Numerical simulation of finite dimensional multibody nonsmooth mechanical systems

Bernard Brogliato, Tony ten Dam, Laetitia Paoli, Frank Génot, Michel Abadie

Numerical simulation of finite dimensional multibody nonsmooth mechanical systems

B Brogliato

INRIA Rhône-Alpes, ZIRST Monbonnot, 655 avenue de l'Europe, 38334 Saint Ismier, France; Bernard.Brogliato@inrialpes.fr

AA ten Dam

National Aerospace Laboratory NLR, Department of Mathematical Models and Methods, Information and Communication Technology Division, PO Box 90502, 1006 Amsterdam, The Netherlands; tendam@nlr.nl

L Paoli

Université Jean Monnet-St Etienne, Equipe d'Analyse Numérique UMR CNRS 5585, 23 rue du Dr Michelon, 42023 Saint-Etienne cedex 2, France; paoli@anumsun1.univ-st.etienne.fr

F Génot

INRIA Rocquencourt, BP 105, 78153 Le Chesnay cedex, France; Frank.Genot@inria.fr

M Abadie

Schneider Electric, Avenue des Jeux Olympiques, S2E, bureau 204 38000 Grenoble, France

This review article focuses on the problems related to numerical simulation of finite dimen-sional nonsmooth multibody mechanical systems. The rigid body dynamical case is examined here. This class of systems involves complementarity conditions and impact phenomena, which make its study and numerical analysis a difficult problem that cannot be solved by rely-ing on known Ordinary Differential Equation (ODE) or Differential Algebraic Equation (DAE) integrators only. The main techniques, mathematical tools, and existing algorithms are reviewed.

1 INTRODUCTION

This article focuses on the problem of numerical simulation of systems composed of interacting rigid bodies. It aims at showing the state of the art in this field and, at the same time, introduces the main features, difficulties, and proposed solutions to simulation of nonsmooth multibody dynamics. This topic is an important part of the multibody systems dynamics general problem, which is itself a major field of investigation [1–3]. It concerns many domains of applications: interactive computer graphics, virtual prototyping [4,5] (electrical devices, watch mechanisms design), aerospace (control of space manipulators, liquid slosh phenomena in satellites), diesel engines, landing gear, analysis of vibrations in machines [6,7] assembly and disassembly processes [8], design of optimal protection from collisions [9], granular matter—dynamics of sandpiles, gravels, planetary rings—[10–12],

dynamics of buildings and structures [13], modeling and control of buildings, hopping [14], walking machines [15], kinematic chains with backlash [16,17], manipulators performing complete robotic tasks (deburring, grinding, polishing, chamfering) [18–20], etc. It also involves various scientific communities: Mechanical Engineering, Systems and Control, Applied Mathematics, Robotics, and Physics. In real-world applications, there exists a trend to use one and the same simulator for design, engineering, and training purposes [21]. A typical example is the simulation environment for a space manipulator [22]. The challenge for model developers is to provide users of simulators with accurate and realistic responses of simulated nonsmooth systems with sufficient level of detail [23]. Issues like repeatability and uniqueness of solutions are clearly of paramount importance. If real-time simulations are desired, the numerical algorithm must also perform real-time (ie, it has to calculate the motion faster than the real process evolution), possibly in a trade-off

with accuracy requirements. The choice of model, eg, rigid body versus compliant, becomes important. We will deal with these issues in the remainder of this paper. With the growing availability of commercial and research code for nonsmooth mechanical systems, the capabilities of these codes become of interest, especially since those packages tend to limit the time needed to code an application model.

The choice between rigid body or compliant models is a problem in itself. At this stage, the reader should understand these two definitions in a very rough fashion since a rigorous definition of rigidity and compliance requires some care [24]. This will be discussed in this article. Let us just point out that the choice of the model in general strongly depends on one's goals and field of application. It may be the result of an iterative process with successive simplifications or complications. There are applications (eg, Systems and Control, Robotics) in which one needs simple enough models (eg, that allow the designer to construct a stable feedback controller). At the same time, such models have to predict the real motion reasonably well, hence keep enough physical information inside them. In this setting, rigid body models seem quite suitable in many applications, where a compromise between accuracy and tractability has to be found.

As is well known, the principal qualities of a numerical algorithm are its accuracy and its stability (conditional stability, A-stability, or unconditional stability). Accuracy is related to the order of the scheme. Stability is related to the boundedness of the numerical solution for fixed steps of integration, as time progresses. A third property is the convergence of the numerical solution towards some function, on a fixed time interval, as the integration step goes to zero. One question that comes to mind is: are these properties transported from ODEs (free-motion systems) or DAEs (bilaterally-constrained systems) towards unilaterally constrained systems? For example, does a multistep algorithm that is known to converge for index 2 DAEs ([25], Ch. VII.3), still converge for nonsmooth systems? As we shall expose throughout this paper, the answer is not trivial, and might be negative in many cases. For instance, the Euler method is known to be not very accurate for ODEs or DAEs ([26], page 247). However, the occurrence of topology modifications and/or impacts (state re-initializations) may also render the higher-order methods useless, or at least much less powerful [27–29]. Moreover, some widely used tricks like Baumgarte's stabilization of constraints become quite inadequate in the case of inequality constraints, where the sign of the Lagrange multipliers is of primary importance. To summarize, the challenge could be formulated as follows: Consider the multibody system in [25] Ch. VII.7 (a seven-body mechanism), on which six different DAE codes are compared. Add some unilateral constraints with friction (say 15, which is a reasonable number). How much is the problem complicated? What does there remain to be done before getting a reliable and accurate simulation software, starting from the proposed algorithm (DASSL, DOPRI, RADAU, etc)? The answer to both questions is: a lot! We will lay the foundation for this answer in the remainder of this paper.

In the following, we shall generically denote the impact times as t_k, and the set of admissible positions as

$$K = \{ q : \phi(q) \geqslant 0 \} \tag{1}$$

with $\phi(\cdot)$ a differentiable vector valued function. The general dynamics of the systems we deal with may be written as follows:

$$
\begin{cases}
M(q)\ddot{q} = Q(q,\dot{q},t) + \nabla \phi(q)\lambda + P(q,\dot{q},\lambda) + \nabla \phi_b(q)\lambda_b \\
\phi(q) \geqslant 0, \quad \lambda \geqslant 0, \quad \phi^T \lambda = 0 \\
\phi_b(q) = 0 \\
\text{State re-initialization (Impact rule)} \\
\text{Coulomb friction.}
\end{cases} \tag{2}
$$

The matrix $M(q) \in \mathbb{R}^{n \times n}$ is the symmetric positive definite inertia matrix, $Q(q,\dot{q},t)$ accounts for inertial (centripetal, Coriolis), external (eg, control inputs), and frictional (Rayleigh dissipation) generalized torques, $P(q,\dot{q},\lambda)$ accounts for frictional (Coulomb) generalized forces, $\phi_b(q) \in \mathbb{R}^{m_b}$ is a set of bilateral constraints, λ_b is the corresponding Lagrange multiplier vector, $\phi(q) \geqslant 0$, $\phi(q) \in \mathbb{R}^m$ is the set of unilateral constraints which indicate in which domain of the configuration space of the system q is allowed to evolve, and λ is a vector of Lagrange multipliers which represent the normal part of the interaction between the system and its environment. As a convention in this paper, ∇ will always denote the gradient in the Euclidean metric (ie, $\nabla \phi^T = \partial \phi / \partial q$), though we shall make use of other metrics. The complementarity (or orthogonality) between $\phi(q)$ and λ indicates that physical interactions producing forces without contact (magnetic effects) and gluing forces are not taken into account in the model. In the following we shall generally consider that there are no permanent bilateral constraints in order to focus our attention on inequality constraints only. Systems as in (2) are sometimes named *complementary slackness* mechanical systems by analogy with the mathematical programming and convex analysis language ([30] page 226). Assume frictionless constraints, and that λ is an impulsive term, ie, $\lambda = p_k \delta_{t_k}$ with δ_{t_k} the Dirac measure at time t_k. Let $f(t_k^+)$ and $f(t_k^-)$ denote the right and left limits of $f(\cdot)$ at $t = t_k$, respectively. Then the Lagrange equations at t_k become [18]:

$$M(q(t_k))[\dot{q}(t_k^+) - \dot{q}(t_k^-)] = \nabla \phi(q(t_k)) p_k. \tag{3}$$

The mathematical problems associated to the dynamics in (2) are not trivial. As we shall explain later in this paper, the dynamics in Eqs. (2) and (3) represent a Measure Differential Inclusion (MDI). Its well-posedness (existence and uniqueness of solutions) is still an open problem in the general case. One of the very first contributions to this field can be found in [31]. Other contributions have been made by Monteiro-Marques [32], Stewart [33,34], Mabrouk [35,36], Carriero and Pascali [37,38], Heemels *et al* [39,40], van der Schaft and Schumacher [41,42], Lötstedt [43], Percivale and Buttazzo [44-48], Ballard [49,50], and Schatzman and Paoli [51–57]. In particular, systems with multiple contacts and/or friction create deep modeling and analysis difficulties. It is not the goal of this survey to provide many details on the mathematical aspects of the continuous-time dynamics. However, it is worth being aware of them since it seems difficult, even hazardous, to simulate a system reliably that is not well-posed (see Section 6.2)! In particular, one should be

aware of the fact that multiple impacts generally imply discontinuity of the solutions with respect to the initial data [18,32,39,58]. This may have consequences on the numerical as well as the control aspects. A non well-posed problem has little chance to yield a well-posed numerical algorithm when it is time-discretized. In [59], it is shown that a modified Coulomb's law yields non-uniqueness of solutions in a 1D system. It is also shown that the discretization with an implicit Euler scheme leads to difference equations that possess several solutions.

It is noteworthy that the tools and models described in this paper concern essentially mechanical systems subject to unilateral constraints with or without dry friction. However, they may also be suitable to the simulation of other nonsmooth physical systems, like electrical networks [27,42]. Indeed, some components like diodes can be modeled with the complementarity conditions which we describe later, and the state may possess discontinuities in complex electrical networks, requiring reinitialization rules similar to the mechanical collisions [39]. As we shall see, all these nonsmooth dynamical systems can be recast in the framework of Measure Differential Inclusions (MDIs). Infinite-dimensional as well as quasi-static or static problems with unilateral contacts are not examined in this paper, see, eg, [60] for a survey.

In this survey, we shall first recall in Section 2 the basics of modeling nonsmooth systems so that the paper contains some introductory material, before passing to more specific developments for numerical simulations. In Section 3, we introduce some peculiarities of nonsmooth multibody systems simulations when compared to simulation of smooth systems. Such an introduction is thought to be necessary in order to fix and clarify the fundamental discrepancies between systems as in (2) above, and various other types of switching, discontinuous, or impulsive dynamical systems that are often encountered in the literature (Applied Mathematics, Systems and Control, Computer Science). Section 4 is concerned with a discussion on the use of rigid and compliant models for simulation of contact of nonsmooth systems.

The main focus of the paper is on numerical analysis in Sections 5, 6, 7, and 9. In Section 5, a brief overview is given of numerical issues for bilaterally constrained mechanical systems. This analysis is used in Section 6 to discuss algorithms for unilaterally constrained mechanical systems. The numerical methods that are discussed are either so-called event-driven or time-stepping methods. A comparison of the methods is made. Convex analysis and mathematical programming tools are at the core of many works in this field and are therefore recalled. In Section 8, we briefly return to mathematical issues in modeling nonsmooth systems. An overview of commercial packages capable of simulating nonsmooth mechanical systems is presented in Section 9. Finally, concluding remarks can be found in Section 10.

2 DYNAMICS OF BILATERALLY CONSTRAINED MECHANICAL SYSTEMS—DAEs

2.1 Formulation of the unconstrained continuous dynamics

When a system contains a number of subsystems, it may not be practical to obtain a single system of differential equations that describes the behavior of the dynamical system. The subsystems may be of a different dynamical character. Take for example a satellite that contains liquid (cooling or fuel). Both the motion of the liquid and of the satellite can be described by differential equations. Another example is a robotic arm that contains flexible as well as rigid bodies. So, at first, a system description will contain submodels that capture specific (local) dynamics that can already contain operational constraint descriptions. By adding environmental constraints, one obtains a closed-chain description. In addition, feedback control laws lead to closed-loop, closed-chain system descriptions. There are two approaches to formulate models of unconstrained continuous dynamical systems: those that compute the mass matrix (Lagrange or Hamilton equations, recursive Newton-Euler), and those that do not (Featherstone recursive algorithm). Both approaches yield formulations that can be applied to open (serial or tree) chains and to closed-loop chains [26]. For closed-loop chains the trick is to cut the loop, introduce a multiplier, and then work with the multiplier method or reduce to independent coordinates. Recursive methods seem to have lost interest in recent years ([26] Remark, page 273). One of the reasons may be the availability of increasing computer power at ever lower prices. The advent of so-called $\mathcal{O}(N)$ methods, where N stands for the number of bodies in a dynamics model, happened at a time when computers were not as powerful as today. In many applications, special effort was paid to saving computational effort, and also to parallelization of algorithms. For large N, algorithms of $\mathcal{O}(N)$ became of interest in view of there efficiency. Some theoretical analyses suggested that the speed of performance of an $\mathcal{O}(N)$ algorithm would become superior over that of a $\mathcal{O}(N^3)$ algorithm for a sufficiently large number of bodies. Featherstone [61] is $\mathcal{O}(N)$ and supersedes $\mathcal{O}(N^3)$ for more than 10 bodies. In [22], it is reported that a special $\mathcal{O}(N)$ algorithm is faster, already, for two bodies. This may be due to tailoring the algorithm to the application at hand: a robotic manipulator with six rigid bodies and two flexible bodies. We refer to [26] for a summary of the different methods. A note on Jourdain's principle, and its relation to D'Alembert's principle and the so-called *Kane's dynamical equations* can be found in [22], Appendix A.

2.2 Reduction into independent coordinates

Since the presence of (equality) constraints leads to limiting the degrees of freedom of a system (model), several authors have proposed to find and select independent coordinates and reduce the number of dynamical equations. See, for example, McClamroch and Wang's (see [18]), Kane's methods (both more or less based on implicit functions), Wehage and Haug [62]. Haug's method does not aim at reducing the dynamics

in closed-form, but just reorders the generalized coordinates in view of correcting the drift away from the constraints [5]. The choice of (dependent) coordinates is not unique, and the reduction to independent ones is neither. In general, the reduction is a difficult task [63]. Advantages of the reduction are that the reduced-order dynamics usually require less computation time and the formulation is free of drift from the constraint surface. Reduced-order techniques are often available in commercial code-generation packages (see Section 9). A drawback is that addition of a new constraint again leads to a differential/algebraic description, and the interesting theoretical issue arises whether one should obtain a new reduced-order formulation from the original description or from the already obtained reduced-order formulation.

A decomposition into submodels simplifies modeling and makes the overall system model more flexible and easy to adapt. This modular approach to modeling facilitates the possibility of exchanging models for subsystems, which is a necessity in system design. Therefore, in the remainder of this paper we will concentrate on the Lagrange multiplier approach to modeling.

Some softwares [64] include heuristic approaches which allow one to decrease significantly, in certain cases, the number of algebraic conditions. An example of a gearbox with 11 clutches and algebraic (bilateral) constraints including 212 unknowns is presented in [64]. The algorithm reduces the number of variables to 23 unknowns. This is an NP-complete[1] problem. The implementation of such fast methods that allow one to eliminate useless constraints is also employed in collision detection, see Section 6.5. Clearly, a software must incorporate such algorithms to be efficient and fast enough.

2.3 Lagrange multipliers methods

Lagrange multiplier formulations of constrained mechanical systems can be obtained via so-called first-principles modeling. It is of interest to note that for constrained mechanical systems, Lagrange formulations can be derived that are particularly useful for simulation. See, for example, [25,65] (proposition 9.4.1) [66], and [67].

Let us view system (2) in the absence of friction, where we assume that Q does not depend explicitly on time.

$$\begin{cases} M(q)\ddot{q} = Q(q,\dot{q}) + \nabla\phi(q)\lambda \\ \phi(q) \geqslant 0, \quad \lambda \geqslant 0, \quad \phi^T\lambda = 0. \end{cases} \tag{4}$$

For systems as in (4), information on where and how contact with the constraints can be made and release can take place can be derived off-line using the algorithms in [68], for the linear case, and in [65] (Ch 7) for the nonlinear case. For simulation studies, it is useful to first look at (4) in case of equality constraints. The bilaterally constrained mechanical system in Eq. (5) is obtained.

$$M(q)\ddot{q} = Q(q,\dot{q}) + \nabla\phi(q)\lambda \tag{5}$$

$$\phi(q) = 0.$$

The Lagrange multiplier now ensures that the motion of the mechanical system satisfies the constraint equation. Differentiating the constraint equation gives $\nabla\phi^T(q)\dot{q} = 0$. Denote $(x_1^T, x_2^T)^T := (q^T, \dot{q}^T)^T$. Then starting from Eq. (5), the following first-order formulation can be obtained.

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = M^{-1}(x_1)Q(x_1,x_2) + M^{-1}(x_1)\nabla\phi(x_1)\lambda \\ \phi(x_1) = 0 \\ \nabla\phi^T(x_1)x_2 = 0. \end{cases} \tag{6}$$

In [65] proposition 9.4.1 (see also [29] Ch. 2.5.3), it is shown that any trajectory of the system in (6) is also a trajectory of the system in (7) and vice versa.

$$\begin{cases} \dot{x}_1 = x_2 + \nabla\phi(x_1)\mu \\ \dot{x}_2 = M^{-1}(x_1)Q(x_1,x_2) + M^{-1}(x_1)\nabla\phi(x_1)\lambda \\ \phi(x_1) = 0 \\ \nabla\phi^T(x_1)x_2 = 0. \end{cases} \tag{7}$$

The formulation in (7) is similar to the formulation in (6), apart from the presence of the Lagrange multiplier μ. It allows one to reduce the index of the DAE from 3 to 2 ([25], Ch VII.1 [67]). The actual analysis that leads to the formulation in (7) is beyond the scope of the present paper, and we refer to [65], also for other equivalent (overdetermined) formulations. It must be emphasized that the Lagrange multiplier μ has no physical meaning. From the analysis in [29], [65], and [66], it actually follows that $\mu = 0$, but the importance of formulation (7) lies in its use to obtain numerical solutions for bilaterally and unilaterally constrained mechanical systems. The two continuous-time representations are equivalent. However, they do not result in equivalent discrete-time formulations (see also Section 5.3).

All formulations (5), (6), and (7) can be used as starting points for analysis and simulation of constrained mechanical systems. In Section 6.3, these formulations will be used in what we will call event-driven simulation schemes, where decisions whether or not contact or release of a constraint takes place must be based on physical interpretations of λ and information on the state of the system, and not only on the position.

3 DYNAMICS OF UNILATERALLY CONSTRAINED MECHANICAL SYSTEMS—MDIs

3.1 The major problem of multiple contacts—LCPs

In multibody systems with multiple contacts, the major difficulty is that the change in one contact generally implies changes at the other contacts, and this is true for detachment conditions as well as for impacts. How to treat this within the rigid body approach? This is the subject of the next sections. In this section, we will describe the effective formalisms that are used in view of numerical simulations. Since there generally exist several manners to formulate the same problem, it is quite important to choose the most suitable one (eg, the

---

[1]See eg, *Computer Theory FAQ* at http://www.cs.unb.ca/alopez-o/comp-faq/faq.html for an introduction.

simplest one if CPU time is an issue). In Section 6.6, we shall focus on the numerical tools used to solve these problems.

Detachment conditions (Delassus' problem): In the case of a single frictionless contact point, the detachment condition can be watched by looking at the sign of the Lagrange multiplier: if $\lambda = 0$, then a sufficient condition for subsequent detachment is that the normal acceleration be >0. Things complicate when several contact points persist on a nonzero time interval. It is well-known since Delassus [18,69] (see also [7] for a simpler example) that *a priori* assuming (in an arbitrary way) certain values of the acceleration (for instance, one decides that the contacts persist) may lead to a contradiction because the calculated multipliers $\lambda_i < 0$. Therefore, one is led to test all the possible combinations—there are 2^m combinations for m active contacts. Such enumeration is cumbersome! However, it is a simple matter to use the constrained Lagrange dynamics in order to express the acceleration $\ddot{\phi}(q(t))$ in terms of λ, see Eqs. (10) and (35) below. Doing it this way and noticing that, provided the m contacts have been active on the interval $[t - \epsilon, t)$, which allows one to state that the complementarity conditions in (2) imply the same conditions with ϕ being replaced by $\ddot{\phi}$ (see (10)), one is able to construct a *Linear Complementarity Problem* or LCP, of the form:

$$A\lambda + B \geqslant 0, \quad \lambda \geqslant 0, \quad \lambda^T (A\lambda + B) = 0 \qquad (8)$$

where the matrices A and B generally depend on q, and q, $\dot{q}$ respectively, and on the system's physical parameters. The unknown of the LCP in Eq. (8) is λ, and we shall denote it as LCP(λ). When friction is present at the contacts and in dimension 3, one gets a Nonlinear Complementarity Problem (NCP, or NLCP) due to the 3D friction cone, which is a much more tricky object, see Section 3.3. Moreau [70,71] was the first to formulate the contact law this way, hence greatly improving Delassus' work. In case of friction it is known that unexpected phenomena can occur, which are due to the lack of physical modeling. However, these phenomena do represent the behavior of the system when the contact stiffnesses are high, and will occur when doing a simulation. For instance, a planar disc may remain wedged in an angle 2α if the friction coefficient at both contacts μ satisfies $\mu > \arctan(\alpha)$. This is related to the pointedness of the total friction cone $C_1 + C_2$, see [33]. Also, the LCP may not possess any solution. In conclusion, let us state that the contact status management is one of the major issues of multibody systems simulation. But contrary to collisions which involve a lot of physical modeling, it is mainly a mathematical problem. As we shall recall later, LCPs possess several equivalent formulations (quadratic programs, complementarity functions) that may be used for numerical implementation.

Notice that to form LCPs that will monitor the topology or transition modifications, index 1 formulations are more convenient (but this does not imply that higher index approaches cannot be settled). Hence we focus on two cases: *Velocity-impulse schemes:*

$$\nabla \phi^T \dot{q} \geqslant 0, \quad \lambda \geqslant 0, \quad \lambda^T \nabla \phi^T \dot{q} = 0. \qquad (9)$$

Acceleration-force schemes:

$$\ddot{\phi}(q) = \overrightarrow{\nabla \phi^T \dot{q}} + \nabla \phi^T \ddot{q} \geqslant 0, \quad \lambda \geqslant 0, \quad \ddot{\phi}^T \lambda = 0. \qquad (10)$$

It is important to keep in mind that the dynamics in (2) with the position complementarity conditions is not equivalent to the dynamics with the velocity or the acceleration complementarity conditions. Actually, if $\phi(q(\cdot)) = 0$ on $I \triangleq [t - \epsilon, t)$, then the position complementarity implies the velocity complementarity on I. If $\phi(q(\cdot)) = 0$ and $\dot{\phi}(q(\cdot)) = 0$ on I, then it implies the acceleration formulation of the complementarity conditions. And so on with the higher derivatives. This yields the notion of Dynamic Complementarity Problems [41]. As we shall see the index, reduction is crucial in discretized schemes since it permits to formulate complementarity problems at each step of integration. Index problems in DAEs are major problems [72]. The reader may expect that the difficulties will be magnified when unilaterally constrained systems are considered. See also Section 7.2.4 concerning the choice between position and velocity complementarity conditions.

Remark 1 (Hyperstatic systems): When the system is hyperstatic (ie, rank($\nabla \phi$) $> n$, implying $m > n$) it may become necessary to solve an LCP at each instant (in the numerical integration) to manage the detachment events. In case of non-hyperstatic system, like when $m = 1$, the Lagrange multipliers can be computed uniquely and it suffices to solve an LCP when they approach zero. As long as they keep their signs strictly positive and no constraint is attained, the system can safely be considered as a DAE and integrated as such. Unfortunately, hyperstatic systems are by far the more commonly met category in practice! An example is the rocking block [18] when both contact points of the base stick [7].

3.2 Complementarity formulation of restitution mappings

A multiple impact occurs each time the system collides the boundary ∂K at a singularity. If the codimension of the struck subspace is m, the impact is called an m-shock or m-impact. In the rocking block example, the shock is a 2-impact. Multiple impacts pose deep problems: continuity of the solutions with respect to the initial data may be lost[2] (hence a high sensitivity with respect to the choice of the initial conditions), and even their formulation (ie, how to get an impact rule that maps the pre-impact velocities to the post-impact ones) is not trivial. Various approaches have been presented in the literature, some of which will be described in the sequel. We can already state that the problem of multiple contacts with or without friction is one of the major problems of nonsmooth mechanics (at the date of writing of this paper). Ivanov [58,73] argues that as soon as the codimension of the struck subspace is larger than 3, impacts should be modeled in a statistical way. This may be related to the *hybrid* manner of designing the software by introducing a

---

[2]This is easily seen with a 2D particle striking in an angle. See eg [51] Section 3.a.

tableau containing the probabilities of events to occur. See Section 10 for more details. Actually, the study of multiple impact rules via the limit of penalized problems or the limit of sequences of simple impacts seems to be a hard task in general [74,75]. Collecting statistical data from experiments may be an alternative path.

