterations are a little denser in the neighbourhood of some faces of \( C \cap A \), hence in the neighbourhood of the edges which are their intersections.
The framed enlargements on Figure 2 correspond to two different modalities in Gauss-Seidel algorithm. On the right, the initial guess of iterations is chosen at random, so that the possibility is open of attaining every solution of the problem. On the left, the initial guess is always zero, but from one computation to the others, it is the order of the cyclic review of contacts which is changed at random. Both procedures yield the same polygonal areas, but with different visions of ghost edges.

3.2 Granular flow on a slope

On Figure 3 are shown some results from a two-dimensional model of dense gravity-driven granular flow.

The ground, presented horizontal on the drawing, actually is inclined 26.5° down to the right and made rugous by glued grains created by the same random generator of polygonal shapes as the flowing ones. In order to imitate in some way a layer of infinite extent, the standard trick of periodic boundary conditions is used: only grains with mass-center located in some rectangular
cell are the object of calculation. Grains close to the upstream and downstream edges of the cell can have contact interaction with virtual grains, images by translation of grains belonging to the cell. When a grain exits the cell by crossing one of these edges, it is replaced in computation by one of its images, with conservation of the velocity variables.

A division into 12 strips is superimposed to the drawing. Over each of them the average of the corresponding component of velocities is calculated, hence the velocity profile shown on the left part of the figure.

The calculation of contact forces in such a case of flowing grains involves less indetermination than in equilibrium situations. In fact when sliding occurs at a contact point, Coulomb’s law connects the direction of the contact force with that of the sliding velocity and this makes a richer information than merely stipulating that the force belongs to the friction cone. But the presence of sticking contacts should also be expected.

The procedure of varying the Gauss-Seidel iterations has been applied in order to estimate the amplitude of the indeterminations. On the right part of the figure are shown the grains F1 and F2, glued to ground, and a few grains of the flow.

When two polygonal grains touch each other along a line segment, the program retains only two contact points, at the extremities of this segment (this amounts to imagine that polygons exhibit on each of their edges some infinitesimal concavity). From each contact point, the program draws a line segment representing, with a fixed graphical unit, the calculated contact force.

Grains F1 and F2 constitute a single fixed rigid body that grain A, whose velocity is found zero, touches in three points. The fans observed at each contact point give some idea of the indetermination.

On the contrary, the two contact points declared between grains A and B show no force dispersion. These grains have a nonzero sliding velocity and the two contact forces have the same direction corresponding to the value 0.3 of the friction coefficient. Also at some vertex-to-edge sliding contacts no force dispersion is visible.

Figure 3: Flow of polygonal grains on a rugous slope
References