It is possible to formulate the restitution mappings (or more generally the mappings that rule the re-initialization of the state vector when the system jumps from one mode to another one [42]) through a complementarity law. The following *plastic* impact rule:

$$\dot{q}(t_k^+) = \arg \min_{\nabla \phi^T z \geqslant 0} \sigma_{\dot{q}}^T M(q) \sigma_{\dot{q}} \tag{11}$$

with $\sigma_{\dot{q}} \triangleq z - \dot{q}(t_k^-)$, has been apparently first formulated (with words) by Carnot [76]. Its link with Gauss' principle and complementarity laws has been first discovered by Moreau [70,71], and used in [27]. Actually, solving Eq. (11) is equivalent to solving the following LCP whose unknown is $\dot{q}(t_k^+)$:

$$\nabla \phi^T(t_k) \dot{q}(t_k^+) \geqslant 0, \quad p(t_k) \geqslant 0,$$

$$p^T(t_k) \nabla \phi^T(t_k) \dot{q}(t_k^+) = 0 \tag{12}$$

where p_k is defined in (3). This can be shown using the algebraic shock dynamics in (3), and the Kuhn-Tucker's conditions (in which case the percussion p_k is to be interpreted as a slack variable). Notice that for an m-shock with $m \geqslant 1$, $p^T(t_k) \nabla \phi^T(t_k) \dot{q}(t_k^+) = 0$ is equivalent to $p_j^T(t_k) \nabla \phi_j^T(t_k) \dot{q}(t_k^+) = 0$ for each $j = 1, \cdots, m$, with all m components of each vector non-negative (hence the set of complementarity conditions in Eq. (12) is equivalent to m 1D complementarity slackness conditions). Then it is easy to see that in the one degree-of-freedom case, this corresponds to a plastic shock $e = 0$. The physical validity of these impact rules is discussed in [77], where it is argued that some simple multiple impacts may not be described by such laws. Basically, this is due to the physical fact that impulsive forces may appear at contacting points that were previously lasting and which detach after the shock. Therefore, both the post-impact velocity and percussion are nonzero at the impact time, precluding any complementarity between them. However, this rule may apply in other cases, as proved in [75] where the convergence of a penalized problem is studied. Moreover, a proper definition of the pairs of complementarity variables allows one to treat detachment of previously contacting points (on a non-zero time interval) to describe such motions as rocking [78].

Notice that one can replace $\dot{q}(t_k^+)$ in (12) by a combination of $\dot{q}(t_k^+)$ and $\dot{q}(t_k^-)$ to allow for elastic impacts, see Sections 3.4.1 and 3.4.3 (eg, replace $\dot{q}(t_k^+)$ by $\dot{q}(t_k^+) + e\dot{q}(t_k^-)$ and check that this implies that $\nabla \phi^T(t_k) \dot{q}(t_k^+) \geqslant 0$ whenever $\nabla \phi^T(t_k) \dot{q}(t_k^-) \leqslant 0$, $e \in [0,1]$).

3.3 Constraints with Coulomb friction

Let us assume that there is one active constraint where dry friction acts at a contact point $\mathcal{A}$. At $\mathcal{A}$ one attaches a local frame in which Coulomb's law is expressed. The problem

now is how to write the generalized torque $P(q, \dot{q}, \lambda)$ in (2) in order to recover a complementary slackness formulation that includes both normal and tangential reactions [7,41,79-83]. Several particular cases have to be considered: 2D versus 3D friction cones, and all possible transitions between stick and sliding modes. Let us denote the tangential velocity components at the contact point $\mathcal{A}$ as V_t, the normal one as V_n, so that $V = \begin{pmatrix} V_i \\ V_n \end{pmatrix}$, and the reciprocal contact force as $F = \begin{pmatrix} F_i \\ \mathbf{N} \end{pmatrix}$. Consider $m = 1$ in (2), ie, $\phi(\cdot)$ is a scalar function. If $V_n = \nabla \phi^T \dot{q}$, then $\lambda = \mathbf{N}$. Then one has [82,83]:

$$\begin{cases} V_t = 0 \Rightarrow |F_t| \leqslant f\mathbf{N} \\ V_t \neq 0 \Rightarrow |F_t| = f\mathbf{N}, \quad F_t = -\alpha V_t, \quad \alpha \geqslant 0. \end{cases}$$

$$\begin{cases} V_t \neq 0 \Rightarrow F_t = -f\mathbf{N}V_t/|V_t| \\ V_t = 0: \begin{cases} \dot{V}_t = 0 \Rightarrow |F_t| \leqslant f\mathbf{N} \\ \dot{V}_t \neq 0 \Rightarrow |F_t| = f\mathbf{N}, \quad F_t = -\alpha \dot{V}_t, \quad \alpha \geqslant 0. \end{cases} \end{cases} \tag{13}$$

The first set of conditions describes dry friction at the velocity level, and the second set at the acceleration level. The latter is used to monitor the transitions from stick to sliding (rolling constraints [82,89]) in what we shall call event-driven algorithms. The reader may notice that the way to go from velocity to acceleration formulation of Coulomb's law is exactly the same as for the normal direction when one replaces position by velocity or acceleration complementarity conditions. Here we assume that there is only one contact to avoid cumbersome notations, but the formulations can be readily extended to the case of multiple contacts (see, however, Sections 7.2.2 and 7.4 for comments about the discretization of multiple contacts with friction). The next two Sections, 3.3.1 and 3.2.2, deal with how to transform the models in (13) into complementary slackness forms, and Sections 3.3.3 and 3.3.4 describe how to replace $P(q, \dot{q}, \lambda)$ by a set of complementarity conditions (linear or nonlinear) that enables one to monitor all the possible mode transitions for frictional contacting points. As we will see later, this enables one to discretize the whole set of dynamical equations and conditions and construct an LCP (or an NCP, or more generally any set of equations and/or conditions that can be solved with available algorithms) at each step, which permits us to advance the numerical algorithm to the next step.

3.3.1 Two-dimensional friction

In this case $V_t \in \mathbb{R}$. The transformation of the velocity conditions in Eq. (13) uses the fact that a relay characteristic between two variables v and z (that may be expressed as $v = \text{sign}(z)$, with $-1 \leqslant v \leqslant 1$ if $z = 0$), can be formulated with complementary slackness conditions as [7,79,81]:

$$z = y_1 - y_2, \quad v = \tfrac{1}{2}(u_2 - u_1), \quad u_1 + u_2 = 2$$

$$u_i y_i = 0, \quad u_i \geqslant 0, \quad y_i \geqslant 0, \quad i = 1,2. \tag{14}$$

The result follows by observing that $F_t = f.\mathbf{N}\text{sign}(-V_t)$. In particular, if $V_t = 0$ one gets $-1 \leqslant F_t/f.\mathbf{N} \leqslant 1$. The quantities y_1 and y_2 are the positive and negative part of $z(=-V_t)$, respectively. This derivation can also be understood from (23) (see below) noting that the projection $\mathcal{D}$ of

the friction cone on the tangent plane at $\mathcal{A}$ reduces to an interval, so that the results of convex analysis concerning normal cones of polyhedral sets (see [30], Ch 23) and complementary slackness conditions can be used directly. The same transformation can be performed for the acceleration law in Eq. (13) [82,83,89] introducing the positive and negative parts of $\dot{V}_t$.

3.3.2 Three-dimensional friction

This case is more difficult since $V_t \in \mathbb{R}^2$ and the friction cone is not polyhedral. Consequently (23) is still true, but cannot be transformed into complementarity conditions by introducing additional slack variables as in (14). The basic idea [33,79,80,82–88] is to approximate the cone, ie, to perform a polygonalization. Some more comments on this will be made in Section 7.4. Notice that this is the same as transforming the disc $\mathcal{D}$ in (23) below into a (convex) polyhedral set $\hat{\mathcal{D}}$ [80,83]. The pyramid approximation obviously corresponds to $\hat{\mathcal{D}}$ being a square [89]. A finer approximation corresponds to adding edges to the polygon $\hat{\mathcal{D}}$. Then a result of convex analysis ([30], Ch 23) allows one to generalize (14) and to express this approximated Coulomb's law with complementary slackness conditions of the form (we choose the acceleration formulation in (13)):

$$-\dot{V}_t = \sum_{i=1}^{\mathbf{e}} \mathbf{e}_i \xi_i, \quad \sigma_i = f.\mathbf{N} - \mathbf{e}_i^T F_t$$

$$\sigma_i \geqslant 0, \quad \xi_i \geqslant 0, \quad \sigma_i \xi_i = 0, \quad i = 1, \cdots, e \tag{15}$$

where $\mathbf{e}_i$ is a unit vector normal to the ith edge of $\hat{\mathcal{D}}$, $\mathbf{e}$ is the number of edges, and ξ_i are slack variables. A similar derivation can be found in [89] for the pyramid approximation, where the variables σ_i and ξ_i are directly the positive and negative parts of the corresponding quantities, provided the local frame and $\hat{\mathcal{D}}$ are properly oriented (hence the $\mathbf{e}_i$s simplify). Notice that almost all authors choose an outer approximation of the cone, although there does not seem to be any fundamental reason for such a choice that yields an overestimation of frictional effects. Actually the most important feature is to find out a formulation that involves as few slack variables as possible, so as to simplify the subsequent numerical procedure.

Let us assume that $\dot{V}_t = J_t \ddot{q} + h(t)$. At this stage one has $P(q, \dot{q}, \lambda) = J_t^T F_t$, and Coulomb's law is expressed through complementary slackness conditions as in Eqs. (14) or (15). The next step is: how to get an LCP that allows one to calculate the unknowns (the normal and tangential contact forces) at each time?

3.3.3 Sliding contacts

The velocity formulation in Eq. (13) holds, which is linear in the contact force [89]. There is a direct relationship between F_t and $\mathbf{N}$, whatever the dimension. Then $P(q, \dot{q}, \lambda) = J_t^T(q)f.\mathbf{N}V_t/|V_t|$. It is therefore a simple matter to express $\ddot{q}$ as a function of q, $\dot{q}$, and $\mathbf{N}$, then to replace $\ddot{q}$ in $\ddot{\phi}(q)$ to get a LCP as (8) with unknown $\mathbf{N}$. In the case of $m_s \leqslant m$ sliding contacts, $\lambda^T = (\mathbf{N}^{(1)}, \cdots, \mathbf{N}^{(m_s)})$ in (2). Due to friction, the matrix $A \in \mathbb{R}^{m_s \times m_s}$ (see (8)) is not symmetric in general and it may also lose its copositivity property. Then existence and uniqueness of λ at time t may no longer be assured. This

yields Painlevé problems, see Section 8.1. Notice that there is no inversion of the matrix $\nabla \phi^T M^{-1} \nabla \phi$ in this procedure.

3.3.4 Sticking (or rolling) contacts

In the 3D case, one has to resort to the polygonalization of the friction cone to avoid nonlinear formulations, see Section 3.2.2. Obviously in the 2D case the cone is already polygonal ($\mathcal{D}$ is an interval) so we shall consider this case as a subproblem of the 3D one once the cone has been polygonalized. Notice that the acceleration friction model is highly nonlinear in the contact force, contrary to the velocity formulation for sliding contacts. Various techniques have been employed in the literature to formulate from (2) and (14) or (15) an LCP whose unknowns are slack variables (including λ in (2)) that allow one to monitor the contact status (detachment/ contact, sticking/sliding). The method proposed in [79] does not accomodate hyperstatic systems. Glocker [7,83] uses the formalism in Eq. (15) and the fact that $\dot{V}_t = J_t(q)\ddot{q} + h(t)$, to get at each contact point an $(\mathbf{e} + 1)$-dimensional LCP in λ plus e slack variables σ_i, ξ_i taken in such a way that hyperstatism problems are avoided. No inversion of the matrix $J_t^T M^{-1} J_t$ is needed. A similar method is proposed in [82,89] for the pyramidal approximation, together with some existence result for the LCP. Without polygonalization, a Nonlinear Complementarity Problem (NCP) of the form

$$y \geqslant 0, \quad g(y) \geqslant 0, \quad y^T g(y) = 0 \tag{16}$$

has to be solved for some slack variable y [83], or even some non-standard NCPs (generalized complementarity problems) [82,89–91], see Section 3.4.2. NCPs are more complex to solve than their linear counterparts, see Section 6.6, and potential users have to develop their own codes. However they may involve less variables than polyhedral approximations [90]. In [83], it is shown how to get a standard 4D NCP per contact. The all-rolling contacts problem is formulated with a Quasi-Variational Inequality in [90] and results on existence of solutions are given. The work in [92] is not based on an approximation of the friction cone and treats the acceleration as the unknown, see Section 6.7.1 for more details.

Remark 2: From a hybrid dynamical system point of view [42] Coulomb friction adds modes or discrete-event states to the system.

3.3.5 Shocks with friction

The problem of frictional impacts is complex, and we shall not insist on it in this article, see [18,93] for details. In the complementary slackness framework, it is of interest to extend the frictionless rule in Eq. (12) to the frictional case. Several authors proposed complementary formulations of frictional impacts [7,27,86,94]. The basic idea of Lötstedt [27] is to extend directly Eqs. (11) and (12) as:

$$\min \tfrac{1}{2} P_k^T M^{-1} P_k + P_k^T \dot{q}(t_k^-)$$

$$p_k \geqslant 0, \quad -f p_k \leqslant p_{t,k} \leqslant f p_k \tag{17}$$

where p_k is the normal percussion in (3), $p_{t,k}$ is its tangential counterpart, and P_k denotes the total percussion vector (ie, the right-hand-side of (3) for frictional impacts). The second line in Eq. (17) is the direct extension of Coulomb's law at

the percussion level, a model that may be energetically inconsistent, see Section 7.1.2. If $f=0$ then (17) reduces to (11) using Kuhn-Tucker conditions. There are different ways to rewrite Eq. (17), using the dual QP and (3), see also (54). Some details on the works in [86] will be given in Section 7.4. For 2D problems, it is similar to the one in [7], where the collision is decomposed into a compression and restitution phases. Concerning [94] see Section 7.2.4.

3.4 Convex analysis in nonsmooth mechanics

In this section we present compact forms of the MDIs in Eqs. (2) and (3). The formulations that follow in this section are based on mathematical tools from convex analysis [30,95]. It is not our goal in this article to provide details on such mathematical tools. However, it seems mandatory to make a short presentation of these approaches, since they are used to develop well-posedness and numerical issues of many algorithms. A tool that will be needed in the following is the definition of a subdifferential. The subdifferential of a convex function $f(\cdot)$, denoted as ∂f, is the set of subgradients of f at y, ie, of vectors γ satisfying $f(x)-f(y)\geq\gamma^T(x-y)$ for all x. This is the generalization of the derivative for non-differentiable convex functions. Let K denote a convex set (not necessarily the one in (1), though this may be the case as we shall see further). The function $\psi_K(y)$ is called the indicator function of the set K, and is defined by $\psi_K(y)=0$ if $y\in K$, $\psi_K(y)=+\infty$ if $y\notin K$. The values inside and outside the set K are consistent with the definition of subgradients.

The subdifferential of $\psi_K(\cdot)$ generalizes the notion of normal directions to ∂K at y. Roughly speaking, this represents the outward normal directions to the boundary of K at some point y, and if the boundary is smooth this is the usual outward normal half line. If the boundary is nonsmooth, it generalizes to a cone, called the *normal cone* and is denoted as $N_K(\cdot)$. When $y\in\text{Int}(K)$, then $\partial\psi_K(y)=\{0\}$, and when $y\notin K$, then $\partial\psi_K(y)=\emptyset$. One also defines *tangent cones* $V(y)$ as $V(y)=\{v:\forall\quad x\in N_K(y),\ x^Tv\leq0\}$ when $y\in K$. Thus if $y\in\partial K$, $V(y)$ is another convex cone. In this case tangent cones reduce to (inward) half spaces for differentiable codimension one boundaries ∂K, see Fig. 1a,c. If $y\in\text{Int}(K)$, then $V(y)=\mathbb{R}^n$, and if $y\notin K$, then $V(y)=\emptyset$. When K is defined as in (1), one may define the tangent cone to K at q as [96]:

$$V(q)=\{v\in\mathbb{R}^n:v^T\nabla\phi_i\geq0,\quad\forall\ i\in J(q)\}\quad(18)$$

with $J(q)=\{i\in\{1,\ldots,m\}:\phi_i(q)\leq0\}$. One notes that this definition coincides with the original definition of the tangent cone as long as $y\in K$. The fact that $V(y)$ is given a meaning outside K is useful for numerical applications in which the admissible domain of configurations in (1) may be violated. In order to motivate the reader, let us note that complementarity conditions between two scalar variables λ and y:

$$\lambda\geq0,\quad y\geq0,\quad\lambda.y=0\quad(19)$$
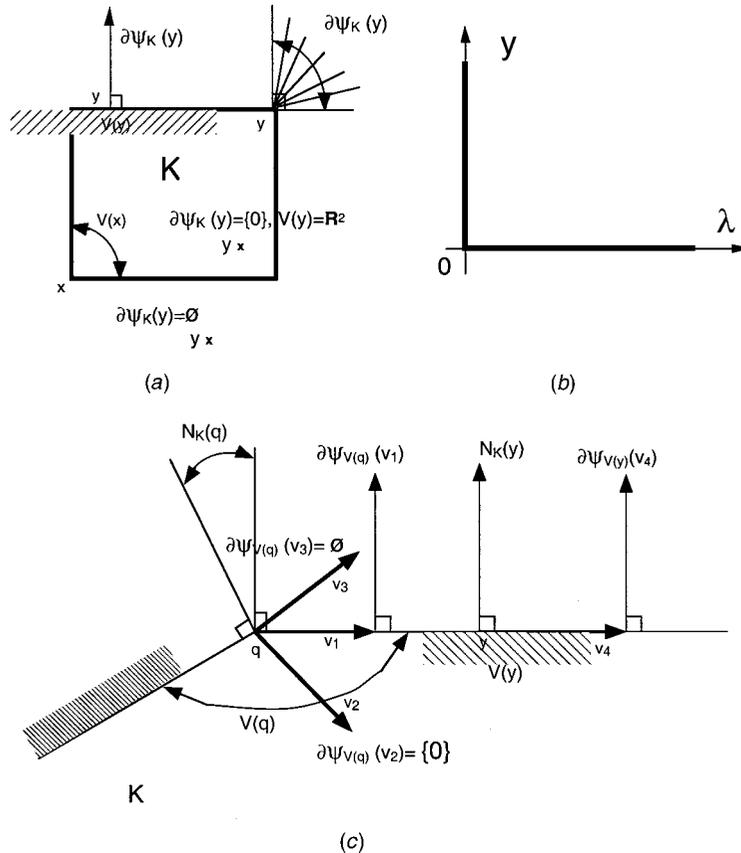
can be equivalently formulated as

Fig. 1 Tangent and normal cones, and the corner law

$$-\lambda \in \partial \psi_K(y), \quad \text{with} \quad K = \mathbb{R}^+. \tag{20}$$

This means that the corner law whose graph is depicted in Fig. 1b, is actually the graph of $-\partial \psi_{\mathbb{R}^+}(\cdot)$. Evidently the role of λ and y can be reversed in (20). This generalizes to the case of complementarity conditions as in (2) whenever K is convex. In this setting, (4) is equivalent to $-M(q)\ddot{q} + Q(q,\dot{q}) \in \partial \psi_K(q(t))$. Both Eqs. (19) and (20) are also equivalent to any one of the formulations [30] page 215; [97]:

$$\lambda \in K, \quad \forall \quad z \in K, \quad \langle \lambda, z-y \rangle \geqslant 0 \tag{21}$$
$$\lambda \in K, \quad \lambda = \text{proj}_K[\lambda - \rho y], \quad \text{for any} \quad \rho > 0.$$

The first formulation in Eq. (21) is a variational formulation of the complementarity conditions. This illustrates that there are several manners to express the same physical law, and such a flexibility is likely to be very useful for various reasons (mathematical studies, effective dynamic formalisms in view of numerical applications). Quite interestingly, the Coulomb friction model can be expressed in a similar fashion, as we already saw. This comes from the fact that by introducing slack variables [7,41] one can recast Coulomb friction into a complementarity framework, see Eqs. (14) and (15). The same holds for many systems with piecewise linear characteristics [39,42]. Some works [64] rely on the parametrization of Eq. (19) by introducing an additional slack variable s. These facts clearly indicate why the dynamics in (2) can be rewritten as a special type of differential inclusion, containing measures. This will give rise to Measure Differential Inclusions (MDIs), as introduced by Moreau [98].

Remark 3: One very good reason for expressing the dynamics in terms of convex analysis is that it yields compact and powerful mathematical formulations and that these properties are preserved in the numerical schemes derived from them. For instance, the graphs of the complementarity (or Signorini) conditions and of Coulomb's law are monotone. It has also been used to derive well-posed impact rules in [99].

3.4.1 Moreau's sweeping process

The theoretical setting exposed in this section has been developed by Moreau [96,98,100] (we refer to [101] for a introduction to the sweeping process, including the first order case and to [102] for a general exposition of its mathematical properties). Moreau's sweeping process is a velocity-impulse formulation, and its main interest lies in its compactness. This property has consequences for its numerical analysis, see Section 7.2. Let us present it briefly in the case of frictionless constraints. What follows does not pretend to possess any mathematical rigor, but is to be considered as introductionary material only. The starting point is:

$$-M(q)dv + Q(q,v,t) \in \partial \psi_{V(q(t))}(v(t^+)) \tag{22}$$

The variable v satisfies $q(t) = q(0) + \int_0^t v(\tau)d\tau$, ie, it equals $\dot{q}(t)$ almost everywhere in the Lebesgue measure sense. Notice that Eq. (22) is an *implicit* formulation in $v(t^+)$. In particular, if v has a jump $\sigma_v = v(t^+) - v(t^-)$ at t then $dv = \sigma_v \delta_t$. When $v(\cdot)$ is time-continuous, then one just replaces its right-limit by $v(t)$ in (22). Let us intuitively ex-

plain what the differential inclusion in (22) means. The interested reader may have a look at [18,96,100] for further details and examples with figures. The set $V(q(t))$ is the tangent cone to K (see (1)) at $q(t)$. When $V(q(t)) = \mathbb{R}^+$ then the interpretation of Eq. (22) is obvious from Eqs. (19)–(21) (this is, for instance, the case for a one-dimensional bouncing ball touching the ground a time t). The term $\partial \psi_{V(q(t))}(z)$ is a subdifferential and can be understood as follows when $q(t) \in \partial K$. If the vector z[3] points inward $V(q(t))$, then $\partial \psi_{V(q(t))}(z) = \{0\}$, so that (22) reduces to the time-continuous Lagrangian equations. If z belongs to the boundary of $V(q(t))$, then $\partial \psi_{V(q(t))}(z)$ is the normal cone to $V(q(t))$ at z. If z points outward $V(q(t))$ (thus outside K) then $\partial \psi_{V(q(t))}(z) = \varnothing$: this situation is forbidden, and one sees that the implicit formulation in (22) *implies* that $v(t_k^+) \in V(q(t_k))$. Now recalling that at $t = t_k$, the variable z is the post-impact velocity, one obtains that (22) becomes $-M(q(t_k))\sigma_v \in \partial \psi_{V(q(t_k))}(v(t_k^+))$. This is equivalent to a quadratic problem as the one in (11) or (12) and this implies that $v(t_k^+) \in \partial K$: this is a generalization of plastic impacts for codimension one constraints. Another interpretation is in terms of *proximation*, that is $v(t_k^+) = \text{prox}[V(q(t_k)), v(t_k^-)]$, where the proximation is to be made in the kinetic metric. $v(t_k^+)$ is therefore the closest element to $v(t_k^-)$ in $V(q(t_k))$, in the kinetic metric. If $q(t) \in \text{Int}(K)$, then $V(q) = \mathbb{R}^n$ so that $\psi_{V(q)}(z) = 0$ for all $z \in \mathbb{R}^n$ and $\partial \psi_{V(q)}(z) = \{0\}$: (22) is a classical Lagrange equation.

Remark 4:
- From Eqs. (19) and (20), one suspects that (22) implies complementarity conditions between the contact reaction and $v(t^+)$. This is indeed the case, see [18], [96].
- It is possible to show that the set $\partial \psi_{V(q(t))}(v^+) \subseteq N_K(q(t))$. This can be seen by drawing a 2D angle and plotting the various cones and half-spaces involved, see Fig. 1c where $\partial \psi_{V(q(t))}(v_1) \subset N_K(q(t))$. This means that (22) is not strictly equivalent to (4).
- It is possible to include a restitution coefficient e in this formulation by replacing v in both sides of (22) by an averaged velocity $u = v(t^+) + ev(t^-)/1+e$ [35,96], see Section 7.2. It is equivalent to inclusion of u in Eq. (12), as pointed out after (12) (notice that in Eq. (22), $(1+e)$ in the denominator is needed to encompass continuous velocities as well).
- The '$-$' sign in the left-hand-side of (22) is only a consequence of some standard notations in convex analysis, see (20).
- Mechanical systems with position constraints have index 3 [25]. Indeed one needs to differentiate the constraint $\phi(q) = 0$ three times to recover a set of ODEs with state $(q, \dot{q}, \lambda)$. Now consider the simple dynamics $M\ddot{q} = \lambda$, $\phi(q) = 0$. Let us write it in a velocity-impulse form, ie, $M(\dot{q} - \dot{q}(0)) = p$, with $p(t) = \int_0^t \lambda(\tau)d\tau$. Then one needs only to differentiate $\phi(q)$ twice (with respect to time) to get an ODE with state (q, p). Thus the constrained modes of (22) correspond to index 2 problems.

---

[3]Usually one draws the sets $q(t) + N_K(q(t))$ instead of $N_K(q(t))$ [102], so z should be understood as the vector emanating from $q(t)$. Similarly for the other sets.

It is important to notice that the essence of the dynamics in (22) is that the complementarity conditions in (2) *and* in (12) are stated through the subdifferential in the right-hand-side of (22), which is a function of the velocity (for fixed position). Velocity spaces are linear spaces (tangent to the configuration space at each q), whereas configuration spaces are seldom linear. Consequently, the use of velocity spaces allows one to recover all the advantages of linear spaces (for instance the tangent cone $V(q(t))$ is always convex, whereas K in (1) may not be). Mathematicians use such Nonsmooth Analysis tools in their existence and uniqueness of solutions studies, which are of high importance for numerical analysis. Indeed, on the one hand, well-posedness results are often based on time-discretization of the continuous dynamics (similarly to the fundamental result for ODEs that uses a Euler discretized scheme whose piecewise constant solutions are shown to converge to solutions of the continuous dynamics when the vector field is continuous [103]). This is what is usually called the *time-stepping* method and will be described later in this article. On the other hand, a necessary first step (but not sufficient to get a robust simulator!) is to establish the existence and uniqueness of solutions before designing a numerical simulator.

3.4.2 Coulomb friction and the principle of maximal dissipation

Moreau [104] has shown that a Coulomb friction model actually satisfies a *maximal dissipation principle*. When there is sliding between two bodies at the contact point, it is intuitively clear that this model obeys such a principle. Indeed it says that the tangent reaction F_t has to be in the opposite direction to the tangent velocity V_t, see Eq. (13), and the product $-F_t^T V_t$ is maximal when one seeks F_t inside its admissible domain $\mathcal{D}$. Here one has $\mathcal{D} = \{R \in \Pi : |R| \leqslant f.\mathbf{N}\}$ where Π is the common tangent plane at the contact point. Incidentally, it is understood here that $\mathbf{N}$ is known. One sees that $\mathcal{D}$ is the projection of a section of the friction cone $\mathcal{C}$ on Π. In particular, $\mathcal{D}$ is *convex*, which allows one to use the tools presented in the introduction of Section 3.4, replacing K by $\mathcal{D}$. With this in mind, it is possible to show (see, eg, Eqs. (14), (19), and (20)) that Coulomb friction can be formulated as

$$-V_t \in \partial \psi_{\mathcal{D}}(F_t) \tag{23}$$

and this inclusion permits encompassing the sticking modes as well. The reader may wish to check that the inclusion in (23) really represents Coulomb's law of friction, using the definition of the subdifferential of an indicator function (that is nothing else but the normal cone to $\mathcal{D}$ at F_t). It is also possible to rewrite Eq. (23) in terms of a dissipation function $\varphi(V_t) = f.\mathbf{N}.|V_t|$ so that $-F_t \in \partial\varphi(V_t)$. The function $\varphi(\cdot)$ is called a superpotential or pseudopotential (by analogy with the classical definition of forces that derive from a potential $F = \nabla V$, where this time the equality is replaced by an inclusion). In generalized coordinates as used in the works described in Section 7.4, the maximum dissipation principle is formulated as $F_q = \arg\max_{z \in \mathcal{C}(q)}(-z^T \dot{q})$, where F_q is the generalized contact reaction impulse and $\mathcal{C}(q)$ is the gener-

alized friction cone. The velocity $V^{(j)} \in \mathbb{R}^3$ at a contact point $\mathcal{A}_j$ and the generalized velocity are related as $V^{(j)} = J^{(j)}(q)\dot{q}$ whereas from the principle of virtual work $F_q = J^{(j),T}(q)F^{(j)}$, where in our notations $F^{(j)} = \binom{F_t^{(j)}}{\mathbf{N}^{(j)}}$. $J^{(j)}(q)$ is a Jacobian matrix.

Remark 5:

- More general frictional characteristics can also be written in a form as in (23). However, when the considered set K is convex one can rewrite Eq. (20) as a variational inequality as in (21). If K is not convex as it is the case for more complex friction models, one has to resort to hemivariational inequalities, as introduced by Panagiotopoulos [105].

- The second set of conditions in (13) with $V_t = 0$ is equivalent to $-\dot{V}_t \in \partial\psi_{\mathcal{D}}(F_t)$ [83,92] and (15) $\Leftrightarrow -\dot{V}_t \in \partial\psi_{\hat{\mathcal{D}}}(F_t)$. The acceleration formulation is then useful to derive the dynamics of rolling (sticking) contacts, see Section 3.3.

- As we saw above, the disc $\mathcal{D}$ can be approximated by a convex polytope $\hat{\mathcal{D}}$ [80,83]. This is an idea that is used in some algorithms to be presented later, see Section 7.4. However, the formulation in Eq. (23) remains valid, even with the modified friction cone. The advantage is that one deals with LCPs instead of NCPs. The case when $\hat{\mathcal{D}}$ is a square (so that the cone is approximated by a pyramid) corresponds to the set K in Fig. 1a. Actually basic results from convex analysis can be used to derive complementary slackness conditions (the best way to understand this is to look at [30] (Ch 23) corollary 23.8.1 and bottom of page 226, and [83]) from the polyhedral approximation of (23) or its acceleration formulation. Some event-driven and time-stepping algorithms that we will describe later hinge on such convex analysis tools to derive suitable (from a numerical analysis point of view) formulations of contact laws.

- The maximum dissipation principle is extended in [90] to more general friction laws. Some existence results (in the quasistatic case) are provided for $(q, \mathbf{N}, F_t)$. As we shall see in Section 7.4, some authors formulate the problem in the configuration space, and $P(q, \dot{q}, \lambda)$ in (2) is written as $D(q)\beta$, with β satisfying $\psi(\beta) \leqslant f.\lambda$. The function $\psi(\cdot)$ (that defines $\mathcal{D}$) should be convex. Then the maximum dissipation principle reads [91]

$$\beta = \arg\min_{\psi(z) \leqslant f.\lambda} [\dot{q}(t^+)^T D(q)z]. \tag{24}$$

This can be shown to be equivalent to

$$0 \in fD(q)^T \dot{q}(t^+) + \mu \partial\psi(\beta)$$
$$\mu \geqslant 0, \quad f.\lambda - \psi(\beta) \geqslant 0, \quad \mu[f.\lambda - \psi(\beta)] = 0 \tag{25}$$

where μ is a Lagrange multiplier (or a slack variable), and λ is as in (2).

- The impact rule in Eq. (12) satisfies, also, a maximum dissipation principle, see (11). Its expression within the sweeping process framework has been recalled in Section 3.4.1. It can therefore also be transformed into a (convex)

quadratic program. Notice that both (11) and (12) can be written in terms either of the contact percussion p_k or of $\dot{q}(t_k^+)$, using (3).

When combined with a velocity-impulse formulation, the principle of maximum dissipation allows one to avoid some inconsistency problems during the numerical integration that are encountered in rigid body dynamics with friction, see Section 7.4. Let us note that (23) is often formulated with $V_t(t^+)$ (the right limit of $V_t(t)$), a choice motivated by numerical implementation [94].

3.4.3 Second-order formulations

The formulations presented below concern frictionless rigid body dynamics, ie, $P(q,\dot{q},\lambda)=0$ in (2). The free motion of the system is described in generalized coordinates by the ODE

$$M(q)\ddot{q}=\bar{Q}(q,p,t),$$

where $p\triangleq M(q)\dot{q}$ is the generalized momentum and $\bar{Q}(q,p,t)\triangleq Q(q,M^{-1}(q)p,t)$. It should be noted that this equation allows dissipative terms. Let us assume that $\phi(\cdot)$ is a smooth function, ie, ∂K (see (1)) is a smooth codimension one surface. From (2), one gets the equation

$$M(q)\ddot{q}=\bar{Q}(q,p,t)+\mu \tag{26}$$

where μ is a measure such that

$$\mu=\nabla\phi(q)\lambda,\ \ \lambda\geqslant 0,\ \ \text{almost everywhere} \tag{27}$$

and

$$\text{supp}(\mu)\subset\{t:\phi(q(t))=0\}. \tag{28}$$

Roughly speaking, relation (28) means that the contact forces are switched on only when the constraints are saturated and is therefore equivalent to the complementarity conditions in (2). The measure differential equation (26) does not provide a complete description of the motion and we must add an impact law in order to describe the transmission of kinetic energy at impacts. We assume a Newton's law for generalized momenta. More precisely, whenever the constraints are saturated, the tangential part of p is conserved while the normal part is reversed and multiplied by a restitution coefficient $e\in[0,1]$. Here the normal and tangential part of p are taken in the sense of the local metric for the space of generalized momenta, which is defined by $M^{-1}(q)$. Let us denote by $\Pi_q p$ the projection of p on the tangent plane Π to ∂K at q with respect to the local metric, ie, $\Pi_q p=p-\nabla\phi^T M^{-1}p/\nabla\phi^T M^{-1}\nabla\phi\nabla\phi$. The complete formulation is finally given by Eqs. (26), (27), (28), and

$$p(t_k^+)=\Pi_{q(t)}p(t_k^-)-e(\mathrm{I}-\Pi_{q(t)})p(t_k^-)\ \ \text{if}\ \ \phi(q(t))=0 \tag{29}$$

and

$$\text{supp}(\mu)\subset\{t:\phi(q(t))=0\},$$

$$\mu=\lambda\nabla\phi(q),\ \ \lambda\geqslant 0,\ \ \text{almost everywhere}. \tag{30}$$

One notes that the impact law in (29) is equivalent to the ones presented in Section 3.2 ([18] claim 6.1). Let us denote

by $T(t_k^\pm)$ the kinetic energy at an impact time $t_k^\pm$ and define $|p|_{q(t)}^{*}=p^T M^{-1}(q)p$, the norm of p in the local metric of generalized momenta. One has:

$$
\begin{aligned}
T(t_k^+)&=\tfrac{1}{2}|p(t_k^+)|_{q(t)}^{*2}\\
&=\tfrac{1}{2}\{|\Pi_{q(t)}p(t+0)|_{q(t)}^{*2}+|(\mathrm{I}-\Pi_{q(t)})p(t_k^+)|_{q(t)}^{*2}\}\\
&=\tfrac{1}{2}\{|\Pi_{q(t)}p(t_k^-)|_{q(t)}^{*2}+e^2|(\mathrm{I}-\Pi_{q(t)})p(t_k^-)|_{q(t)}^{*2}\}\\
&\leqslant T(t_k^-)
\end{aligned}
\tag{31}
$$

and the equality holds if and only if $e=1$. Hence this model is energetically consistent.

In the particular case of a constant mass matrix, we can consider the generalized coordinates defined by $u=M^{1/2}q$, which makes sense since $M=M^T>0$. Then $p=M^{1/2}\dot{u}$ and with this new coordinates the impact law reduces to

$$
\begin{aligned}
\dot{u}(t_k^+)&=\dot{u}_T(t_k^-)-e\dot{u}_N(t_k^-)\\
&=\text{whenever the constraints are saturated}
\end{aligned}
\tag{32}
$$

where $\dot{u}_T$ and $\dot{u}_N$ are the tangential and normal parts of $\dot{u}$, with respect to the Euclidean metric. The set of admissible positions is then described in terms of these new coordinates by $K=\{u:\phi(M^{-1/2}u)\geqslant 0\}$. If K is convex with respect to u, we can replace the relations (26)–(28) by the following differential inclusion:

$$-\ddot{u}+f(t,u,\dot{u})\in\partial\psi_K(u) \tag{33}$$

where $\partial\psi_K$ is the subdifferential of the indicator function of K and can be interpreted similarly as the right-hand-side of (22), replacing $V(q(t))$ by K and the right-velocity by the position, see Fig. 1a. Therefore, Eq. (33) is a second-order differential inclusion, and the total dynamics is given by (32) and (33), or equivalently (29) and (27), (28). Notice from (33) that the inertial forces are therefore a subgradient of the indicator function $\psi_K(u)$. From (22), one sees that Moreau's idea has been to consider them as subgradients of $\psi_{V(q(t))}(v(t^+))$ instead.

4 RIGID BODY VERSUS COMPLIANT MODELS

4.1 Rigid body models: Some properties

First of all it is necessary to make it very clear what is meant by *rigid* and by *flexible*, see [24] for a discussion. By rigid one may mean any model that uses velocity-impulse formulation for the collisions (like the Darboux-Keller or Routh approaches). However, this does not necessarily imply an algebraic form of the shock dynamics as in Eq. (3). In other words, what is called the rigid body approach does not preclude the incorporation of deformation effects, though the way they appear in the dynamics may differ from the usual spring-dashpot formulations. This is the case of the Darboux-Keller's shock equations [18]. Clearly, in general, a model may mix rigid body modelling features (complementarity conditions, restitution coefficients) and compliance (deformation of the bodies that contact). In this section compliant models are to be understood essentially as those models that give rise to ODEs with switching vector field and constant

dimension, but not to MDIs as in Eq. (2). Studies based on continuous mechanics for the bodies modelling and contact complementarity conditions [60] therefore do not fall into the type of compliance discussed here.

Hyperstatic systems: In the case of hyperstatic systems (95% of all real systems) the Lagrange multipliers cannot be calculated uniquely, as is well known. However, the accelerations are uniquely determined [43,70,71]. The reader may think of a chair on four legs standing on rigid ground. Clearly, for reasonable applied external forces–ie, no *strange* forces that create right-accumulations of impacts like in Bressan's counter-example [106]—the acceleration of the chair is unique since it detaches if the normal force compensates for gravity, or it stays at rest if it does not. This shows that if one is not interested in computing exactly the contact forces, such inconsistency is not bothering.

Energetical behavior: The energetical coherence can be in some cases guaranteed by the dissipativity properties of the schemes. At least this holds for certain choices of the restitution coefficients ($e=0$ in case of sweeping process with friction [96], or same value of normal e and tangential e_t restitution coefficients for Anitescu *et al* [7,84] in case of multiple impacts). In [55], an estimation of the (discrete-time) energy is also made for all $e \in [0,1]$.

Impact detection: Some schemes do not require an explicit calculation of the impact times: convergence of the discretized solutions towards a solution of the continuous problem is assured so one does not need a special module of detection. Decreasing the (constant) integration step automatically guarantees the improvement of the numerical result. This is the case for the time-stepping schemes described in Sections 7.2, 7.3, and 7.4. In other words, such schemes do not require (nor provide) the accurate determination of the times t_k and a specific (possibly time consuming) procedure of *local* computation. They will therefore be suitable when one does not desire a very accurate knowledge of the dynamics at the contact points and/or the time of contact, but instead is content with a global picture of the system (eg, detection of particular orbits or attractors).

Restitution laws: The restitution coefficients need not at all to be constant. They can incorporate dependence with respect to initial relative velocity, material properties, shapes, dimensions etc, which take into account the influence of global and/or local deformations and sources of kinetic energy loss during the shock. The derivation of such coefficients has been and is still the object of research works. In particular, vibrational effects and transmission of energy in multiple impacts are quite fundamental issues. Some recent results are very promising [107–109].

LCPs: The resolution of LCPs may create some problems, however this is a topic that supersedes mechanics [110,111] and which is the object of many research works, see Sections 3.1 and 6.6.

Global motion analysis: It is known that nonsmooth complementarity modeling allows one to simplify the analysis of motion and to permit the study of first-return 1D Poincaré maps, which would not have been possible with compliant models of contact [112].

One major and still largely open modelling problem is that of multiple impacts with or without friction.

4.2 Compliant models

Spring-dashpot model: A compliant model that is often used is a spring-dashpot model. The linear spring-dashpot model possesses some strange properties which are often ignored or tolerated in the literature. Let us consider the following dynamics:

$$m\ddot{q} = \begin{cases} -kq - d\dot{q} & \text{if } q < 0 \\ 0 & \text{if } q \geq 0. \end{cases} \tag{34}$$

Let us consider $m=1$, $d=3$, $k=1$, $q(0)=0$, $\dot{q}(0)<0$. The values of the parameters are chosen such that the mass m crushes the spring-dashpot and there is no subsequent detachment: this is a plastic impact. Numerical calculations yield $F = -kq - d\dot{q} = 6.86\exp(-2.62t) - 0.14\exp(-0.38t)$ and this contact force can take negative values which would correspond to some "gluing" effects (which obviously are not the goal of this model). However the impulse of $F(t)$ is positive, which explains why this model provides a limit (as k and d diverge to $+\infty$ in a certain way) that is compatible with rigid body collision modeling [18,53].

Compliance location: The first point is about which sort of compliance is to be modeled: local deformation at the contact point? Or global deformations distributed in the bodies or in the kinematic chain (finite-element-like models [60,109,113–116], or elasticity at joints [117] for instance) to model the vibrational effects? Or both? A discussion on the relationships between compliant models and rigid body models can be found in [24]. In [118], the domains of application of various rigid body (also called stereocontact) models and compliant contact/impact models (Newton's, Poisson's restitution, Hertz', Saint Venant's impacts) are discussed from their respective time-scales. It is argued that multi-rate/multi-method integration codes provide better results. A simulator has been implemented that uses Gear's multi-rate integration [119]. This type of arguments is consistent with [120] who proposes to mix multi-rigid body with finite elements methods, and also with the general message at the end of this survey.

Impact detection: Notice that the impact detection also exists in compliant models since one has to compute the position at which contact is made so as to switch to the new vector field. In case of stiff equations and too large integration steps the energetical behavior has to be carefully checked, because long run simulations may lead to wrong results (energy loss or energy gains).

Integration step: The integration step h has to be chosen satisfying $h \ll \mathcal{O}(1/\sqrt{k})$ where k is the contact stiffness since the time of penetration (say, of deformation) is itself $\mathcal{O}(1/\sqrt{k})$ [53]. If one wants, for instance, to calculate 100 points during the shock and with $k=10^{10}$ N/m (a physical value in many instances like gears, pinions, systems with joint clearance [121]), then $h = 10^{-7}$ s is required. If some real-time applications are required or if n is very large (it can be $\geq 10^4$ in granular matter, think of a mere sandpile), then this may be an obstacle. Notice that some authors recom-

mend the calculation of at least 1000 points for each collision to assure a good accuracy of long run simulations [10]. Clearly, real-time applications—or even fast enough numerical tests—are impossible with such integration steps, see example 1 below for numerical values. Another important point is the fact that, in practice, impact times may be quite close one to another [107], [[54], Ch 4], due for instance to micro-collisions. If one desires to approximate correctly such high frequency impacts, one is led to choose a sufficiently large k in order to numerically detect enough impacts. Such a choice for k is often difficult.

Example 1: Numerical results are reported in [121] who chooses a Hertz contact model with stiffness $k = 6.58$ 10^{10} N/m$^{1.5}$ (let us recall that such contact model supposes low approach velocities and is valid for central impacts only). A slider crank mechanism with joint clearance is simulated. A predictor-corrector routine is used based on Adams-PECE formulas. The simulation is performed over 30 s and its duration is 3.2 hours. A Baumgarte technique (see Section 5.2) is used to stabilize the kinematic constraints. The results are not compared to any experimental data, and they do not concern the long-term dynamics: how does the scheme behave on longer simulation times? Is such an approach realistic for long run prediction and for more complicated systems with several clearances and other unilateral contacts, both from the simulation length and the outcome accuracy? (Let us recall that certain mechanical systems like *simple* circuit breakers may include from 15 to 30 unilateral contacts, and from 7 to 20 bars: the identification of contact parameters and the sensitivity of a long run simulation outcome with respect to their numerical value may be a big problem). The simulation duration may be a serious obstacle in certain applications. If one wants to simulate the system on several minutes—and this may be needed in some instances—then the simulation length becomes a real problem.

Implicit algorithms may be preferred to explicit ones as it is known that they often provide better accuracy and stability for stiff problems. When an explicit code encounters stiffness, the integration step needs to be decreased to keep stability ([25] page 21). Implicit schemes have the tendency to filter out the high frequencies and therefore treat stiff ODEs as DAEs. Larger integration steps are allowed for the same accuracy. However, real-time applications may require fast enough algorithms that preclude the use of implicit discretizations (consequently of iterative algorithms).

Physical parameters identification: In practice, the identification of the physical parameters (damping, stiffness, or any other coefficient that appears in the contact model) may be quite difficult. Furthermore it should be recalled that some well known models like Hertz hinge on a particular stress law and are valid only for central impacts, for a simple shock between two bodies (ie, the line joining centers of gravity coincides with the normal direction at the contact). Collisions may be far from central in some applications, like colliding kinematic chains. Moreover, the influence of the parameters numerical values on the long run simulation outcome may be important. In certain simple cases of multiple

shocks (such as Newton's cradle), it can be shown that the collision outcome is quite sensitive to such parameters values [18].

Remark 6 (Coulomb friction): The same type of discussion could be done about the friction model. All the algorithms described in this paper and which incorporate frictional effects use the Coulomb friction model. On one hand it is already difficult enough to properly identify and discretize this model, so that there is no need, from this point of view, to complicate things more. And, as we pointed out in Section 3.2, the main problem that we face is not the complication of the dry friction model for one contact, but its formulation and time-discretization for multiple contacts. On the other hand, this model, though simple, incorporates already a lot of physical phenomena and proves to be sufficient in many cases. In view of the large spectrum of mechanical systems that have been simulated with the simplest model, ranging from granular matter to buildings made of blocks and deep drawing processes, there seems to be little need to use a more complex model. In particular, the sticking modes are quite important in many applications (... *the physics points to real discontinuities, and there is little advantage numerically in smoothing the discontinuity. The discontinuity is here to stay* [91]). The simulation of a pile of rigid blocks will usually not require anything much more complex than Coulomb's law, because the phenomena other than sticking and sliding do not play a major role in the motion. However, it will necessarily involve complementary slackness conditions for both normal and tangential directions, and consequently the need for complementarity problem solvers.

Any more sophisticated model than Coulomb's law that is to be incorporated in the framework of the developments in this paper, should at least satisfy the principle of maximum dissipation, both for physical reasons (what is the physical validity of a model that is not dissipative with respect to the supply rate $V_t^T F_t$?), and for mathematical programming reasons (underlying complementary slackness conditions and variational formulations). Models of this sort have been studied in [122,123]. More discussions on friction models in non-smooth multibody problems can be found in [91].

4.3 Conclusions

The principal drawbacks of the rigid body approach are indeterminate and inconsistent configurations, plus the need for discovery of multiple impact rules. First of all, since the rigid body model has to be seen as a limit model (in a sense it is the model that contains the least physical information on the process) these peculiarities are not surprising: the less information, the worse prediction. However, the other approaches are not perfect, either, and possess their own drawbacks. In [57,124], calculations show that elastic impact of a particle in an angle with a penalization, implies essentially unpredictable results after two reflections. Moreover, physical details can be introduced in a rigid body model, through suitable restitution coefficients that may incorporate vibrational effects (global deformations) like in [107], local properties at the contact point, etc. See [18] and references therein. A lot depends on one's goals and on the domain of

application which most often requires finding a compromise between the model complexity and its tractability: eg, designing a stable feedback control algorithm or simulating the long-run motion of a complex kinematic chain using sophisticated compliant contact models seems unrealistic in most cases. On the other hand, fine calculation of contact forces is impossible with rigid body approaches since they rely on a two-time scale of the dynamics and do not care about forces during the shock but only about the impulses of forces and their effect on the velocity. Notice, however, that for some of the reasons listed above (physical parameters identification, unpredictable motion for multiple shocks, etc), using compliant models may not at all solve this problem.

In summary, some reasons that may motivate one to use rigid body modeling approaches are:

- Stiff ODEs are to be avoided for real-time applications or all tasks which require fast simulation (because stiff ODEs usually require implicit algorithms). The principal source of stiffness is the contact model. Hence, rigid body contact/impact modeling allows one to avoid stiff differential equations.
- Estimation of contact parameters for compliant models (stiffness, damping, etc) can be quite difficult in practice, and the collision outcome may be highly sensitive to such physical parameter values.
- Simple compliant models of contact may essentially be useful for mathematical aims (study of convergence of penalized problems[4]) and justification of rigid body restitution laws, [18,51,53,74,75,125], but not for numerical simulation due to some fundamental drawbacks [126].
- Rigid models serve as a convenient model for control design purpose (backlash, bipedal locomotion, manipulators during complete robotic tasks, liquid slosh phenomenon, etc) especially when the impact cannot be controlled because the input values should then be chosen too large (consequently the collisions are autonomous).
- Provide good predictions for long term motion simulation, especially if one is not interested in details of local collision (or contact) behavior that involve complex material characteristics, but rather in the effect of collision on a global scale, for example the manner in which vibrations progress from end-effector to base in a flexible robotic manipulator due to impact forces or impulses.
- It can provide the basis for a tool that is easier to use by design engineers.
- When n is too large (like in granular matter) penalization models with high stiffness become impossible to simulate due to too long simulation time (to say nothing about first and second items).

To end this part, let us cite Baraff in [127]: *Although the penalty method is useful in some contexts (namely largely static environments) it has become increasingly apparent that the performance of spring-and-damper systems for*

simulating rigid body motion is inefficient and has unpredictable accuracy in dynamic settings... Notice that compliant and rigid body approaches may also be mixed, see, eg, [120] where rigid body and finite element approaches are mixed for collision detection.

5 NUMERICAL ANALYSIS OF BILATERALLY CONSTRAINED MECHANICAL SYSTEMS

5.1 Introduction

From a general point of view, there are two main classes of impacting systems: the ones whose orbits undergo collisions at separated instants (ie, there exists $\delta > 0$ such that $t_{k+1} \geq t_k + \delta$ for all $k \geq 0$), and the ones for which this is not true (see also Section 6).

In the first case, there exist time intervals of strictly positive measure on which the dynamics are ODEs or DAEs. One also finds in this category, eg, simple systems with periodic impacting orbits (vibro-impact systems [6], bipedal locomotion systems [15], etc). However, even for very simple nonsmooth systems one cannot always simulate the motion by combining contact detection with DAE or ODE solvers. Consider a rigid ball falling to the rigid ground, where one models the elastic collision with Newton's collision law $0 < e < 1$. Then there exist an infinite number of collisions in a finite time-interval. In a variation to Zeno's argument, one can then argue that such a situation is impossible, since it requires an object to pass through an infinite number of collisions in a finite amount of time.

Obviously, the second class is the most general one, and the one which creates the most interesting challenges in terms of mathematical analysis and modeling. We shall come back later to which ingredients one needs to construct a good simulation software. But it is useful in this setting to recall first the particular features of DAE simulation.

5.2 Simulation of DAEs: General methods

For unconstrained dynamical systems, there are many discretization methods available. Discretization formulas for linear systems can be found in [128]. For nonlinear systems, there are many numerical methods available to solve Ordinary Differential Equations (ODEs), see, eg, [129,130]. This is not the case for Differential Algebraic Equations (DAEs): establishing solvers for DAEs is still a very active research area. Overviews of the theory can be found in [25,29].

Numerical simulation of the conventional formulation of a constrained mechanical system exhibits severe stability problems, already, for simple systems. Since simulation of multibody systems is an activity with a long history, several computational procedures have been proposed to overcome the stability problems. These include techniques where a distinction is made between dependent and independent variables (a solution is sought through singular-value-decomposition), equilibrium correction strategies [131], penalty formulations [132,133], coordinate partitioning methods [62], predictor/corrector algorithms [134], a differential algebraic approach [66,67], and projection methods [135].

---

[4]A fine examination of the planar impact of a particle in an angle [74,75] shows that a penalized problem (damped linear spring-dashpot) yields Moreau's rule Eqs, (11) and (12) for generalized soft impacts in almost all the cases (in a measure theoretical sense).

In the literature on numerical integration, DAEs are often characterized by their index [25,29,136]. Roughly speaking, the index equals the number of times the constraints must be differentiated to arrive at a set of ODEs. The index can be viewed as a measure of how far a DAE is from being an ODE. DASSL, designed by Petzold in the early eighties, is capable of solving DAEs which have a low index [137]. Constrained mechanical systems often have an index equal to three [29]. It is well known that index three systems can not be solved directly by standard ODE solvers [135,138]. The dynamics in Eq. (7) have been used in [67] to reduce the index. A multistep (BDF) method with variable step-size, combined with a Newton iteration root-finding algorithm, has been shown to converge. For constrained mechanical systems, stable numerical algorithms are few, and are usually available as research code only. A summary of DAEs simulation algorithms applied to bilaterally constrained mechanical systems can be found in [26] (Ch 7.2).

In engineering practice, the constraint stabilization technique presented by Baumgarte [131] is often applied because it is conceptually simple and easy to implement. Differentiating the position constraint twice gives $\ddot{\phi}(q) = 0$. It is well known that the numerical solution of this equation can be unstable, and can lead to values of $\phi(q)$ and $\dot{\phi}(q)$ that are far from the desired value zero. The modified acceleration equation: $\ddot{\phi}(q) + 2\alpha\dot{\phi}(q) + \beta^2\phi(q) = 0$, is (asymptotically) stable for $\alpha > 0$. The additional terms in the latter equation can be seen to act as a proportional/derivative control with gains equal to 2α and β^2. Baumgarte also presented the proportional/integral counterpart, for the asymptotic stabilization of holonomic constraints [139].

One problem can readily be seen from the formulation of the stabilization technique: how to choose the coefficients α and β? Since the stabilization term can be interpreted as a proportional/derivative control law, it is noted that the use of the stabilization term shifts the poles of the system and alters its dynamic behavior. The choice of α and β is merely a matter of how fast we want to damp out the constraint violations. Large values of α and β lead to high-gain feedback laws. Note that the choice $\alpha = \beta$ yields a critically damped system. It is this choice that is commonly used when Baumgarte's method is applied. In [140], the gains 2α and β^2 are related to the step size that is applied in the numerical algorithm. There it is remarked that their choice of gains tends to damp out constraint violations faster than any other choice, but accumulation of (integration) errors cannot be prevented. Furthermore, decreasing the step size results in larger gains. As a result, the damping terms dominate the numerical solution process: they make the system become numerically stiff. A further analysis of Baumgarte's method can be found in [141] and [142]. In spite of these drawbacks, the constraint stabilization technique of Baumgarte is often applied since it avoids iterative solution of algebraic constraints. This is in contrast to, for instance, a predictor/corrector algorithm and some of the other methods. For instance, the projection method proposed in [135] uses a combination of the numerical solvers known as Backward Difference Formulas (BDFs) and a Gauss-Newton projection method. These algorithms require iteration processes to obtain values within a certain predefined error level: a number of corrector steps must be applied.

5.3 Application to mechanical systems

During simulation studies of unilaterally constrained mechanical systems an expression of the Lagrange multiplier is often used as a model for a force sensor, or simply as a nonlinear expression for the contact force, expressed in terms of the motion on the constraint manifold and the control input [143]. Let us start from Eq. (5) and add a control input w in the right-hand-side. Under the assumption that $\nabla\phi^T(q)$ has full-row rank, and omitting friction terms, one can obtain:

$$\lambda = -[\nabla\phi^T(q)M^{-1}(q)\nabla\phi(q)]^{-1}(\nabla\phi^T(q)M^{-1}(q)$$
$$\times(w + Q(q,\dot{q})) + \overbrace{\nabla\phi^T(q)\dot{q}}).$$

(35)

In [66], it is noted that, although Eq. (35) provides an analytical expression for the Lagrange multiplier, it is not always a good starting point to obtain a numerical solution. The presence of numerical errors leads to violation of the constraints and eventually leads to a drift-off from the constraint manifold. As a consequence, one obtains physically meaningless solutions. Usually, for consistent initial conditions, it will take a longer period of time for the drift-off to become noticeable. And if the drift-off remains small, then the corresponding approximate solution may well be acceptable. But, generally speaking, a growing drift-off can not be accepted. Striking examples of drift-off can be found in [66] and [135].

For ease of notation, let us examine drift-off in a first-order formulation (which can be obtained, for example, starting from (6) by setting $x^T = (q^T, \dot{q}^T)^T$, and redefinition of terms).

$$\dot{x} = f(x) + g(x)w + g(x)C^T(x)\lambda$$
$$p(x) = 0.$$

(36)

Instead of finding a discrete-time expression for λ directly from (36) another sequence of steps is advocated in [66]. This sequence of steps can be described as *discretize first–substitute next–combine later*. This approach to simulation has been applied to restricted ODEs in [66] and to boundary value problems of Partial Differential Equations (PDEs) in [144]. But the original idea can already be found in [145] where it is applied to index one systems with linear, stationary constraints in combination with the Forward-Euler integration method.

First the equations in (36) are discretized. We will use the Forward-Euler method only to illustrate the concepts, although in general it is not advisable to simulate mechanical systems using the Forward-Euler method. This gives

$$\begin{cases} x_{i+1} = x_i + h[f(x_i) + g(x_i)w_i + g(x_i)C^T(x_i)\lambda_i^d] \\ p(x_{i+1}) = 0. \end{cases}$$

(37)

Note that the constraint is treated in an implicit manner. The notation λ_i^d is used to distinguish the discrete Lagrange multiplier from the analytical one in eg, (35). The idea now is to obtain a discrete formula for λ_i^d directly from Eq. (37). Let ϵ_i denote the numerical error at step i, and let us assume here that ϵ_i is a constant ϵ. First assume that the matrices C and $H = \partial p / \partial x$ are constant. The discrete generalized Lagrange multiplier λ_i^d is given by

$$\lambda_i^d = -(CC^T)^{-1}\left(H[f(x_i) + g(x_i)w_i] + \frac{p(x_i)}{h} \right). \quad (38)$$

We will refer to $p(x_i)/h$ as a compensation term in the remainder. One can now obtain [66]:

$$p(x_{i+1}) = hCC^T\epsilon. \quad (39)$$

No error accumulation can take place. Indeed, if $h \rightarrow 0$ one has $p(x_{i+1}) \rightarrow 0$, as desired.

In contrast, if one used the analytical expression of the Lagrange multiplier as in Eq. (35) it is shown in [66] (see also [65], Ch 9) that one obtains, with neglect of the higher-order terms:

$$p(x_{i+1}) = p(x_0) + h \sum_{j=0}^{i} C(x_j)C^T(x_j)\epsilon_j. \quad (40)$$

From the latter equation, it follows that once an error is made the solution is not on the constraint manifold. Now again consider the special case where $\epsilon_i = \epsilon$, $\forall i \in \mathbb{N}$, with ϵ constant, and that C is a constant matrix. This gives

$$p(x_{i+1}) = p(x_0) + t_i CC^T\epsilon. \quad (41)$$

Note that it makes no sense to let $h \rightarrow 0$ since h does not even appear in Eq. (41). And if $t_i \rightarrow \infty$, for instance because we are interested in an equilibrium solution, one even has that $p(x_{i+1}) \rightarrow \infty$! Even if the initial conditions are consistent with the equality constraints, error amplification is inevitable due to the presence of numerical errors. Each error source will contribute to the drift-off. This is one of the reasons why many simulation codes project the state on a time t_{i+1} back to the constraint surface for example using Newton-Raphson, before proceeding the time simulation.

The expression for the discrete generalized Lagrange multiplier λ_i^d is useful also in combination with other explicit and implicit integration routines, and for nonlinear constraints as well. For this one treats the compensation term as a constant on an interval $[t_i, t_{i+1})$ and evaluates all other functions in the points needed by the numerical method that is applied. For application to mechanical systems, the expression for the Lagrange multiplier on intervals $[t_i, t_{i+1})$ now reads:

$$\lambda_i^d = -[\nabla \phi^T(q)M^{-1}(q)\nabla \phi(q)]^{-1}\bigg(\nabla \phi^T(q)M^{-1}(q)$$

$$\times (w + Q(q,\dot{q})) + \overline{\nabla \dot{\phi}^T(q)}\dot{q} + \frac{\nabla \phi(q(t_i))\dot{q}(t_i)}{h} \bigg).$$

$$(42)$$

Using Eq. (7) one also obtains:

$$\mu_i^d = -[\nabla \phi^T(q)\nabla \phi(q)]^{-1}\left(\nabla \phi^T(q)\dot{q} + \frac{\phi(q(t_i))}{h} \right). \quad (43)$$

Observe that analytically the constraints are strictly satisfied, and one has $\lambda_i^d = \lambda$ and $\mu_i^d = \mu$.

Using the "discretize first—substitute next—combine later" sequence of steps, a numerical method can be obtained that has the property that it is robust with respect to errors in the initial conditions, and stable with respect to errors made during numerical integration. Moreover, the use of the discrete generalized Lagrange multipliers λ_i^d and μ_i^d does not yield numerically stiff equations when the time-step is reduced: the term h in the denominator is canceled again. Since the compensation term is fixed on each time-interval, it is easy to combine the method with variable stepsize methods (see [146,147] for a variable stepsize method). The numerical method can be extended to cover constraints $\phi(q,t) = 0$, ie, constraints that depend explicitly on time. These constraints arise, for instance, in dynamic path planning of robotic manipulators, and also in flight-path management of airplanes. Most of the theory remains valid, but in the case of, for example, Runge-Kutta-4 there are additional requirements on higher derivatives of the constraints that need to be satisfied. This still leads to stable numerical simulation of bilaterally constrained dynamical systems.

6 NUMERICAL ANALYSIS OF UNILATERALLY CONSTRAINED MECHANICAL SYSTEMS

6.1 General motivations for new specific schemes

The main problem in simulating DAEs is the stabilization of the constraints. For MDIs, additional difficulties occur:
- Complementarity conditions are an essential part of the dynamics and have to be carefully incorporated into the numerical scheme. In particular, Baumgarte's stabilization technique is not suitable since the resulting multipliers' signs are physically meaningless, and cannot be used (even in the codimension one constraint case) to detect release.
- Detection of contact instants and re-initialization of the state. In the case where one is able to detect analytically what the precise impact point and time are, there are still problems in the numerical approximation of these instances. First, one must be sure that an impact is not missed (see Section 6.5). Next, since machine zero is not identical to zero, there will almost always be a small mismatch between, for instance, actual impact location and simulated impact location. It must then be decided whether a collision will take place or that the bodies remain in contact with each other. Clearly, *a priori* determination of all possible contact points, necessarily including the velocity information as done in [148] and [149], is an asset. Last but not least, one must decide whether or not, after constrained motion a release takes place. Since zero force level can not be maintained numerically and since DAE solutions do not yield exact satisfaction of a constraint, also in this case *a priori* determination of all possible release points, including the velocity and force information

is a benefit [149]. In Section 6.4 we will come back to these difficulties in relation to event-driven software schemes.

- Accumulation point of the sequence $\{t_k\}_{k\geqslant 0}$ occur frequently. One has to impose a threshold or derive local analytical forms of the impact Poincaré map [150]—which are similar to the bouncing ball dynamics. For $n=m=1$, the transition to the permanently constrained mode can be decided via a threshold. For $n\geqslant 2$ and/or $m\geqslant 2$, things complicate because generally the coordinates are dynamically coupled. If one constraint is decided to be saturated because numerically one is unable to detect rebounds any longer (ie, computer accuracy may be attained), then the overall motion may be drastically modified. In general, the methods based on impact detection will provide good results for:
  —Periodic motions with finite number of impacts per period, for all n, m.
  —Periodic motions with finite or infinite number of impacts per period, for $n=m=1$ [151–153].

The multiplicity[5] of the shock is not easy to calculate, see works on Newton's cradle [154] and on impact of a particle with an angle [73,155].

6.2 MDIs are not concatenation of ODEs, DAEs, MDEs

It is important to realize that an MDI as in Eq. (2) is quite different from ODEs, DAEs, DIs *à la* Filipov, MDEs[6] of the form

$$\dot{x}=f(x,t)+g(x,t)\dot{w},\quad x(0^-)=x_0 \tag{44}$$

where $x\in\mathbb{R}^n$, and the control input $w\in\mathbb{R}^m$ is of bounded variation, and even MDEs as in [156], which are defined as:

$$\begin{cases} \dot{x}=f(x) & \text{if } t\neq t_k \\ x(t_k^+)=x(t_k^-)+a(x(t_k^-)) & \text{if } t=t_k \\ x(0^-)=x_0 \end{cases} \tag{45}$$

for a sequence $\{t_k\}_{k\geqslant 0}$, where $t_k\rightarrow +\infty$ as $k\rightarrow +\infty$, t_k may or may not depend on x[7]. The following fundamental discrepancies between the last two differential equations and our MDIs are [18,87]:

- It is clear that the overall dynamics with complementary-slackness mechanical systems is far from being only a differential equation with impulsive inputs. The way the set of indices for active constraints and sticking/slipping contacts are refreshed is a fundamental part of the (hybrid) system [157]. Only the very simplest dynamics of systems as in Eq. (2) seem to resemble Eq. (45). But they are of different nature, as explained below.

- The commutativity property of the vector fields $g_i(x)$ $\in\mathbb{R}^n$, $i\in 1,\ldots,m$ allows one to decouple the MDE $\dot{x}=f(x)+g(x)\dot{w}$ into simpler systems for which existence and uniqueness of solutions are easier to prove. These commutativity properties are always satisfied for mechanical systems, but they do not at all solve the problem of multiple impacts [18].

- If $f(x)$ and $a(x)$ in Eq. (45) are linear then the whole nonsmooth system is linear as well. On the contrary, linearity of the continuous vector field in a mechanical system subject to a unilateral constraint does not at all imply linearity of the total system (see [18], Ch 1).

- The system in Eq. (45) and the developments in [156] do not allow for finite accumulation points of discontinuities.

- The product $g(x)\dot{w}$ is quite meaningful for unilaterally constrained mechanical systems because $g(x)$ contains only positions and only the velocities are discontinuous [18,87]. Otherwise specific tools have to be developed to render Eq. (44) meaningful. The formulation in Eq. (45) is a way to overcome such problems.

- The abundant literature on (Lyapunov) stability properties of impulsive differential equations as in Eq. (44) deals only with the stability of the fixed point of $\dot{x}=f(x)$, where the remaining terms are considered as disturbances. Stability studies of unilaterally constrained mechanical systems require other tools. Mainly because [18–20] *i*) they are hybrid dynamical systems with different modes (in other words there is a natural Discrete Event part whose states correspond to the modes associated to the index sets of sticking and sliding contacts), *ii*) the fixed point of the overall system may not at all correspond to the one of the continuous dynamics (think of the bouncing ball example in which the vector field $\ddot{q}=-g$ does not even possess any fixed point!), and *iii*) the stabilization goals may be quite different depending on the task (stabilizing jugglers or hopping robots [14] is totally different from stabilizing a manipulator along a time-varying trajectory including free-motion, constrained motion and impacting phases [19,20]).

- Nonsmooth mechanical systems as in Eqs. (2) and (3) have solutions which are generally discontinuous with respect to initial conditions [51,39]. This is not the case for MDEs as in Eq. (44) as can be easily seen taking, for instance, $g(x,t)=g(t)$ [158].

- A lot of work has been dedicated to simulation of bilaterally constrained mechanical systems, since they are index 3 DAEs (see [25] for further references and see Section 5). But as we said above for hyperstatic systems, LCPs have to be solved at each step and integrating switching DAEs is far from being sufficient. Baumgarte's method is not suitable since it does not guarantee that the sign of the multiplier has any physical meaning.

In summary, the MDIs in Eqs. (4), (22), (26)–(28), and (32)–(33) are quite different in nature from the other types of MDEs. They are not a simple concatenation (except in very simple cases) of ODEs and/or DAEs separated by re-initializations of the state (the velocity). In order to reinforce

---

[5]The multiplicity is to be understood here as the number of shocks that the system undergoes when it strikes in a neighborhood of the singularity of ∂K. It should perhaps better be called the order O of the collision, which can be simple or multiple. In this sense, the ball falling on a ground undergoes, in finite time, a simple shock of order infinity. A 2D particle striking an angle undergoes a 2-shock of finite or infinite order depending on the angle, initial data, restitution coefficients. Actually, these two notions require more accurate definitions: if $O<+\infty$ then one just faces a series of simple shocks. So the only relevant case is $O=+\infty$.

[6]Measure Differential Equations.

[7]Evidently k here is an index and has no relationship with the stiffness in Section 4.2.

these ideas, let us consider the following third order system ([39], Ch 7.3; [159]), which belongs to the class of Linear Complementarity Systems [42,160]:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \dot{x}_3 = \lambda \\ \lambda \geqslant 0, \ x_1 \geqslant 0, \ \lambda x_1 = 0 \\ x^T(0^-) = (0, -1, 0). \end{cases} \quad (46)$$

The state has to jump initially to the value $x^T(0^+) = (0,0,0)$, and with $\lambda(0) = 0$. Then the system stays at rest keeping the same state for all $t \geqslant 0$ (note that the DAE corresponding to Eq. (46) reduces to the trivial system $x \equiv 0$, $\lambda \equiv 0$). Let us apply the following backward Euler (implicit) scheme to the system in (46):

$$\begin{cases} \dfrac{x_{i+1} - x_i}{h} = A x_{i+1} + B \lambda_{i+1} \\ x_{1,i+1} \geqslant 0, \ \lambda_{i+1} \geqslant 0, \ x_{1,i+1} \lambda_{i+1} = 0. \end{cases} \quad (47)$$

The matrices A and B are easily identifiable from (46) and x_i denotes the discretized value of x at step i. Roughly, one constructs a LCP whose unknown is λ_{i+1} by inserting $x_{1,i+1}$ into the complementarity conditions (which requires that $(I - hA)$ be invertible, which is true for h small enough). This way of doing is classical in time-discretization procedures, see Sections 7.1 and 7.4. Under certain conditions (satisfied for this example) the constructed LCP(λ_i) in Eq. (47) possesses a unique solution for sufficiently small h. The initial state jump is computed by solving the LCP at the first step, and one finds $(x_{1,0}, \lambda_0) = (0, 1/h^2)$. Then the solution should converge to the null solution. The application of the algorithm in Eq. (47) yields [39]: $(x_{1,i}, \lambda_i) = (i(i+1)/2h, 0)$ for $i \geqslant 1$. Moreover it can be shown that the nonimpulsive part of x_1 satisfies $\|x_1\| \geqslant N_h(N_h - 1)/2 h^{3/2}$, where N_h is the integer part of T/h for some $T > 0$ and $\|\cdot\|$ is the Euclidean norm. Therefore x_1 diverges as $h \to 0$!

Following [72], let us consider linear DAEs. If the initial values are inconsistent (ie, they do not satisfy the constraints), then fixed-step k-step backward differentiation formulas (BDF) produce solutions which may diverge during the first steps as $h \to 0$, for $k \leqslant 6$. However, they are $\mathcal{O}(h^k)$ after a maximum of $(\tilde{m} - 1)k + 1$ steps, where $\tilde{m}$ is the DAE index. Therefore, the example in Eqs. (46) and (47) shows that one should be very cautious when applying an implicit Euler discretization to a nonsmooth system as in (46): the results of convergence which are available for ODEs or DAEs may absolutely fail when considering complementarity systems. Indeed it is really the complementarity conditions (and consequently the LCP(λ_i)) which create the divergence phenomenon. The initial jump drives the state away from the constraint, but convergence is not recovered afterwards. Fortunately, this sort of problem is shown not to occur for linear mechanical systems in [39,159], and more generally for dissipative systems (see [117] for details on dissipative systems theory). The example in Eqs. (46) and (47) demonstrates that the existence of solutions to the (discrete-time)

LCP(λ) does not imply at all the stability of the overall integration scheme nor convergence of its solution as $h \to 0$. This fact is quite similar to what occurs for the continuous-time dynamics, [18] Remark 5.18. Another example that demonstrates the importance of the (discretized) complementarity conditions can be found in [91]. It concerns a bouncing ball, whose dynamics is in Hamiltonian form with state (q,p), discretized by a implicit Euler midpoint rule. If one uses $0 \leqslant \lambda_i \perp (q_{i+1} + q_i)/2 \geqslant 0$, then the effective restitution is not elastic and the energetical behavior is desastrous. If one uses $0 \leqslant \lambda_i \perp (p_{i+1} + p_i)/2 \geqslant 0$, the energetical behavior is perfectly elastic (this had also been noticed in [28,96]).

It has been stated [161] that "Differential/Algebraic equations are not ODEs." We conclude that: *Measure Differential Inclusions that represent complementary-slackness mechanical systems are not concatenations of ODEs and DAEs, neither MDEs.*

6.3 Simulation algorithms

It is difficult to classify the existing methods of numerical integration since there are many different criteria that can be used to fulfill such a task: acceleration-force/velocity-impulse; computation of the contact forces or not; resolution of LCPs/NCPs for all times, at certain times, never; frictionless constraints; Coulomb friction, regularized Coulomb; Signorini conditions; polyhedral approximation; revolution cone of friction; 2D/3D order of the discretization, contact point; finite-element methods, quasistatic/dynamic, inclusion of complementarity conditions in the discretization; calculation of impact instants or not; proof of convergence towards a solution of the rigid body dynamics or not; inclusion of multiple shocks; implicit or explicit form of the numerical scheme etc. Following [101], we choose the following classification in this paper (another classification is chosen in [115]):

- Event-driven schemes: one integrates the DAEs between state re-initializations. One needs some basic *ingredients* that we will describe below.
- Time-stepping schemes: one considers a time-discretization of the nonsmooth dynamics (MDI) including complementarity conditions and impact rules. The whole set of discretized equations/conditions is used at each step to calculate the future state.
- Penalized-constraint schemes: the unilateral constraints are replaced by some compliant model of the contact.

Consequently, we will first describe the main ingredients that any good event-driven software should incorporate, and then we shall describe the most advanced methods that have been proposed in the recent years by several authors and which incorporate (implicitly or explicitly) such modules. Evidently we do not deal with penalized schemes and stiff ODEs here, as announced before. The problem of discretization of the continuous dynamics (ODEs or index 3 DAEs) has been discussed in Section 5, as well as the problem of constraint stabilization.

6.4 Overview of modules for event-driven schemes

It is clear that there is a need for simulation tools that provide routines that can be used to simulate MDIs. In the remainder of this section we will discuss some of the ingredients that should be incorporated in event-driven simulation softwares. It must be stated that even for bilaterally constrained mechanical systems, commercial tools are not abundant and in many cases the numerical code that implements the transition rules and discontinuities due to uncontrolled collisions must be supplied by the user. In this section an overview of different modules is given.

In order to simulate unilaterally-constrained dynamical systems, one needs to know whether or not there will be a jump in the state or its derivative (the acceleration), and if so, how this jump is made. A simulation of contact must use models for the collision maps, the transition rules and the dynamics. For affine nonlinear systems, [65,149] give algorithms that compute all possible contact and release sets off-line. Using these sets, explicit expressions for the collision maps can be made off-line as well. During simulation, once the contact point is known, a simple check followed by one function evaluation then suffices to obtain the desired results. Due to discretization however, there remain a number of problems with respect to simulating uncontrolled and controlled contact. The problem with (un)controlled contact is the approximation of the time that contact with the boundary set is made. The problem with controlled contact is the validation of the transition rules for contact and release. Since real-time simulation often involves trade-off between accuracy and computation time, in [65] (Ch 9), parameters are introduced to aid the user of a simulation program to customize simulation to his/her specific needs.

The introduction of the Lagrange multiplier makes detection of contact and release during simulation studies nontrivial, as pointed out in Section 3.1. For instance the decision that a release takes place cannot be based on a check whether or not $\phi(q) > 0$, since an active Lagrange multiplier will ensure that $\phi(q) = 0$. It must be decided when to (de)active the Lagrange multiplier λ. The transition from free motion to constrained motion, and vice versa, or plastic collision detection can be based on the following analytical rules [43,149] (for codimension one constraints).

Contact rule: if $\phi(q) = 0$ and

$$\nabla \phi^T(q) \dot{q} = 0 \quad \text{then activate } \lambda \quad (48)$$

Release rule: if $\lambda < 0$ then deactivate λ.

First we discuss the problem of determining the time that a trajectory makes contact with the boundary set. Assume that at time t_i there holds $\phi(q(t_i)) > 0$. If contact is made in the interval (t_i, t_{i+1}) then this can be detected only if $\phi(q(t_i)) \cdot \phi(q(t_{i+1})) \leq 0$. This also implies that if a controlled contact and a controlled release take place in the interval (t_i, t_{i+1}), (and the motion stays on a boundary for a small period of time), this can be detected numerically only if the time-step is small enough. We assume that this is the case.

Depending on the characteristics of the numerical solution, it may well be that $\phi(q(t_{i+1}))$ is significantly smaller than zero. This implies that even with a fixed step-size, so-called step back is necessary. Step back means that the time-step h is adjusted and the simulation is started again from time t_i. This means that the same dynamic equations are solved again. And if the new estimate of $q(t_{i+1})$ is not satisfactory, again step back is necessary. Clearly such an iterative procedure may increase computation time significantly.

The method of false position or regula falsi can be used to establish the time-instant t^* where a trajectory makes contact with the boundary set. Since we assume stable numerical integration, the numerical approximated trajectory and the analytical trajectory are related through the accuracy of the numerical ODE solver that is used, and to the approximaton error on the initial data (that is needed after each impact). This means that it makes sense to search for a time instant $\hat{t}$ that is close, but not necessarily equal to t^*. The step-size is adapted to a value $\hat{h} \in (0, h)$. This gives $\phi(q(t_i + \hat{h})) \approx 0$. This approximation of the contact point may already be acceptable to the user of the simulation program. And using this value of $\hat{h}$, a linear (or nonlinear) interpolation with the numerical value of the velocities at times t_i and t_{i+1} may give an acceptable value of the velocity component at contact. But we can go on by restarting the simulation from time t_i by using $\hat{h}$ as the new time-step. Now the obtained numerical approximation is checked again to see if $\phi(q(t_i + \hat{h})) \leq 0$. If this holds then the procedure outlined above can be repeated until a point is reached for which the position constraint is not satisfied. The approximation of the time of contact, $\hat{t}$, is now set equal to $t_i + \hat{h}$, where $\hat{h}$ is the last update such that $\phi(q(t_i + \hat{h})) \approx 0$. Next, the time $t_i + \hat{h}$ is also used to approximate the velocity vector at the time of contact. Subsequently we set the time-step to its old value h again as the time-step $\hat{h}$ may be too small to maintain real-time simulation. A parameter γ should be introduced to set an a priori bound on the number of iterations. The trade off between accuracy and computation time is then put in the hands of the user. In [162,163], h is divided by two until $\phi(q(t_k))$ is smaller than a given tolerance. Another adaptive step-size method is proposed in [164]. Various methods similar to what is described above to compute t^* are discussed in [151–153] as well as their influence on the scheme order when combined with Newmark, RK24 and Dormand-Prince RK (the benchmark is a one-degree-of-freedom system with a single constraint).

Even though the procedure outlined above is simple, it still uses an iteration process. If the time needed to execute the procedure violates real-time simulation, the initial linear interpolation step may be replaced by a higher-order interpolation method using information at velocity level. Another promising approach is to use in the iteration process itself only a part of the dynamics equation: $M(q)\ddot{q} = w$. This choice is motivated by the fact that the contact set is independent of the system matrices M and Q and the importance of the matrix M in calculating the impulse [149]. This is still an active research area for real-time simulations of unilaterally constrained robotic manipulators (see also Section 6.5).

Next it must be determined whether one is dealing with a collision or not. In case plastic collisions are considered, after detection of contact, the Lagrange multipliers are activated and the motion remains on the surface, see Eq. (48). When elastic collisions are possible, there is another numerical problem since $d/dt\{\phi(q)\}(t^*)$ will in general not be zero numerically. This problem is similar to the problem in bilaterally constrained mechanical systems when the reaction force is approximately zero. A heuristic approach is the following. If at a contact point one has that $\nu \leqslant d\phi/dt(q(t^*)) \leqslant 0$, for a user defined value of the parameter ν, then the motion is assumed to proceed on the boundary. If $d/dt\{\phi(q)\}(t^*) < \nu(\leqslant 0)$ then the motion is to proceed by the use of an uncontrolled collision map. To detect release from the surface a check is made whether or not $\lambda_i^d < 0$, where λ_i^d is in Eq. (38). If this inequality holds true, then the Lagrange multiplier is deactivated. Since the compensation terms are usually much smaller then the value of λ_i^d itself, these compensation terms have no influence on this decision [65].

There remains a problem when the desired (simulated) contact force during motion on the boundary is very small. But then controller design is also difficult: the slightest deviation from the desired path will mean that release takes place, although control is aimed at maintaining contact. This is the reason that in practice a certain amount of (constant) normal force on the surface is chosen.

6.5 Collision detection between bodies

This section and the next one concern a very important module: the management of contacts status.

Approximation of the shapes and approximate calculation of impact times are generally CPU-time intensive tasks [120]. Many works have been dedicated to collision detection, eg, [148,165–167] to cite a few. Roughly this module requires to calculate, explicitly or implicitly, the expressions for $\phi(q)$ and solve $\phi(q)=0$ (ie, determinate the points that are going to touch–which are not necessarily the ones which are the closest at the instant of the computation, so several pairs of points have to be *watched* simultaneously). Even in very simple cases such as one degree-of-freedom systems, various numerical methods may be used to calculate the times t_k such that $\phi(q(t_k))=0$ [151–153]. Their influence on the algorithm properties (consistency, order) may be significant.

The main problem is that an exact analytical description of the objects shapes, even when this is possible, is quite time consuming. Secondly one has to calculate with a suitable numerical routine the times t_k. In case of accumulation of impacts and for multiple contacts, the problem is harder because the influence of deciding the end of the series $\{t_k\}_{k\geqslant 0}$ according to the machine accuracy, is not easy to quantify. Micro-collisions phenomenon [107] prove that it is possible in some cases that there is a large quantity of rebounds, but finite number of collisions, and an escape out of ∂K after a finite time. Things even complicate for multiple impacts. What is the influence on the long run motion if one decides instead that one constraint becomes active?

As pointed out in Section 2.2, it may not be possible to define all the constraints $\phi_j(q) \geqslant 0$: there would be too many! Hence one usually employs procedures that eliminate *useless* constraints, ie, those bodies which are too far one from each other to be likely to collide in the next future steps of integration. Consequently one implements *rough tests* that select the bodies which may collide, and *fine tests* to compute the collision times [120]. Rough tests usually consist of surrounding the bodies by simple volumes (spheres, boxes) and watching whether they overlap or not. Concerning the finest tests, the main approaches are (see [166] for a review):

- Classification of typical contacts and geometries [163,168–172]. In other words, process the real surface of the objects and the type of contact (circle-circle, circle-line, angle-line etc). These methods are essentially studied in the Mechanical Engineering literature. They are restricted to certain types of geometries contained in the available *library* developed for the software. If the bodies surfaces are simple enough to be described by analytical curves, one gets an explicit function $\phi(q(t))$ (see for instance the developments in [83] for the derivation of $\phi(q)$ from a local frame at the contact point). The next step is to solve numerically $\phi(q(t))=0$-which can be done with a Newton-Raphson method or a polynomial root finding routine, since in case of several roots Netwon-Raphson may compute the wrong zero and there is penetration before the algorithm decides that contact has occurred. Others [162,163] use a time step halving process until $\phi(q(t_k)) = 0$ is satisfied within a specified tolerance. We have already outlined this problem in Section 6.4. For instance, for two bodies with parametric surfaces $\phi_i(u_i,v_i,t)=0$, $i=1,2$, one has $\phi(q,t)=\phi_1(u_1,v_1,t)-\phi_2(u_2,v_2,t)$ and one faces a nonlinear 5-dimensional root-finding problem [148]. These methods are however less fast and more complex to implement than the 2-dimensional ones [127].

- For 2D systems, one can approximate the bodies B_i, $i \in \{1,\cdots,N\}$ by polygons made of edges and nodes N_i. Two main methods are used [120]: the node-in-polygon test (NIPT) and the ray-crossing approach (RCA), see Fig. 2a,b, respectively. Let n_i be the number of nodes N_i in polygon i, and α_{ii} the angle (N_jN_i,N_jN_{i+1}). Then if $\Sigma_{\text{nodes of }B_i}\alpha_{ii}=0$, the node $N_j \notin B_i$, if $\Sigma_{\text{nodes of }B_i}\alpha_{ii}=2\pi$, then $N_j \in B_i$: the bodies intersect. The RCA consists of looking at the number $\tilde{n}$ of intersections of a straight halfline (a ray) emanating from N_j, with the polygon ∂B_i. If $\tilde{n}$ is odd then $N_j \in B_i$, if $\tilde{n}$ is even then $N_j \notin B_i$. The RCA is more robust than the NIPT. Both methods are $\mathcal{O}(n_in_j)$ for two bodies B_i and B_j. However, their generalization to 3D systems is not easy [120].

- Approximation of the objects surfaces and of the impact times by bounding boxes methods [166] are more efficient for 3D systems. These methods are essentially studied in the Computer Science literature. If the bodies are convex and subject to gravity (or more generally to any vector field that is integrable) it is possible to approximate the distance $\phi(q)$ and to calculate a lower-bound on the impact time [165]. The approximation can be refined as much as the constraints (desired accuracy, speed of computation)
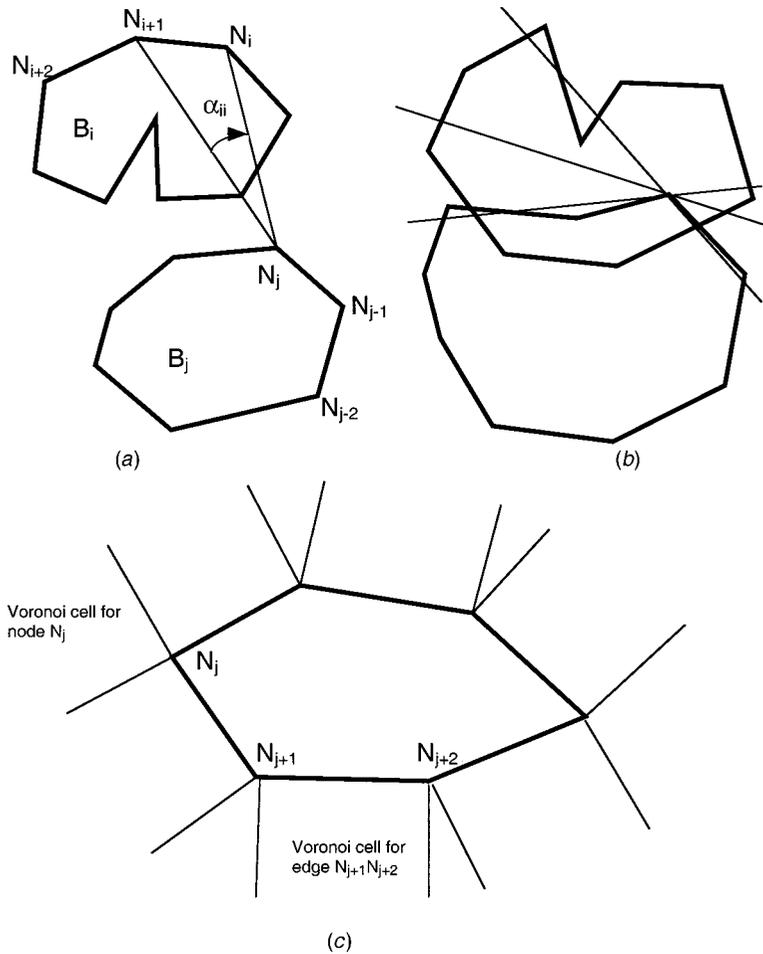
Fig. 2   Collision detection methods

permit to do it. In [148], it is pointed out that just watching positions to determine collision times cannot work since contact may occur between two sampling instants t_i and t_{i+1} while $\phi(q(t_i)) > 0$ and $\phi(q(t_{i+1})) > 0$. So including the velocity information in the algorithm is mandatory. Adaptive subdivision of the bodies into simple volumes (polygons or polyhedra [120], spheres [165,166], rectangular prisms [148]) and incorporation of a Lipschtiz boundedness condition on $\phi(\cdot)$ allows one to approximate the collision times [148,173]. This method is called *bounding box schemes*: each object is surrounded by bounding boxes. When these boxes overlap, the objects must be close one to each other. Then a more accurate collision test is made once more. Bounding box schemes allow one to avoid testing all possible contacts ($=\mathcal{O}(N^2)$ for N bodies), but to focus on objects in close proximity only. Roughly speaking, the Lipschitz bounds permit to approximate the next step motion of each simple volume (or surface) and to determine if a collision has occurred. A refinement of the mesh can be used to increase the accuracy of the collision time computation, in an adaptive way. These methods apply well to convex bodies. Nonconvex bodies can be decomposed into convex parts to be treated. Voronoi regions for polytopes [167,174] are used to maintain a list of clos-

est distances during the simulation[8]. The change in Voronoi cells from one step to the next one is usually small, facilitating the calculations. An implementation of the Lin-Canny algorithm with a running time linear in N, can be found in [175]. Baraff [176] proposes a coherence based bounding box test that is $\mathcal{O}(N)$.

6.6   LCP and NCP solvers

LCPs: As we saw in Sections 3.1, 3.2, and 3.3, the dynamics in Eq. (2) with or without friction, can be transformed in a form involving LCPs or NCPs which is useful to cope with detachment, sticking-sliding, impacts. Notice from Eq. (8) that if $B \geqslant 0$ (componentwise), then $\lambda = 0$ is a solution to the LCP. Problems arise when $B_i < 0$ for some i. The most popular algorithm used to solve LCPs, Lemke's algorithm, is motivated by this observation [110,177,178]. Basically this is a pivoting method that converges in a finite number of pivots provided the LCP matrix A in (8) is a copositive matrix. In the worst case, Lemke's algorithm is exponential but is expected to be polynomial in m. Lemke's algorithm is initial-

[8]A Voronoi cell associated to an object consists of the set of points whose distance to this object is the smallest. The object can be a node, an edge, a face. See Fig. 2c.

21

ized using a so-called covering vector [89], or supplementary variable [178], or initial ray [177], that is chosen arbitrarily or even randomly [89]. Lemke's algorithm is formulated in [177] as finding the zero of a piecewise linear function, using a homotopy method. Even in the nice case where A is copositive, Lemke's method is known to fail when the dimension is too large (because of accumulated errors of pivoting), or even for medium-sized problems if the diagonal of A has null entries [115]. Further interesting properties of Lemke's algorithm have been proved in [179]. Other methods to solve LCPs exist, such as Mangasarian which consists of using a nonlinear (complementarity) function whose zero is the LCP solution [180], then a root-finding algorithm. Another homotopy method called Katzenelson's algorithm exist [181]. Mangasarian's method has been used in [182]. As we pointed out above, LCPs and Quadratic Problems under constraints are equivalent when A is copositive (then the QP is convex and its solution satisfies the so-called Karush-Kuhn-Tucker conditions which are complementarity conditions between $\mathbf{N}$ and a Lagrange multiplier—a slack variable). In [183], Interior Point algorithms are described that solve LCPs as QPs under constraints. In [89], a comparison between Lemke's and a new specific Interior Point algorithm is proposed. 35 different problems are tested, randomly generating 328 data sets, with LCPs of dimensions ranging from 2 to 170. The IP seems to supersede Lemke in terms of the number of problems it can solve when Lemke's covering vector is chosen arbitrarily. See [115], [178], and [184] for a comparison between Lemke, Gauss Seidel and augmented Lagrangian methods, using distinct-element models of various mechanical systems: Lemke's algorithm is shown to provide the smallest CPU computation time.

Baraff developed a Dantzig algorithm [185]. Let there be m frictionless contacts with a set of complementarity conditions between $\dot{V}_N^{(j)}$ and $\mathbf{N}^{(j)}$ at each contact. Initially the contacts are classified into 2 sets: $\dot{V}_N^{(j)}=0$ and $\dot{V}_N^{(j)}>0$, whereas one sets $\mathbf{N}^{(j)}(0)=0$, $\forall j \in \{1, \ldots, m\}$. Dantzig algorithm is a pivoting method of the indices between the 2 sets. One starts by computing suitable values of $\dot{V}_N^{(1)}$ and $\mathbf{N}^{(1)}$, then one passes to contact 2 and adjusts both contacts 1 and 2. Indeed modifying the data $\mathbf{N}^{(2)}$ and $\dot{V}_N^{(2)}$ must in general modify $\mathbf{N}^{(1)}$ and $\dot{V}_N^{(1)}$, see Eq. (8). Then contact 3 is adjusted, taking into account 1 and 2 as well. And so on until contact m is attained. Then all contacts satisfy the complementarity conditions and the LCP(λ) in (8). The algorithm terminates if A is copositive. Extensions to frictional contacts are possible, in 2 or 3 dimensions. It is noted in [185] that in many practical cases $\nabla \phi$ is constant, consequently B in (8) is in the column space of A: this property may be used for singular A (eg, hyperstatic systems). Moreau uses a Gauss-Seidel method [78], see Section 7.2.2.

In general, users may be tempted to directly connect their software with available codes. However the use of packages has the following drawbacks [185]:

- Interfacing the numerical software package with the simulation software may not be easy.
- The available codes are often implemented as research codes, especially for QPs.
- The tuning of the package adjustable parameters (numerical tolerances, iteration limits etc) may not be easy.

- The packages are often *black boxes* whose codes are not accessible to the user. The origin and remedy of numerical problems may therefore be difficult to isolate.

Algorithms for 2D frictional contact problems in the quasistatic case have been presented and compared in [186]. It is shown that this reduces to finding the zero of a function in a closed domain. Newton's method seems to supersede interior point ones.

NCPs: The NCP in Eq. (16) is equivalent to solving $\min[y, g(y)]=0$. Such nonsmooth equations can be treated with nonsmooth Newton methods [187]. As a generalization of the variational expression in Eq. (21), (16) is also equivalent to the variational inequality [111]: find $y \in \mathbb{R}^+$ such that $(z-y)^T g(y) \geqslant 0$ for all $z \in \mathbb{R}^+$. From a theoretical point of view NCPs are still an active research area, (see eg, [188] and [189] and references therein).

6.7 Event-driven algorithms

Basically these are algorithms which integrate the motion between events (shocks or stick-slip transitions) and use some event detection procedure. They are of the force-acceleration type and require the computation of contact forces. The works therefore focus on the improvement of the modules in Sections 3.2, 3.3, 6.5, and 6.6. The time-discretization problems are not treated explicitly, but the authors generally merely use available DAEs or stiff ODEs solvers. Roughly, the main discrepancy between what we call event-driven and time-stepping schemes is that the former are closer to the continuous-time description of the dynamics in Eq. (2) with a two-time-scale dynamics (which is close in spirit to the hybrid dynamical systems point of view [42]). The latter are a real difference equation approach, and therefore, better lend themselves to convergence analysis, a crucial property. On the contrary event-driven algorithms do not seem amenable for convergence analysis. This may have important consequences for the simulation of complex dynamics. The first three groups of algorithms in Sections 6.7.1, 6.7.2, and 6.7.3 clearly supersede the others in the way they formulate the nonsmooth dynamics. They are briefly presented here since most of the specific material they contain has been presented elsewhere in this paper.

6.7.1 Pfeiffer and Glocker's formulations

These authors proposed in [7] a general formulation of the dynamics in Eq. (2), at acceleration-force level. The basic convex analysis tools described in Section 3.4 are used to derive complementarity conditions and write friction and impact rules with LCPs or NCPs [83]. Two-dimensional friction (planar systems) are treated in [7], Lemke's algorithm is used. Extensions to 3D systems is studied in [83], using polytope approximations of the friction cone and convex analysis tools which allow one to express such approximate Coulomb's law with complementary slackness conditions. It encompasses collisions with friction and tangential restitution, see Section 3.3. The work in [92] uses the second formulation in Eq. (21) to express the complementarity conditions in Eq. (10) and dry friction in dimension 3 (see (23) and its acceleration counterpart). The dynamics in Eq. (2) is

transformed in an implicit equation for the acceleration and $\mathbf{N}$, F_t, to be solved at each step by a root finding algorithm (a subroutine NEWT from [190] is used, and the dynamics are integrated with a RK-Fehlberg with automatic control of the integration step). Again convex analysis tools are at the core of the proposed method. The numerical results are compared to experiments (a planar mass on an oscillating oblique plane). In another work [182] the same authors use an augmented Lagrangian approach to solve the same problem, with a linearized friction cone. The numerical predictions fit well with experiments.

In most of these schemes, the integration of the continuous dynamics is done with a RK scheme. The existence of solutions of the obtained LCPs (Painlevé paradoxes) is discussed but not treated. Some of these results have been implemented in a commercial software [191,192].

6.7.2 Modified Moreau's scheme

The algorithms presented in [4] and [5] may be seen as an adaptation of Moreau's schemes (in particular the Gauss-Seidel method to solve the multi-contact problem with friction). They provide the user with the event occurences and the contact forces values, which are of primary importance for engineers doing virtual prototyping. Due to these industrial constraints, an event-driven force-acceleration formulation is more suitable than a time stepping velocity-impulse one. Also low-order algorithms (Euler) may not be very accurate when applied to systems with sparse events, a drawback that is well-known for free-motion systems [26]. The algorithms have been thoroughly tested on the company Schneider Electric circuit breakers (low and average tension) and have proved to supply the design engineers with much more reliable results than the available penalty-based softwares, see Section 9. In particular, a significant advantage is that a change of topology in the mechanisms does not necessitate a re-estimation of the contact parameters as it is the case when spring-dashpot contact models are used.

6.7.3 Baraff's algorithms

Baraff, motivated by problems in computer graphics and animation, essentially focused on the calculation of contact forces and development of specific methods to increase the speed of calculation and the resolution of LCPs or NCPs. Many different aspects of rigid body simulations are discussed in [179], like Painlevé paradoxes, the formulation of friction in 3 dimensions, as well as quite interesting developments on Lemke's algorithm. Lemke's algorithm is used in the early works, but Baraff developed a Dantzig algorithm in [185], See Section 6.6.

6.7.4 Other schemes

The works in [64], [162], [163], [168–172], [191], and [193–198] essentially focus on the dynamics formulation (Section 2.1), the collision detection (Section 6.5), the impact rule. The type of integrator that is used for the continuous phases of motion is generally not provided (except [169,196] RK 4, [162] DADS 2D, [198] Newmark). The problems raised in Sections 3.1, 3.2, and 3.3 are usually ignored, except [171] and [193] who analyze multiple impacts.

It is for instance not clear how the capture and finite accumulations of impact times t_k are treated numerically. An extension of DASSL (called DASSLRT) has been used in [199] to simulate several discontinuous phenomena that affect robot motion. Basically these authors consider that the dynamics in Eq. (2) is a time concatenation of DAEs (no mention is made of inequality constraints and related notions). A root finding algorithm is used to compute transition instants. DASSLRT and the ADAMS routine supplied in SIMULINK-with a coordinate reduction, see Section 2.2—are compared on an example (a double pendulum with a singular configuration, ie, $\nabla \phi_b(q)$ in Eq. (2) is not full column rank). Other routines failed. DASSLRT seems to be quite fast, $\mathcal{O}(0.1 \text{ s})$, while ADAMS took 20 min, due to the singular configuration, in the neighborhood of which acceleration diverges[9]. No numerical result incorporating detachment and collisions is presented. The specific multiple contacts modelling problems and complementarity conditions are almost totally ignored in most of these studies. The study in [151–153] may be a first step towards a better understanding of interconnections between the integrators of ODEs or DAEs and impact detection algorithms, despite it is limited to a very simple one degree-of-freedom system (with however possible finite accumulations of impacts).

7 TIME-STEPPING ALGORITHMS

We have described above some various (and classical) manners to discretize DAEs and ODEs. The question one may ask to oneself is: how does the addition of complementarity conditions and state re-initialization modify the properties of such schemes? In other words, many schemes are known to be consistent (ie, the discrete piecewise constant solution converges—in a certain sense—towards the solution of the real dynamics). As we shall see further, many time-stepping schemes, though not all, consist of discretizing simultaneously the continuous dynamics and the complementarity conditions, consequently forming a LCP or a NCP to be solved at each step. The simulation is then easily advanced in time by solving the LCP, using the available solvers, see Section 6.6. Additionally, state re-initializations are needed. The interest of time-stepping methods over event-driven ones, is that they aim at providing a difference-equation approach to the simulation problem, which is suitable for convergence proofs. This may be quite important in case there are many impacts, because it remains to be shown that event-driven schemes are robust with respect to the accumulation of initialization errors (after each impact). This fact combined with the problem of sensitivity with respect to initial data, may render their use quite delicate.

Actually the methods presented in this section all belong to the time-stepping schemes family. But as we shall explain,

---

[9]Notice that this is an artifact. Indeed consider Eq. (35). Clearly, if one tries to compute the multiplier using this formula, difficulties arise whenever the inverted matrix loses its rank. However, as shown in [43] and [71], if $\nabla \phi \in \mathbb{R}^{n \times m}$ has full column rank, then the multiplier vector and the acceleration are unique. If the matrix loses its column rank, then the acceleration remains unique, but not the multiplier vector. Loss of column rank may occur because of hyperstatism or at a singular configuration. Special techniques can be developed to cope with this problem [4,7]. Some authors [83] recommend to avoid inverting matrices of the type $W^T M^{-1} W$.

there are significant discrepancies between the schemes within this family. For instance, the ones in Section 7.2 and Section 7.3 do not explicitly require the calculation of the contact forces, contrary to the ones in Sections 7.4 and 7.1. Moreover the schemes in Section 7.2 and Sections 7.3, 7.4 do not rely on an accurate determination of the shock instants: they work with constant h. These schemes are, therefore, true difference equations of the MDI in Eq. (2), with however, possible iterations within a step due to their implicit form. The schemes in Section 7.4 work with h constant, but the shock equations (in some of them) are treated apart when a velocity reinitialization is needed. The schemes in Section 7.1 are based on an accurate calculation of the times of contact with a linear interpolation, similarly to event-driven algorithms. The algorithms in Section 7.2 and Section 7.4 are of the velocity-impulse form. This has some consequences on the integration when friction is present, see Section 8.1.

7.1 Lötstedt's algorithms

In [27], Lötstedt introduced time-discretization procedures, at the acceleration/force level. We choose to classify these algorithms into the time-stepping section, because Lötstedt explicitly discretized the dynamics and the complementarity conditions (on the velocity) to form a LCP or a NCP whose unknown is the multiplier λ_i at each step i. However the discontinuity instants (on the velocity—when there is a shock— and the acceleration—when there is a transition between stick and slip phases—) are computed by an inverse linear interpolation, similarly to an event-driven algorithm. The detection of these instants is made by monitoring the impulse (considered to be zero under a certain threshold as in Section 6.4). Lötstedt's work can with no doubt be considered as an important pioneering work in the field of time-discretization of nonsmooth mechanical systems.

7.1.1 The frictionless case

Let us consider first the frictionless case. The following numerical scheme is proposed to compute the state at step i:

$$
\begin{cases}
q_i = \dfrac{1}{\alpha_0^1} b_i^1 \\[2ex]
\dot{q}_i = \dfrac{1}{\alpha_0^2}[h\beta_0^2 M^{-1}(Q_i + \nabla\phi_i\lambda_i) + b_i^2] \\[2ex]
\nabla\phi_i^T\dot{q}_i \geq 0, \quad \lambda_i \geq 0, \quad \lambda_i^T\nabla\phi_i^T\dot{q}_i = 0
\end{cases}
\tag{49}
$$

where

$$
b_i^1 = h\sum_{k=0}^{r}\beta_k^1\dot{q}_{i-k} - \sum_{k=1}^{r}\alpha_k^1 q_{i-k}
$$

$$
\tag{50}
$$

$$
b_i^2 = h\sum_{k=1}^{r}\beta_k^2 M^{-1}(Q_{i-k} + \nabla\phi_{i-k}\lambda_{i-k}) - \sum_{k=1}^{r}\alpha_k^2\dot{q}_{i-k}.
$$

Clearly, the complementarity relations in Eq. (49) correspond to the active constraints at step i, see Section 3.1. They encompass the persistent contact as well as plastic impacts phases. The formulas in Eqs. (49) and (50) correspond to two linear r-step methods [25] Section VII.3. $\nabla\phi_i = \nabla\phi(q_i)$ and

$\phi_i = \phi(q_i)$. The coefficients α_k^1 and β_k^1 are determined from an Adams-Bashforth family of explicit formulas ([26], page 250), denoted as AB-r, $\beta_0^2 = 1$, $\beta_k^2 = 0$ for all $k = 1,2,\cdots,r$. The second equation in (49) is a backward difference formula, denoted BDF-r. Notice that the mass matrix M is assumed to be constant (hence the Coriolis and centrifugal torques are zero), which restricts the application to simple mechanical systems with Euclidean configuration space (like collections of particles). It is however argued that this is just a matter of convenience to allow for an easy factorization of M, and that the extension towards $M(q)$ is possible ([27], Section 6). The torque $Q_i = Q(q_i, \dot{q}_i, t_i)$ therefore contains gravity, viscous friction, and external actions (like control inputs). The integration step is chosen constant, equal to h.

When $Q = Q(t,q)$, it is shown in [27] that a LCP whose unknown is λ_i can be formulated from (49). As we already pointed out, this LCP can be rephrased as a quadratic program:

$$
\min_{\lambda_i \geq 0} \frac{1}{2}\lambda_i^T\nabla\phi_i^T M^{-1}\nabla\phi_i\lambda_i + h^{-1}\lambda_i^T\nabla\phi_i^T[b_i^2 + hM^{-1}Q_i].
$$

$$
\tag{51}
$$

Consequently, the set of equations in (49) allows one to advance the solution in time from $i-1$ to i. The methods AB-1 (forward Euler)-BDF-1 and AB-2-BDF-2 are chosen in [27], where it is recalled that it is useless to use methods of order ≥ 3 (linear multistep A-stable methods have an accuracy of order ≤ 2, ie, at most $\mathcal{O}(h^2)$ [26] pages 250–251). After discontinuities in $\dot{q}_i$ or $\ddot{q}_i$ (which are detected from the value of the impulse on one step-with a threshold under which it is considered to be zero), the AB-1-BDF-1 algorithm is used during two steps to restart the simulation (it is known that multistep methods are not self-starting and require the help of a single-step algorithm initially).

When $Q = Q(q, \dot{q}, t)$, then the LCP formulation is lost. However Lötstedt proves that provided the matrix

$$
A(\dot{q}) = M - \frac{h}{\alpha_0^2}\frac{\partial Q}{\partial \dot{q}}(q, \dot{q}, t) \tag{52}
$$

is full rank and $\nabla\phi_i^T A^{-1}\nabla\phi_i$ is positive definite, then (49) still possesses a unique solution so that the algorithm can be used to safely advance the solution in time. However this time λ_i is generally the solution of a NCP (a quick look at the second equation in (49) allows one to realize this). The condition in (52) can be used with the implicit function theorem to express $\dot{q}_i = g_i(\lambda_i)$ for some function $g_i(\cdot)$. The second condition is used to prove the existence of a solution to the NCP. A way to solve the NCP is proposed, based on functional iteration. Certainly this could be improved using new tools, see Sections 3.1 and 6.6. In summary, Lötstedt's algorithm is given as follows [27]:

- Compute q_i using AB-1 or AB-2, with h such that the local error in q_i is smaller than $h\epsilon$ for a prescribed tolerance ϵ (ways to estimate such a h are given in [27]).
- Calculate $\nabla\phi_i$ to a prescribed accuracy and calculate $\ddot{q}_i = M^{-1}(Q_i + \nabla\phi_i)$ and $\dot{q}_i$ by BDF-1 or BDF-2.

24

- Test whether velocities and accelerations are discontinuous between t_{i-1} and t_i, either due to an impact (detected from a nonzero value of the impulse) or to the activation of a new constraint ($\phi_k(t_{i-1})>0$ and $\phi_k(t_i)\leqslant 0$ for some k), or to the deactivation of a constraint. The time of such jumps is calculated by inverse linear interpolation. After a shock a new velocity $\dot{q}_{i+1}$ is computed by a collision rule (the rule in Eq. (12) is used). Then restart the algorithm at the first step with AB-1 and the new set of active constraints.
- Test the detachment conditions by checking whether one entry of the vector λ_i passes through zero, and whether the corresponding entry in the normal velocity $\nabla\phi_i^T\dot{q}_i$ is positive. Then refresh the set of inactive constraints if needed.
- End.

Remark 7:
- The algorithm is based on velocity constraints to reduce the index and form an LCP. The drawback is evidently a possible drift away from ∂K during permanent contact phases (DAE simulation). The stabilization on the constraints ∂K during permanent contact phases needs special attention [22,66], see Section 5.
- Lötstedt also shows that the LCP(λ_i) can be reformulated as the minimization problem

$$\min \; \left\|\sum_{k=0}^{r} \alpha_k^2 \dot{q}_{i-k} - hM^{-1}Q_i\right\|_M, \quad \nabla\phi_i^T\dot{q}_i\geqslant 0. \tag{53}$$

- The velocity can be calculated directly from Eq. (53). However it is argued in [27] that it is better to compute the multiplier and then insert it in the dynamics because this simplifies the initialization of the algorithm (a feasible λ_0 is easy to find), the QP in Eq. (51) is a particular case of the QP constructed for the case with friction (see below with comments), and it may be interesting to know the value of λ_i explicitly, since λ_i can be used as a measure of force. This is of interest for control design of real-world systems [21].
- Dissipativity of Eqs. (49), (50), and (51) plus the impact rule and convergence of the algorithm are not proved.

7.1.2 Constraints with 2D friction

As recalled briefly in the introduction and Section 8, the dynamics in Eq. (2) is much more complicated when friction is considered. Friction may create some unexpected phenomena [18,96,200,201] as the divergence of the contact force (but with bounded impulses) or so-called *Impacts without Collisions*. These phenomena are not to be considered as artefacts due to the model deficiencies. Although they *disappear* when rigidity is relaxed, or when the Coulomb model is replaced by some regularized law, they really represent physical phenomena which occur in real systems. More comments are given in Section 8. Since Lötstedt's algorithms are acceleration/force schemes, these phenomena should be taken into account. Lötstedt was perfectly aware of such problems [202], and therefore, proposed a particular numerical procedure to avoid them.

The same algorithms AB-1-BDF-1 or AB-2-BDF-2 are used as in the frictionless case. When 2D friction is incorpo-

rated in the algorithm, one has to add the tangential contribution of the contact force in the right-hand-side of the second equation in (49). The contact force is split into two parts, see Section 3.3: $\nabla\phi\lambda + P(q,\dot{q},\lambda) = [G_1(q)+H(q,\dot{q})]\lambda_1$. Roughly $G_1(q)\lambda_1$ contains the normal generalized force and the contribution of the sticking contacts, whereas $H(q,\dot{q})\lambda_1$ accounts for the sliding contacts. The vector λ_1 contains the normal multipliers $\lambda_{n,j}$ and the tangential ones $\lambda_{t,j} = \pm f\lambda_{n,j}$ for the jth contact. There are two features in the algorithm proposed in [27]. The first one is the approximation of $\lambda_{1,i}$, the second one is the calculation of the impulses at the shock instants. Let us denote the jth component of λ_1 by λ_1^j and its ith iteration by $\lambda_{1,i}^j$. Then the approximated value is $\bar{\lambda}_{1,i}^j = \lambda_{1,i-1}^j + h_i(\lambda_{1,i-1}^j - \lambda_{1,i-2}^j)/h_{i-1}$, for a variable step of integration h_i. A QP is constructed that allows the computation of the term $G_1(q)\lambda_1$. As noted in [185] (Section 9.2.1), it possesses the advantage of assuring that the tangential force is opposite the tangential acceleration. But it has the strong drawback that sliding generally implies the QP matrix to be non-symmetric, rendering the problem harder to solve. It is clear that the introduction of $\bar{\lambda}_{1,i}^j$ in the dynamical equations modifies the subsequent calculations in a non-physical manner right after the first step[10], and should be avoided. Special procedures are also used after a shock and a discontinuity in the acceleration. The error introduced in $\dot{q}_i$ by the use of $\bar{\lambda}_1^j$ in a permanent contact phase are shown to be $\mathcal{O}(h^3)$ when $h_i=h$, a constant. They are $\mathcal{O}(h)$ after a reinitialization of the velocity or of the acceleration.

The second point (calculation of the impulse at a shock instant), is formulated as follows. Taking frictional effects at impacts into account, let us denote the right-hand-side of Eq. (3) as $P_k=G(q)\Lambda_k$, where Λ_k is a vector of normal and tangential percussions. Then Lötstedt proposes to calculate the impulse from the QP in Eq. (17). The dual version of this QP is:

$$\min \; \frac{1}{2}[\dot{q}(t_k^+) - \dot{q}(t_k^-)]^T M[\dot{q}(t_k^+) - \dot{q}(t_k^-)]$$

$$W^T v = G^T\dot{q}(t_k^+), \quad v\geqslant 0, \quad v^T W\Lambda_k = \Lambda_k^T G^T\dot{q}(t_k^+)=0$$

$$W = \begin{pmatrix} I & 0 \\ f_{imp}I & -I \\ f_{imp}I & I \end{pmatrix} \tag{54}$$

where I is the identity matrix with dimension equal to the number of active constraints, and f_{imp} can be considered as an impulse ratio [93]. The main problem with the calculation in Eq. (54) is that although it looks like the plastic impact rule Eq. (12), it is not like (12): there may be rebounds. In addition, if there is a tangential velocity reversal during the shock (ie, the post and pre-impact tangential velocities have opposite signs), then there may be a kinetic energy gain at the shock instant (this phenomenon is well-known in the literature, see eg, [18,93]).

---

[10]Indeed the state at step 2 is calculated with $\bar{\lambda}_1^j$, so the next λ_1^j is calculated from wrong positions and velocities.

Remark 8:

- Lötstedt discusses the issue of Section 6.5 for collections of polyhedral objects. He points out the need for a selection procedure for the collision detection, but does not provide many details on this part of his algorithm.
- The algorithm in [164] is close in spirit to Lötstedt's one (time-stepping with accurate detection of contacting times). It uses a trapezoidal discretization of the continuous frictionless dynamics (implicit one-step scheme, solved by a Newton method with an initial guess from a Euler's discretization), and an adaptive step size procedure. Several simple examples show that h may decrease to very small values as 10^{-12} s during the simulation. Lemke's algorithm is used to solve the contact force LCP and the impact percussion LCP (see Eqs. (11) and (12)).

7.2 Discretization of Moreau's sweeping process

7.2.1 Frictionless sweeping process

We describe in this section and the next one the time-discretization of the general MDI presented in Sections 3.4.1 and 3.4.2. It has been named the *NonSmooth Contact Dynamics* (NSCD) method by Moreau and Jean, and can be considered as one of the results of the research led by Moreau in Montpellier on Convex Analysis and Nonsmooth systems since the early 60s (see eg, [70,95,104,203,204]). The presented methods constitute the first attempt [28] of time-discretization of MDIs as in (22), ie, simulation of multibody systems without regularization of either the normal or the tangential friction laws of contact/impact. The Contact Dynamics method provides a very general and powerful framework for the simulation of various nonsmooth mechanisms, including granular matter [12,78,205], buildings made of blocks and monuments [13], deep drawing process [94,206,207], robotic systems [208] and kinematic chains [5]. It so happens that the time-discretization of the MDI in Eq. (22) yields an intrinsic implicit formulation. But it can be transformed into an explicit scheme using basic convex analysis [101,209]. We first concentrate on the discretization of (22), then we indicate how friction is treated, and finally we focus on the general NSCD method. Following [28,96] choose in Eq. (22) at step i:

$$
\begin{cases}
t_{m,i} = t_i + \frac{1}{2}h \\
q_{m,i} = q_i + \frac{1}{2}hv_i \\
M_{m,i} = M(q_{m,i}) \\
Q_{m,i} = Q(t_{m,i}, q_{m,i}, v_i) \\
v_{l,i} = v_i + hM_{m,i}^{-1}Q_{m,i} \\
-M_{m,i}(v_{i+1} - v_{l,i}) \in \partial\psi_{V(q_{m,i})}(v_{i+1}) \\
q_{i+1} = q_{m,i} + \frac{1}{2}hv_{i+1}
\end{cases}
\tag{55}
$$

where h is the integration step (that might be varying from one step to the next). One sees that the computation of the values at step $i+1$ depend on intermediate calculations at the midpoint $t_{m,i}$. The midpoint velocity $v_{l,i}$ is equal to the approximated velocity that the system would have at step i

$+1$, if there was no collision on (t_i, t_{i+1}). This is easily seen from the equivalence between the inclusion in Eq. (55) and

$$
v_{i+1} = \text{prox}_{M_{m,i}}[V(q_{m,i}), v_{l,i}]
\tag{56}
$$

where prox means the proximation operation in the kinetic metric (with an underlying projection in the same metric). The tangent cone $V(q(t_i))$ is defined also outside the admissible domain K in order to cope with possible violation of the constraints in the course of the simulation, see Eq. (18). Indeed if there is a contact event between i and $i+1$, which is detected by checking the negative sign of $\phi(q_{m,i})$, one uses Eq. (56) to compute v_{i+1}. This is a quadratic program in the kinetic metric $M_{m,i}$. The reader may recognize again that the formulation in the last line of Eq. (55) encompasses the whole dynamics (continuous motion and shocks). This discrete-time inclusion is a discretization of the MDI in Eq. (22) with $dv \simeq v_{i+1} - v_{l,i}$, $v(t^+) \simeq v_{i+1}$ and $v_{l,i}$ can be interpreted as a left-velocity at t_{i+1}.

The general case $0 \le e \le 1$ can also be handled, as we observed in Section 3.4.1 [35,96]. Let $u = v(t^+) + ev(t^-)/1 + e$ in Eq. (22) (with $v(t^+)$ in the right-hand-side of (22) also replaced by u, see also Remark 4). Then Eq. (22) can be discretized as follows [35]:

$$
\frac{h}{1+e}Q(t_{i+1}, q_{i+1}, v_{i+1}) - M(q_{i+1})\frac{v_{i+1} - v_i}{1+e}
$$

$$
\in \partial\psi_{V(q_{i+1})}\left(\frac{v_{i+1} + ev_i}{1+e}\right)
\tag{57}
$$

which is clearly an implicit formulation but is equivalent to:

$$
\frac{v_{i+1} + ev_i}{1+e} = \text{prox}\left(V(q_{i+1}), v_i\right.
$$

$$
\left. + \frac{h}{1+e}Q(t_{i+1}, q_{i+1}, v_{i+1})\right).
\tag{58}
$$

One notes that there are some differences between Eqs. (57), (58), and (55). This last formulation is used in [32] and [35] for the study of convergence of the algorithm. This indicates that mathematical convergence proofs and real implementation of an algorithm may lead to different discretizations. Equation (58) is an explicit form as long as $Q = Q(t, q)$ and provided one chooses $q_{i+1} = q_i + hv_i$. Indeed in this case introducing q_{i+1} inside the left hand side of Eq. (58) one gets a direct way to calculate v_{i+1} by a proximation tool (ie a quadratic programming approach).

Remark 9:

- *Convergence*: Proofs of convergence of the discretized Moreau's sweeping process can be found in [32], using a discretization as in Eq. (57) and $e = 0$. The scheme in Eqs. (57) and (58) is proved to converge globally for $0 \le e \le 1$ in [35] using techniques inspired from [32], with $\partial K \in C^{1,\beta}$, $\beta > \frac{1}{2}$ and codimension one constraints ($m = 1$).
- *Impact calculation*: The usefulness of the midpoint calculation in Eq. (55) lies for instance in the fact that, as one easily calculates, $q_{i+1} = q_i + \frac{1}{2}h(v_i + v_{i+1})$. Contact is detected if $\phi(q_{m,i}) < 0$, while $\phi(q_i) > 0$. In the case of an elastic collision (think for instance of the simple one-

degree-of-freedom bouncing ball), one gets $v_{i+1} = -v_i$, since the velocity is simply reversed as can be checked from Eqs. (55) and (56): one has $\mathrm{prox}[\mathbb{R}^+, v_i + h/2\,g] = 0$ for h small enough since $v_i < 0$ at a shock instant; modify (56) in the same lines as Eq. (58) to get the result. Thus $q_{i+1} = q_i$, hence the ball rebounds exactly (at the machine accuracy) to the same height, whatever the length of the simulation. This is also pointed out in [91] who discusses symplectic integrators and energy conservation problems on the bouncing ball example. As we shall see some schemes do not possess this property, and may yield energetical inconsistencies. One sees that in the time-stepping scheme in Eqs. (55)–(56), $v_{i,l}$ plays the role of the pre-impact velocity, while v_{i+1} plays the role of the post-impact velocity. This is on contrast with some event-driven schemes, where the impact rule is applied at step i. The midpoint Euler scheme is further justified in [28] (Section 6), by the fact that it assures a much more accurate estimate of q_i than a Euler algorithm. This fact combined with the need of jumps detection led Moreau to choose a discretization as in Eq. (55). Notice that (55) is a one-step method, but is not a second-order explicit RK ([26], page 247). Further sophistications of the algorithm are presented below.

- Constraint drift: When there is only one contact that is made, the proximation in Eq. (56) is equivalent to $\nabla \phi_{m,i}^T v_{i+1} = 0$. This will generally result in a bad stabilization of the constraints during persistent contact phases. The approximation of the tangent cone $V(q_{m,i})$ can be replaced by $V(q_i + hv_i)$. This may help in satisfying the constraint better [96]. In addition if the constraints are not respected, then Moreau [28] proposes to use a projection of q_{i+1}, denoted as q_{i+1}^c, onto ∂K, computed in the kinetic metric as:

$$q_{i+1}^c = q_{i+1} - \phi(q_{i+1})(\nabla\phi M^{-1}\nabla\phi^T)M^{-1}\nabla\phi \qquad (59)$$

  where the last term is evaluated at q_{i+1}.

- **Transition phases and capture:** As we explain in the **Impact rules** paragraph of Section 7.4, the transition between free-motion and persistent-contact phases—ie, the problem of capture—, *via* a shock or a sequence of shocks, is automatically treated by the algorithm in Eqs. (55) and (56) or (57) and (58). This feature is actually shared by other time-stepping schemes as the ones in Section 7.3, and constitute a serious advantage of time stepped algorithms (which are a truly difference equation approach to simulation, contrary to event-driven schemes which are in a sense closer to the continuous-time description for impact detection and collision effects). Note however that in event-driven schemes contact/rebound happens at one time instance t_i, whereas with time-stepping schemes there is always at least a time-delay h between contact and release. Incidentally, notice that replacing the left-hand-side in Eq. (56) by $v_{i+1} + ev_i/1 + e$ does not change much the capture problem once $v_i \approx 0$. Here again one sees that v_{i+1} plays the role of $v(t^+)$ whereas v_i plays the role of $v(t^-)$ when an impact is detected at $t_{m,i}$.
- Actually as shown by Moreau [100,101] the first-order

sweeping process formulation $-dv/dt \in \partial\psi_{C(t)}(v(t))$ with $C(t)$ a moving convex set, cannot be discretized via any Euler explicit scheme of the form $v_i - v_{i+1} \in \partial\psi_{C(t_i)}(v_i)$ since this is equivalent to $v_i = \mathrm{prox}(C(t_i), v_{i+1})$ which is nonsense. On the other hand, $v_i - v_{i+1} \in \partial\psi_{C(t_{i+1})}(v_{i+1})$ makes sense since this is equivalent to $v_{i+1} = \mathrm{prox}(C(t_{i+1}), v_i)$ which provides an explicit way to calculate v_{i+1}. One notices that the integration step does not appear in these expressions: this is due to the fact that since the right hand side is a cone, multiplying it by any positive constant does not change the inclusion. In the second-order sweeping process formulation, h appears explicitly at the velocity level, see Eq. (57).

7.2.2 Constraints with Coulomb friction

At each contact $\mathcal{A}^{(j)}$ the Coulomb friction law is formulated as [94,104]:

$$F_{t,i+1}^{(j)} \in \mathcal{D}_{i+1}^{(j)}, \quad \forall s \in \mathcal{D}_{i+1}^{(j)}, \quad [s - F_{t,i+1}^{(j)}]^T V_{t,i+1}^{(j)} \geqslant 0 \quad (60)$$

where the notations are the same as in Section 3.4.2. In particular, $\mathcal{D}_i^{(j)} = \mathcal{D}(f|\mathbf{N}_i^{(j)}|)$ has an obvious meaning from the definition of $\mathcal{D}$. The discrete form in Eq. (60) is equivalent to $F_{t,i+1}^{(j)} = \arg\max_{z \in \mathcal{D}_{i+1}^{(j)}}(-z^T V_{t,i+1}^{(j)})$, and to $F_{i+1}^{(j)} = \arg\max_{z \in \mathcal{C}_{i+1}^{(j)}}(-z^T V_{i+1}^{(j)})$. In the 2D case, Eq. (60) exactly represents the Coulomb graph between $V_{t,i}^{(j)}$ and $F_{t,i}^{(j)}/\mathbf{N}_i^{(j)}$. In the 3D case, one recovers a friction cone.

The Gauss-Seidel method [78,101] roughly consists of the following *Signorini Coulomb loop* [94]. Suppose that for contact l at step i the data $F_{t,i}^{(j)}$, $\mathbf{N}_i^{(j)}$, $V_{t,i}^{(j)}$, $V_{n,i}^{(j)}$ are known for all $j > l$ from the previous iteration, and from the current iteration for all $j < l$. Then compute the status of contact l by solving the Signorini Coulomb conditions (which monitor the transitions at $A^{(l)}$: sticking/sliding/detachment). Iterate until the last contact. Then apply a convergence test, that may be on the contact forces, ie, $|F_i^{k+1} - F_i^k|/|F_i^k| \leqslant \epsilon$, where k is the index of the iteration of the Signorini Coulomb loop at step i, or on the distance to the Signorini graph. If it is not satisfied redo the calculations for all contacts. One may also choose to stop arbitrarily after a certain number of iterations, since in some practical cases convergence is not at all guaranteed and the algorithm could stuck at one step. The fact that the contacts are treated independently one after each other, guarantees at least that Coulomb friction law is satisfied by $V_{t,i+1}^{(j)}$ and $F_{t,i+1}^{(j)}$ at each contact at the end of the iteration process. This is not necessarily the case for other formulations, see Section 7.4. Such a cyclic procedure is similar to a nonlinear block Gauss Seidel algorithm (which is for instance used to solve QPs [5]). Some convergence results can be found in [113].

Remark 10:
- The dynamics is written from the Lagrange equations as in (2) whereas the dry friction law involves local quan-

tities at the contact points $\mathcal{A}_j$ as in (60). Consequently one has to calculate the Jacobians $J^{(j)}(q)$ to relate both (see Section 3.4.2). It is argued in [94] that $J^{(j)}(q)$ may be evaluated at various intermediate values of q, eg, q_i, or $q_i + h/2\dot{q}_i$, or even a value obtained from an iteration process to approximate q_{i+1} (in case of an implicit formulation of the dynamics one needs to implement a Newton-Raphson like algorithm to compute q_{i+1} and $\dot{q}_{i+1}$ at step i). This is however related to the curvature of the contacting surfaces, which regulates the change in $\nabla\phi$ from i to $i+1$.

- A discretization of the sweeping process with friction that generalizes that in [32] is proposed in [210], for $m > 1$. The inclusion (23) is merely written at each contact point.
- The sweeping process (second order) discretization does not resort to any calculation of the acceleration and the contact forces. Only velocities (and implicitly the contact impulses) are calculated at each time step. This allows one to avoid the problems related to Painlevé's paradoxes when friction is present (see Sections 7.2.2 and 8.1), ie, the computation of unbounded contact forces (for certain friction coefficients and configurations of the system). Indeed as shown in [200,201] on a simple example, the force impulse remains bounded when the orbit crosses so-called critical points. The so-called *impacts without collisions*, or *tangential impacts* (which are of a different nature from the contact force unboundedness problem) are handled via the principle of maximal dissipation: at each time step, the tangential impulse is calculated so that it maximizes the dissipated frictional energy.

7.2.3 Simulation results

Extensive numerical tests have been performed by Moreau on granular matter [78,205]. Simulation of granular matter (sand piles, planetary rings) is a difficult subject, essentially due to the very large number of degrees of freedom. It is also difficult to make experiments and to compare them with simulations (try to follow the motion of a sand grain in a sand pile!). Only macroscopic phenomena may be checked (resulting pressures, average stress tensors, distribution of reaction forces), and may be expected to be robust with respect to numerics [94]. It is even difficult to make numerical simulations, and compliant models may simply be impossible to use, see Section 4.2. On the other hand, this is a topic with major applications in industry and in theoretical physics. In [78], a 2D vertically shaken cylindrical vessel is simulated. It contains 3999 beads with diameter 0.2 cm and one bead with diameter 0.5 cm (ie, $n = 4000$ in (2)!). In [205], another test is made with 2000 beads with diameter 0.2 cm and 200 beads with diameter 0.02 cm. A 3-parameter impact law is chosen (e_n, e_t, f), where e_t is a tangential restitution coefficient [93]. The numerical integration allows one to test the influence of the physical parameters on the global behavior, and to verify if some phenomena like clusterization, fluidization, bulk segregation, convection effects, occur and under which conditions.

Remark 11:
The dynamical effects can be quite important in the overall

motion even for almost static systems, because they permit to simulate possibly sparse events that have a crucial influence on the final configuration.

7.2.4 Related algorithms

NSCD methods: The work in [116], that is part of the NSCD method, is dedicated to the simulation of a continuous medium, approximated by a mesh of nodes. This is called the *distinct elements method*, in order not to confuse with the classical Finite Element Method (FEM). The nodes collide with the obstacle while the whole structure deforms. The inertia matrix is constant. Various discretization procedures based on θ- and θ-Euler methods are compared. Let us recall that the θ-method yields $\int_{t_i}^{t_{i+1}} f(t)dt \approx h[\theta f(t_{i+1}) + (1 - \theta)f(t_i)]$. The θ-Euler method consists of discretizing the Lagrange equation in (2) by a θ-method, and the position by a Euler method. A modified θ-method is also tested which consists of using a θ-method, then replacing q_{i+1} by $q_{i+1} + h(1 - \theta)\dot{q}_{i+1}$ in the contact relations. In particular the influence of the discretization of the complementarity conditions (second line in Eq. (2)) is discussed in [116]. The term $\phi^{(j)}(q(t))$, which represents the local distance between node j and the constraint, is discretized at each node as (see Remark 12):

$$\phi_{i+1}^{(j)} = \phi_i^{(j)} + h\theta[\nabla\phi_{i+1}^{(j)}]^T\dot{q}_{i+1} + h(1 - \theta)$$
$$\times[\nabla\phi_i^{(j)}]^T\dot{q}_i, \quad (\theta\text{-method})$$

$$\phi_{i+1}^{(j)} = \phi_i^{(j)} + h[\nabla\phi_{i+1}^{(j)}]^T\dot{q}_{i+1},$$

$$(\theta\text{-Euler and modified } \theta\text{-methods}). \tag{61}$$

The reader will see that contrary to Eqs. (49), (50), and (69), (70) which base on the analytical form of $[\nabla\phi^{(j)}(q)]^T\dot{q}$ and on the calculation of this expression using q_i or q_{i+1} and $\dot{q}_{i+1}$ or $\dot{q}_i$, the expressions in Eq. (61) are a time-discretization of $\phi^{(j)}(t)$. As shown in [155] this may have important consequences on the numerical results (eg, concerning the calculation of Lyapunov exponents). For the θ and θ-Euler methods, the complementarity conditions are at node j: $\phi_{i+1}^{(j)} \geqslant 0$, $N_{i+1}^{(j)} \geqslant 0$, $\phi_{i+1}^{(j)}N_{i+1}^{(j)} = 0$. For the modified θ-method, they are formulated with $\bar{\phi}_{i+1}^{(j)} = \phi_{i+1}^{(j)} + h(1 - \theta)\dot{\phi}_{i+1}^{(j)}$. It is noted that for the θ-method, $\phi_i^{(j)} = 0$ does not imply that $[\nabla\phi_{i+1}^{(j)}]^T\dot{q}_{i+1} = 0$, but this is the case for the θ-Euler and modified θ methods. In [94], it is pointed out that the position complementarity conditions with $\bar{\phi}_{i+1}^{(j)}$ are not recommended for large collections of bodies (like granular matter with $n \geqslant 1000$), because the correcting impulse that is calculated after a penetration to send back the system in Int(K) may be a non-negligible numerical artifact. Then a velocity complementarity formulation is preferable.

Remark 12: The idea of *consistency of the gap approximation with unilateral condition* is introduced in [94]. It means that $\bar{\phi}_i^{(j)} = 0$ and $\bar{\phi}_{i+1}^{(j)} = 0 \Rightarrow (\nabla\phi_{i+1}^{(j)})^T\dot{q}_{i+1} = 0$ is satisfied. For instance if one chooses simply $\bar{\phi} = \phi$, this consistency is satisfied only when $\theta = 1$. Other choices are possible

as the ones above. When this consistency property is not satisfied, some artificial numerical oscillations may appear [116]. The choice of the gap approximation is not important, provided it has the consistency property.

For the 2D case, dry friction is transformed in [116] via velocity slack variables into a set of two complementarity relations per contact, see Eq. (14). Then the whole stuff is assembled as for the time-stepping schemes in Section 7.4 (a technique already used in [114]). Lemke's algorithm is used to solve the LCP in [116]. Inelastic impacts with friction are treated in [94] by combining the resolution of complementarity conditions between the normal quantities $\bar{\phi}_i^{(j)}/h + V_{n,i}^{(j)}$ and $N_{i+1}^{(j)}$, and the resolution of Eq. (60). In a compact form, they might be written as $\text{SignCoul}(i, V_{i+1}^{(j)}, F_{i+1}^{(j)})$. Using the second formulation in Eq. (21) it may be shown that the mapping $(V,F) \mapsto \text{SignCoul}(i,V,F)$ is piecewise affine in 2D, and continuous raywise in 3D.

Simulation results: Due to its implicit formulation and the Signorini Coulomb loop, the NSCD method consists of two nested iteration loops at each time step i [94]. Although *large* integration steps can be used, it is therefore time consuming, which renders its use for real-time applications less easy. The numerical results obtained by the three discretization methods above, are compared to the exact solution in the case of the impact of two identical elastic bars moving on a line. They are further compared to a Newmark method with $\gamma = 0.5$ and $\beta = 0.25$ (these values are the smallest ones that guarantee A-stability [26] Ch 7.1.5). The Newmark and θ-method with $\theta = 0.5$ generate oscillations during the contact phase. Taking $\theta = 0.55$ damps out the oscillations (increasing θ renders the algorithm *more implicit*, and it is known [94] that implicit schemes behave nicely when applied to nonsmooth problems, although their implementation is more complex). The transition contact-non contact is too slow with the θ-Euler method. So the modified θ-method is the best one for this 1D case. For more complex case studies the θ-Euler method is said to be a good compromise between the θ-method (too smooth velocity) and the modified θ-method (oscillations). A 2D example from [211] (a disc colliding a rigid ground) is simulated with the θ-method ($\theta = 0.55$). The mesh has 99 nodes. The oscillations obtained with a Newmark scheme and an impact detection procedure as in [211], and which cannot be explained by acoustic waves propagation, are damped out. Evidently in these cases only a comparison with experimental results would really be telling (although the validation of a code cannot be made by comparison with experimental results [115], but by comparing various numerical schemes tested on benchmarks, or by comparing the discrete solution with the analytical one in simple cases). In [13], buildings made of blocks are simulated (an arch under various loads). Each block is approximated as in Fig. 3. Comparisons with experimental results with wood blocks led in the Ecole Supérieure de Mécanique de Marseille (F) are encouraging. It is important in this setting to recall that Coulomb friction is the main effect that precludes the existence of a unique equilibrium point (despite its dissipativity, see [18] pages 207-208), so that the obtained state depends on the history of loading. Physical situations are like this. The simulation of a Couette granular flow with 2400, 4000 and 16000 polydisperse disks can be found in [94]. An implicit Euler scheme has been used ($\theta = 1$ in Eq. (61)). The NSCD method seems to be particularly suitable for granular matter simulation and has been also used to study stress transmission and granular pressure [12]. Different quasistatic examples are presented in [115]: a dovetail assembly (4345 nodes, 49 contact nodes), a block sliding on a plane (4193 nodes, 65 contact nodes), and a pressure vessel (674 nodes, 54 contact nodes), a high presssure screw press (11933 nodes, 250 contact nodes), the extrusion of an aluminum cylinder (105 nodes, 21 contact nodes). Emphasis is put on the efficiency of various LCP solvers in terms of CPU times. Some preliminary results have been obtained for kinematic chains [208]. However such systems differ a lot from granular matter and distinct element systems, so that other algorithms (more accurate, with explicit contact forces and events calculation) may be preferred, see Section 6.7.2. The Contact Dynamics method has been implemented in a software called Simem3 [206].

3-parameter impact law: A numerical scheme inspired from [78] and [96] has been proposed in [212]. It applies to a 2D rigid body hitting a wall. Its focus is on the proper use of LCPs for impacts with friction and tangential restitution, so that 3 parameters are used (the 3-parameter impact law (f,e_n,e_t) is one of the most used in the literature [18] for impacts of spheres). A contact with Coulomb friction is treated with two LCPs, see Eq. (14). The LCPs are solved by a pivot algorithm. Experimental results of a rubber ball thrown in a box with wooden walls and steel plate are reported. The real motion of the ball between two impacts is recorded by a high-speed camera. The coefficients f and e_n were measured off-line and e_t fitted afterwards to get a good matching between experiments and simulations. The choice $e_n = e_t$ (which assures some energetical consistency [7,84]) is made in most experiments, although e_t is sometimes varied to get a better result. Further comparisons are made with experimental results available in the literature. The conclusions are mitigated and it is pointed out that a more complex impact model including moment impulse [93] could improve the results. The work is extended to 3D problems in [213], relying on an impact rule proposed by Moreau [78]. Roughly, it uses a variational formulation of this impact rule and of Coulomb friction (similar to the variational inequality in (21) and to (60)) to express the dynamics as a nondifferentiable equation, inspired by the works for quasi-static frictional contact in [214]. The problem is solved using a Newton-type algorithm specially devised for such nondifferentiable equations [215]. We note that such a procedure to formulate the complementarity relations from Eq. (21) is quite similar to what is used in [92,182] for the acceleration calculation. Similar experiments as in [212] are made and compared to the numerical results, showing good agreement when $e_t = e_n$.

7.3 Discretized second-order MDI

A numerical scheme especially suited to the second-order formulation presented in Section 3.4.3 is proposed by Paoli and Schatzman in [54-56,216,217].

7.3.1 Description of the scheme

The scheme is written in terms of positions only. So the relevant local metric is defined locally by the mass matrix of the system. Let us use the same notations as in Section 3.4.3. If we assume that ∂K and $q \mapsto M(q)$ are smooth (the second property being satisfied in most cases), it is possible to define locally a projection on K relatively to the Riemannian metric defined by $M(q)$.

Let F be a continuous function, consistent with Q, ie,

$$F(t,q,q,v,0)=M^{-1}(q)\bar{Q}(q,M(q)v,t) \quad \forall (q,v,t).$$

The scheme is given by:

$$q_{i+1}=-eq_{i-1}+(1+e)P_K\left(\frac{2q_i-(1-e)q_{i-1}+h^2F_i}{1+e}\right)$$

$$F_i=F\left(t_i,q_i,q_{i-1},\frac{q_{i+1}-q_{i-1}}{2h},h\right) \tag{62}$$

provided that q_{i+1} is uniquely defined in a neighborhood of K. The projection P_K is done in the kinetic metric.

In order to understand how this scheme approximates the system (26)–(29), we must say a few words about its construction. Let us consider the simplest formulation which corresponds to the case of convex constraints with a constant mass matrix. The dynamics in Eq. (33) is approximated by the implicit algorithm

$$\frac{q_{i+1}-2q_i+q_{i-1}}{h^2}+\partial\psi_K\left(\frac{q_{i+1}+eq_{i-1}}{1+e}\right) \ni F_i. \tag{63}$$

Remark 13:

The reader may notice the similarity between Eq. (63) and the last equation in (55). However this time the positions are involved, not the velocities. Notice that F_i is premultiplied by h^2 in Eq. (62) whereas the same term is multiplied by h in Eq. (55).

We can transform Eq. (63) by using a result of convex analysis. It can be shown that if K is a nonempty closed subset of $\mathbb{R}^d$ ($d \geqslant 1$), then with $y \in \mathbb{R}^d$ and $\lambda > 0$ given, the equation

$$x+\lambda\partial\psi_K(x) \ni y$$

admits a unique solution $x=P_K(y)$. Taking $x=q_{i+1}+eq_{i-1}/1+e$ in Eq. (63) one gets:

$$q_{i+1}=-eq_{i-1}+(1+e)P_K\left(\frac{2q_i-(1-e)q_{i-1}+h^2F_i}{1+e}\right) \tag{64}$$

which is exactly the first equality in Eq. (62).

Remark 14:

Notice that the proximations in Eqs. (56) and (58) can also be formulated as projections in the kinetic metric but the set of projection (ie, $V(q_{m,i})$ in (56) and $V(q_{i+1})$ in (58)) de-

pends on the step i. Here the set of projection is the same at each step: one always projects on K. Another common point between the schemes (55) and (62) is that they do not require the calculation of contact efforts.

It should be noted that the constraints are satisfied at each time step by the average position $\bar{q}_i=q_{i+1}+eq_{i-1}/1+e$. Moreover, if $\bar{q}_i$ belongs to Int(K), equation (64) reduces to

$$q_{i+1}=2q_i-q_{i-1}+h^2F_i \tag{65}$$

which is a classical second-order approximation for the equation of the free motion of the system. Furthermore, the constraints on positions and the impact law are taken into account at the same time by using the average position $\bar{q}_i$. Due to the choice for F_i in Eq. (62), Eq. (64) is an implicit equation as soon as $\bar{Q}$ depends on $p=M(q)\dot{q}$. In such a case some iteration (Newton-Raphson like algorithm) has to be used at each step to compute q_{i+1} (see eg [109]). Notice that the formulation (57) and (58) used in the convergence proof of the discretized Moreau's sweeping process is also an implicit formulation. For ODEs or DAEs implicit methods are known to possess larger domains of conditional stability ([218], page 239). What about MDIs?

As pointed out above the scheme in Eq. (62) does not require the systematic detection of impact times and does not need to refine the time step when the discrete positions are close to the boundary of K. As an example let us consider the 1D model problem described by $K=\mathbb{R}^+$, $\bar{Q}(q,p,t)\equiv 0$ and the initial data $q(0)=1$, $\dot{q}(0)=-1$. The motion is given by

$$q(t)=1-t \quad \text{if } t\in[0,1], \quad q(t)=e(t-1) \quad \text{if } t\geqslant 1. \tag{66}$$
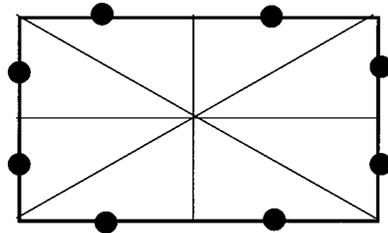
The algorithm (62) yields

$$q_{i+1}=-eq_{i-1}+\max[2q_i-(1-e)q_{i-1},0].$$

Let us choose $q_0=1$ and $q_1=1-h$. For $i\geqslant 1$, let $w_i=2q_i-(1-e)q_{i-1}$. For $i=1$, we have

$$w_1=2q_1-(1-e)q_0=(1+e)-2h$$

and $w_1>0$ if and only if $h<(1+e)/2$. From now on, let us assume that $h<(1+e)/2$. We define $n=\inf\{k\in\mathbb{N}^*:w_k<0\}$. We have $n\geqslant 2$ and for all $i\in\{2,\ldots,n\}$ we get

$$q_i-q_{i-1}=q_{i-1}-q_{i-2}=-h$$

● Candidate for contact

Fig. 3   Elementary block

hence $q_i = 1 - ih$.

One can also observe that $q_{n-1} > 0$ since $w_{n-1} = (1 + e)q_{n-1} - (1-e)h \geqslant 0$. Since $w_n < 0$, we have $q_{n+1} = -eq_{n-1} < 0$. Let us compute w_{n+1}:

$$w_{n+1} = 2q_{n+1} - (1-e)q_n = -2eq_{n-1}$$
$$- (1-e)(-eq_{n-2} + w_{n-1})$$
$$= -e(2q_{n-1} - (1-e)q_{n-2}) - (1-e)w_{n-1}$$
$$= -w_{n-1} \leqslant 0. \tag{67}$$

It follows that $q_{n+2} = -eq_n$ and

$$w_{n+2} = 2q_{n+2} - (1-e)q_{n+1} = -2w_n \geqslant 0 \tag{68}$$

$$q_{n+3} = -eq_{n+1} + w_{n+2} = e^2 q_{n-1} + w_{n+2} \geqslant 0.$$

Moreover, a straightforward induction gives $q_i = q_{n+2} + eh(i - n - 2)$ for all $i \geqslant n + 2$.

The approximate positions q_i do not satisfy the constraints at each time step, similarly to the discretized sweeping process. Of course, the average positions $\bar{q}_i = q_{i+1} + eq_{i-1}/1 + e$ belong to K for all $i \geqslant 1$ but not necessarily q_{i+1} and q_{i-1}. In this 1D example, at least one and at most two approximate positions are outside K.

It can proved that the penetration tends to 0 uniformly with respect to the time step since

$$\|q_i - \bar{q}_i\| \leqslant h\|\dot{q}_h\|_{L^\infty} = \mathcal{O}(h).$$

Moreover, modification has been proposed in [155] that assures penetration on one step only. It follows that the scheme is at most of order 1. This fact is not surprising since we approximate a second-order differential equation in terms of positions only. Such a choice means that we prefer to propose a fast scheme than a very accurate one. We should remind that we have to deal with highly nonlinear problems that are often very sensitive to initial data. Thus the accuracy of the scheme may be less important. But its convergence is always a crucial property.

7.3.2 Convergence

The convergence of this scheme is proved. The result has been established first in the case of convex contraints with a trivial mass matrix in [54] and [216], then an extension to non convex constraints (but still trivial mass matrix) is proposed in [217], finally the general case is studied in [55] and [56]. The proof follows the sketch described in Section 7.5 and is based on a rather natural geometrical idea: with an appropriate choice of local coordinates we can describe the set of admissible positions K by an half-space and obtain a simpler expression of the constraints and the projection on K. Nevertheless, the change of coordinates introduces other quadratic terms in the algorithm due to curvature effects. These new terms interact with the constraints and create serious difficulties in the study of the scheme.

7.3.3 Implementation

The scheme could seem to be difficult to implement since it requires to solve at each time step an implicit equation involving a projection on K for a Riemannian metric. But, in most of the cases, it is possible to describe the system with a choice of generalized coordinates such that the projection term is easy to compute exactly or to approximate. This scheme has been implemented on different examples. In [54], [219], and [220], a 1D model of tight joints is considered; in [54] and [220], the discretization of a guided beam is examined; and in [109], the motion of a slender bar is studied. For the two first cases the mass matrix of the system is constant, but not for the last one since a model in large deformations is considered. The results are compared to other numerical results obtained by systematic detection of impact times for the first and last examples and to experimental results for the last two examples. For instance in [109] free-motion dynamics is discretized with a Newmark scheme (implicit algorithm) with $\gamma = 0.5$ and $\beta = 0.25$ (such values guarantee A-stability [26], Ch 7.1.5), and impacts are detected by a root-finding algorithm (a Newton scheme initialized by the values computed at the preceding step). In all the cases, the performances of the scheme in Eq. (62) were satisfactory: the scheme is substantially faster than the detection method (almost 40% faster in [109], three times faster in [219]) and gives good approximation of the motion even with rather large time steps and/or on long time intervals (see [219]). This scheme has also been implemented in [151–153]. Therein they consider a spring-dashpot system with one degree-of-freedom. They compute its motion for two sets of data leading to periodic motions, by using the time-stepping scheme (62) and different event-driven schemes. They compare the numerical results to the analytical solution. In the two cases the scheme (62) is of order 1 and is faster than the event-driven schemes for a given time-step. On the other hand, due to its low order, the scheme (62) requires a larger computing time than the event-driven schemes, in order to approximate the solution to a given precision.

7.4 Velocity-impulse formulations

These works have been performed by a group composed of Stewart [33], Trinkle [84], Pang [85], Anitescu [86], and Potra [87]: different formulations have been proposed by Stewart *et al* following the works of Moreau and Lötstedt. They may be seen as variants of the semi-implicit Euler method for DAEs (which are very attractive for systems as in Eq. (6) [25], page 524). Some details have been given concerning friction formulation, see Eqs. (24) and (25). In Eq. (70), a polyhedral approximation of the friction cone is used, so that the conditions in Eq. (25) are modified. The algorithms have the general form:

$$\begin{cases} M(q_{i+1})(\dot{q}_{i+1} - \dot{q}_i) = \\ \quad \nabla\phi(q_i)\mathbf{N}_{i+1} + D(q_i)\beta_{i+1} + hQ(q_i, \dot{q}_i) \\ q_{i+1} = q_i + h\dot{q}_{i+1} \end{cases} \tag{69}$$

$$\begin{cases} \mathbf{N}_{i+1} \geqslant 0, \quad \nabla \phi(q_i)^T \dot{q}_{i+1} \geqslant 0, \quad \mathbf{N}_{i+1} \nabla \phi(q_i)^T \dot{q}_{i+1} = 0 \\ \beta_{i+1} \geqslant 0, \quad \lambda_{i+1} \mathbf{e} + D(q_i)^T \dot{q}_{i+1} \geqslant 0, \\ \quad \beta_{i+1}^T [\lambda_{i+1} \mathbf{e} + D(q_i)^T \dot{q}_{i+1}] = 0 \\ \lambda_{i+1} \geqslant 0, \quad f \mathbf{N}_{i+1} - \mathbf{e}^T \beta_{i+1} \geqslant 0, \\ \quad \lambda_{i+1} [f \mathbf{N}_{i+1} - \mathbf{e}^T \beta_{i+1}] = 0. \end{cases}$$

$$(70)$$

It is clear that the complementarity relations in Eq. (70) apply when the constraint is active only (including impact times), ie, if $\phi(q_i + h \dot{q}_i) < 0$. If $\mathbf{N}_{i+1} = 0$, then $\beta_{i+1} = 0$ as well. In Eqs. (69) and (70), $\mathbf{e}^T = [1,1,\ldots,1] \in \mathbb{R}^e$ where e is the number of edges of the polyhedral approximation of the friction cone, and f denotes here the friction coefficient. Hence $\beta \in \mathbb{R}^e$ as well. Indeed the friction cone is approximated by the polyhedral set $\hat{\mathcal{C}}(q) = \{\mathbf{N} \nabla \phi + D(q) \beta, \mathbf{N} \geqslant 0, \beta \geqslant 0, \mathbf{e}^T \beta \leqslant f \mathbf{N}\}$. The columns D^l of the matrix $D(q)$ are vectors that span the tangent subspace at the contact point. It is also assumed that there is always l and j such that $D^l = -D^j$. For instance in the case of a planar point-mass system $D^1 = \mathbf{t}$ and $D^2 = -\mathbf{t}$, and the tangential part of the contact reaction is given by $D(q)\beta = D^1 \beta_1 + D^2 \beta_2$, where $\mathbf{t}$ is the tangent direction at the contact point. Let $q^T = (x, y)$ with x the tangential coordinate. The two sets of complementary slackness conditions in Eq. (70) then become (in continuous time, so drop the indices):

$$0 \leqslant (\lambda + \dot{x}) \perp \beta_1 \geqslant 0$$

$$0 \leqslant (\lambda - \dot{x}) \perp \beta_2 \geqslant 0 \qquad (71)$$

$$0 \leqslant (f \mathbf{N} - \beta_1 - \beta_2) \perp \lambda \geqslant 0.$$

Assume for instance that there is a sliding motion with $\dot{x} > 0$. Then since $\lambda \geqslant 0$, one has $\beta_1 = 0$. Now necessarily $\lambda > 0$ since $\lambda \geqslant \dot{x} > 0$. Thus $\beta_2 = f \mathbf{N}$. If $\dot{x} < 0$ one would find $\beta_1 = f \mathbf{N}$ and $\beta_2 = 0$. Consider now $\dot{x} = 0$. One finds that if $\lambda \neq 0$, then $f \mathbf{N} = \beta_1 + \beta_2$. Since $\lambda > 0$ the third relation implies $\beta_1 + \beta_2 \leqslant f \mathbf{N}$: the contact reaction is inside the friction cone. Such a reasoning generalizes to 3D cases and the pyramid cone formulation. The conditions in Eq. (25), when discretized, yield a highly nonlinear complementarity problem [91].

It is assumed that $Q(q, \dot{q})$ derives from a potential energy. It can be shown that if $M(q) \equiv M$ then the set of discretized equations in (70) can be transformed in an LCP whose unknown is the vector $(\mathbf{N}_i, \lambda_i, \beta_i)$, and this LCP possesses a solution. Thus the algorithm can be advanced in time. Notice that λ_i and β_i are to be considered as impulses since they are proportional to forces times h. The last two sets of complementarity conditions in (70) represent an approximation of Coulomb model, where the friction cone is replaced by a polyhedral set (a pyramid if $e = 4$). They are the Kuhn-Tucker conditions for the maximum dissipation principle in Eq. (72). This is what allows one to express this model as a LCP, using tools from convex analysis, as pointed out in Remark 5. Otherwise in dimension 3, one would end up with a NCP to formulate dry friction. The tangential impulse is represented by β, while λ is a slack variable.

In relationship with Section 3.4.2, let us note that the complementarity conditions in Eq. (70) are equivalent to (Λ_q is the generalized impulse in coordinates q)

$$\Lambda_{q,i+1} = \arg \max_{z \in \hat{\mathcal{C}}(q_i)} \{-z^T \dot{q}_{i+1}\}. \qquad (72)$$

If they were formulated with $\nabla \phi(q_{i+1})$ and $D(q_{i+1})$ then one would get $\Lambda_{q,i+1} = \arg \max_{z \in \hat{\mathcal{C}}(q_{i+1})} \{-z^T \dot{q}_{i+1}\}$, yielding however a NCP instead of a LCP to be solved at each step.

Remark 15: The polyhedral approximation of $\mathcal{C}$ hampers to have the tangential velocity in a direction exact opposite to the tangential force, since the latter belongs to one of the corners of the polygon [89]. If the number of faces of the polyhedral set is increased, this effect is decreased, but the number of LCPs to be solved at each step is increased as well. Hence the interest of looking for solutions that do not approximate the cone $\mathcal{C}$ and minimize the number of equations to be used [83].

In case of multiple m contacts, the (generalized) friction cone (see eg, [18,28,201]) is taken as the sum of the friction cones at each contact, ie, $\mathcal{C} = \Sigma_i^m \mathcal{C}_i$ (in other words an element $R \in \mathcal{C}$ can be written as $R = \Sigma_i^m R_i$ with each $R_i \in \mathcal{C}_i$). The vectors $\nabla \phi$, λ, β, $\mathbf{N}$, are then simply constructed as the concatenation of the vectors for each contact. In the multi-contact case (say 2), it may therefore occur that $\dot{q}_{i+1}$ and $\Lambda_{q,i+1}$ are computed from Eq. (72) inside $\hat{\mathcal{C}}_1 + \hat{\mathcal{C}}_2$, whereas $\nabla \phi^{(1)T}(q_i) \dot{q}_{i+1} > 0$, ie, there is detachment from constraint 1 at t_{i+1}. However as we pointed out above this implies in turn that $\mathbf{N}_{i+1}^{(1)} = 0$ and $\beta_{i+1}^{(1)} = 0$ too. Thus the impulse $\Lambda_{q,i+1}$ is calculated inside $\hat{\mathcal{C}}_2(q_i)$. It is not clear at this stage whether this formulation would allow to get satisfaction of Coulomb law as $h \to 0$ or not. It seems that the main obstacle towards such a result is the possibility of Painlevé-like effects, which have not yet been understood for several contacts. Another problem that might occur as $h \to 0$ is that there may be some *cyclic* situations in which $\hat{\mathcal{C}}$ switches infinitely fast between several values, eg, $\hat{\mathcal{C}}_1 + \hat{\mathcal{C}}_2$, $\hat{\mathcal{C}}_1$ and $\hat{\mathcal{C}}_2$. What happens in the limit then? This problem is—at least at the *philosophical level*—of the same nature as that of multiple impacts. Interestingly enough, the conclusions in [125] about the relationship between the limit of a penalized problem with Coulomb friction and the LCP (as constructed in Section 3.3.3), hold also only for the one-contact sliding mode. Singular perturbation analysis is used and the stability of the so-called boundary layer is no longer equivalent to the existence and uniqueness of solution to the LCP(λ) when $m_s \geqslant 2$.

Impact rules: The impact rules are chosen according to Moreau's *maximal dissipation principle* [104] in [33] and [34], ie, Eqs. (69) and (70) represent a plastic generalized impact as in Eq. (12). In [84,86], the collision rule of Pfeiffer and Glocker [7] is chosen, ie, one solves a LCP to compute the velocity at the end of the compression phase, and another LCP for the velocity at the end of the expansion phase (recall that such an approach relies on Darboux-Keller's model for collisions [18]). Poisson coefficients of restitution are chosen. Energetical consistency holds when all coefficients are chosen equal to a unique value e_p. Roughly one considers

the same set of discrete-time equations as in (69) and (70) at q_i, replacing all quantities indexed by $i+1$ by $i+1/2$ for the compression phase, ie, one writes $M(\dot{q}_{i+1/2}-\dot{q}_i) = \nabla\phi(q_i)\mathbf{N}_{i+1/2}+D(q_i)\beta_{i+1/2}$. For the expansion phase, one replaces $\dot{q}_i$ by $\dot{q}_{i+1/2}$ (ie, one writes $M(\dot{q}_{i+1}-\dot{q}_{i+1/2})$) and the right-hand-side of the Lagrange dynamics becomes $\nabla\phi(q_i)\mathbf{N}_{i+1}+D(q_i)\beta_{i+1}+e_p(\nabla\phi(q_i)\mathbf{N}_{i+1/2} +D(q_i)\beta_{i+1/2})$. Finally, $q_{i+1}=q_i+h\dot{q}_{i+1/2}$ after these two *half-steps*. Let us note that the scheme in Eq. (55) incorporates the term $hM_{m,i}^{-1}Q_{m,i}$ in the calculation of the postimpact velocity, see Eq. (56). This is not the case for the algorithms in [84,86]. The term $hQ(q_i,\dot{q}_i)$ is however present in the right-hand-side for the compression phase calculation in [87]. Actually adding this term permits to better handle the problem posed by the capture of a constraint after an infinite series of rebounds (like the classical ball that rebounds on the ground), which is not the case if one applies an impact rule which neglects all non-impulsive terms. This can be understood with the 1D bouncing ball. In this case, one has $v_{i+1}=\text{prox}[V(q_{m,i}),v_i-h/2\,g]$. When v_i becomes very small, then $h/2\,g$ dominates $|v_i|$ and there is no numerical problem in continuing the calculations. Since the term $-h/2\,g$ is <0, $v_{i+1}=0$ in the subsequent steps. So the threshold parameter γ introduced in Section 6.3 is directly incorporated *via* h. On the other hand this numerical trick hampers the simulation of *reversed* accumulations of impacts [49,106] (a situation that might occur with a particle at rest on a table submitted to some excitation). However, such detachment conditions are met much less often than capture in practice.

There is, therefore, a significant difference between the schemes in [33] and [86]. Indeed in the second case, one integrates the motion and applies the impact rule when a collision has been detected, ie, $\phi(q_i)<0$. Then the algorithm computes $q_{i+1}=q_i+h\dot{q}_{i+1/2}$ after the two steps of the collision rule. When applying the maximal dissipation at i, the algorithm computes the quantities at $i+1$ by modifying abruptly the velocity direction if needed, but the forces $Q_i(q_i,\dot{q}_i)$ are part of the calculation (as in Moreau's scheme).

Remark 16:
- Similar backward Euler methods are used for the simulation of complex electrical circuits [160,221,222]. Notice that event-driven algorithms have also been used in this framework [223,224]. The discretization of so-called Linear Complementarity Systems [18,39,40] with an initial state jump as in Eq. (46) is studied in [159]. Sufficient conditions for consistency (ie, convergence of the discrete-time solution towards a solution of the continuous-time system) are given.
- Concerning multiple shocks with friction, it seems that both the algorithms in this section and Section VII A yield similar results, in the sense that the outcomes they provide are rather unpredictable: although they are formulated as a generalization of the frictionless plastic impact in Eq. (12), they may yield rebound depending on the parameters and data. The extension of the maximum dissipation principle and the generalization of Gauss' principle towards multiple collisions with friction, is far from being a trivial matter.

Several ways to formulate Eqs. (69) and (70) have been proposed by this group of authors. In [86], $M(q)\equiv M$ and $\nabla\phi(q_{i+1})\dot{q}_{i+1}$ is used. In [33], $\nabla\phi(q_i)\dot{q}_{i+1}$ and $M(q_{i+1})$ are used (which in practice leads to solving a NCP), but the work is essentially aimed at convergence proofs. In [85], the normal constraint is formulated as $\nabla\phi(q_i+h\dot{q}_i)q_{i+1}\geq\alpha$, which assures the respect of the constraints for the linearized dynamics. It is argued that the various quantities (mass matrix, $\nabla\phi$, matrix $D(q)$) can be calculated with $q_i+h\dot{q}_i$ or q_{i+1} or $q_i+q_{i+1}/2$. But the last two approaches yield NCPs, no longer LCPs, consequently more difficult to implement, although perhaps more accurate. These variations are not gratuitous, as they may have strong consequences on:
- The implementation of the algorithms and their speed, accuracy, robustness, etc.
- The mathematical analysis needed to prove their properties (consistency, stability, etc).

These influences may have opposite effects! Some of these properties are discussed in the related papers. With respect to Lötstedt's schemes, Stewart *et al* have improved the algorithm in several directions:
- **Friction:** 3-dimensional and 2-dimensional friction (Painlevé paradoxes for one contact are treated without resorting to any numerical trick).
- **Dissipativeness of the numerical scheme:** Stewart [33,88] proves that the algorithm in Eq. (70) with the maximal dissipation principle, is dissipative if $M(q)$ and $Q(q,\dot{q})$ are constant (linearized dynamics, or affine potential energy). The same is proved for the scheme in [85] provided K is convex and $\phi(q_i)\geq0$ for all i (no violation of the constraints). Anistescu and Potra [86] prove a similar result when $M(q)\equiv M$. However Stewart proves that the limit of his scheme is dissipative (which makes sense since otherwise it could not be a solution of the original problem).
- **Convergence:** Convergence (but not uniqueness) of the solutions of Eqs. (69) and (70) towards a solution of Eq. (2) has been proved only for the one frictional contact case [33,34,87], encompassing Painlevé paradoxes. In the multicontact case, it seems difficult to prove the convergence of the piecewise constant solutions of Eq. (70) towards a solution of Eq. (2) that satisfies Coulomb friction law. It is not clear whether this problem comes from the formulation of Coulomb friction itself (through the generalized cone $\mathcal{C}$ and the maximum dissipation), or from the interaction between the discretization and this formulation. In any case the problem of multiple contacts with friction still requires investigations. No convergence proof is available for the schemes in [84,86].

Some questions need to be still investigated: why solving LCPs at each step when the constraints are independent? What happens when systems are hyperstatic? Which problems does the multiplication of LCPs for solving Coulomb friction (polyhedral cone) create for real-time applications? Is the formulation of generalized friction for multiple contacts equivalent to Coulomb friction at each contact? Is it possible to recover sticking and sliding contacts from the

multicontact generalized formulation, in all cases? Solutions proposed for treating Coulomb with the revolution cone and new methods (homotopy etc) to solve NCPs need to be compared carefully with Moreau's Gauss-Seidel method. More generally it seems that there is a strong need for clarifying the domains of application and the performance of the various methods that allow to solve LCPs or NCPs (Lemke, NCP or LCP-functions and homotopy, nonsmooth Newton algorithms, QPs, interior point). See comments above and in Remark 9. From a general point of view, there is a lack of available numerical studies concerning the time-stepping schemes in Section 7.4 and comparisons with other methods.

7.5 Convergence studies

A very important mathematical study is the proof of convergence of the approximate (piecewise constant) discrete solution, towards a continuous function that is the solution of the continuous dynamics. These are the results of Monteiro-Marques, Stewart, Mabrouk and Paoli and Schatzman that we have outlined above. Let us define the approximate solution q_h by

$$q_h(t) = q_i + (t - t_i)\frac{q_{i+1} - q_i}{h} \quad \text{if } t \in [t_i, t_{i+1}).$$

The goal is to prove that one can extract a subsequence, still denoted q_h, which converges uniformly to a solution q of the Cauchy's problem. One shows first that the sequence $(\dot{q}_h)_h$ is bounded independently of h. Hence, Ascoli's theorem implies that there is a subsequence of $\{q_h\}_h$ which converges uniformly. Moreover, one establishes that the total variation of $\dot{q}_h$ is also bounded independently of h and with Helly's theorem it follows that the sequence $(\dot{q}_h)_h$ converges pointwise (except perhaps on a countable set of points) to $\dot{q}$, and $(\ddot{q}_h)_h$ converges weakly to $\ddot{q}$ in the space of vector valued measures.

Remark 17: This notion of convergence is needed for the acceleration, because it allows one to get convergence of functions towards (singular) measures like the Dirac measure. This is not possible with other convergence notions. This is clearly explained in [87] in a way accessible to non-mathematicians.

Then one has to check that the limit $(q, \dot{q})$ satisfies the constraints (ie, $\phi(q(t)) \geqslant 0$ for all t) and the impact and friction laws. Convergence ensures that a scheme gives a good approximation of the continuous motion even when there is an accumulation point in the set of impact times. This is a great advantage of the time-stepping methods presented in Sections 7.2, 7.3 and 7.4 over all other methods, especially over the event-driven algorithms based on a systematic detection of impacts which may fail whenever there is an infinite converging sequence of impacts, without special attention to such phenomena in the schemes.

7.6 General comments

We have seen that Moreau and the NSCD method use a first-order discretization, the works in this section and in Section 7.3 as well, whereas Lötstedt chose multistep methods. Actually, if the goal is to build a LCP(λ) at each step,

this is quite understandable since it may simplify the construction of this LCP. However there is another reason (that is also known in the DAE literature for problems with inconsistent initial data [29]): after each collision the order decreases to one. So if the system undergoes many collisions, why using higher order and/or multistep algorithms? Here comes into play the nature of the system to be simulated. It is argued in [86] that Runge-Kutta methods could be used, without further argumentation however. In [26] (page 264), it is indicated that multistep BDF methods result in a severe reduction of the step size h_i when a discontinuity in the state occurs, because it tries to fit a polynomial through this jump. Reinitialization techniques must be used. This, however, may apply to systems that consist of switching DAEs, but cannot reasonably constitute a general method for MDIs simulation.

The problem of drift away from the constraints is seldom discussed in these works. In [85], (see also [27]) it is pointed out that the normal complementarity conditions could be stated as $\phi(q_i) \geqslant 0$, $\mathbf{N}_i \geqslant 0$, $\phi(q_i)^T\mathbf{N}_i = 0$. However this yields a NCP. It can be solved as a series of LCPs, but this may drastically increase the number of LCPs to be solved at each step (if one considers the ones coming from approximated friction). In [91], it is pointed out that using the conditions $0 \leqslant \phi(q_{i+1}) \perp \mathbf{N}_{i+1} \geqslant 0$ does not work because the behavior is that of a random impact when a shock occurs. Projection of q_{i+1} back on the constraints is also alluded to as a remedy to drift. Lötstedt [27] points out that the velocity complementary slackness formulation in Eq. (49) may yield drift, and proposes to use Baumgarte's method as a remedy.

The integration step can be chosen as time-varying. Actually, the scheme in Eqs. (69) and (70) is of order 1 therefore not very accurate unless h is decreased a lot. Similarly as Moreau's scheme, it should therefore be preferred for systems with a lot of events only.

Notice that it may however be argued that such modifications of time-stepping schemes may make them lose their basic interest, which is to remain simple enough but robust. In other words, they should be able to detect the main characteristics of a system (strange attractors, periodic orbits etc), without necessarily providing very accurate results (for instance if the solutions are very sensitive to initial data, it may be hopeless to get accurate numerical results).

However the reader should keep in mind that some of the presented time-stepping schemes provide a new value of the state and contact forces, at each step i. In many cases, the resolution of LCPs or NCPs passes through a fixed-point calculation (Newton's like, smooth or nonsmooth), similarly to the proximations or projections operations. Except when there is an abrupt change in the system topology (deletion or activation of a constraint, which implies a change in the indices sets and consequently in the LCP dimension), the root at step i should not be too far from that at step $i-1$. Hence the apparent complexity of having to solve LCPs or NCPs at each step may be greatly simplified and accelerated in practice, provided h is taken small enough (and provided the algorithm used permits to fix the initial conditions at will, which is for instance not the case for Lemke's). More generally it is of interest to reuse the data of the previous steps to

decrease the computational efforts in all the modules of the software, like contact force calculation [185] and collision detection [225]. The combined BDF-Newton Raphson algorithm proposed in [67] for systems as in Eq. (7) uses iterations to calculate μ_i, $x_{1,i}$ and $x_{2,i}$ that satisfy the two equality constraints. Interestingly enough, it is pointed out in [67] that one iteration of the Newton-Raphson seems sufficient to get stability, from numerical experiments.

What is it that leads the authors to use one discretization procedure or another one? We already provided the reader with some elements of an answer, concerning Moreau's midpoint scheme. This may be the ability of constructing a LCP(λ_i) or NCP(λ_i) at each step (Lötstedt, Stewart *et al*), the combination of a second-order ODE discretization with shock dynamics (Paoli and Schatzman, Eqs. (65) and (64)). Recall also that the four classes of time-stepping schemes presented above, discretize different models. About DAEs, Petzold said that "BDF is so beautiful that it is hard to imagine something else could be better" [25], page 481. BDF combined with the reduction index in Eq. (7) has been applied to bilaterally constrained mechanical systems in [67] and has been shown to converge. This also may have been a motivation for the choice in Eqs. (49) and (50). There may also be significant discrepancies between the schemes constructed for mathematical convergence proofs, and the ones with more practical goals (compare Eqs. (57) and (55)). Moreau [28] (page 33) noted that *multiple step methods seem a priori inadequate since one looks for algorithms allowing to take impacts into account.* Consequently it seems that single step methods should be preferred to multistep ones. Then the advantage or drawbacks of Euler versus RK algorithms remains an open issue.

8 MATHEMATICAL ISSUES

Even for simple ODEs it is known that numerical methods that converge do not necessarily yield stable and robust results (see eg, [218], Chs 3 and 5). For instance an ODE as simple as $\dot{x} = -ax$ must be simulated with care when a becomes large [218] (page 238) ie, when the problem becomes stiff. Then implicit methods are known to provide much better results [25]. For ODEs analysis of round-off errors influence, problems of conditional stability, are understood. For MDIs, things complicate so drastically that mere convergence is in general very hard to prove [32,33,54,55]. If one wants ultimately to avoid ad hoc tricks in the course of the simulation, schemes with strong mathematical foundations seem mandatory. Additional problems arise with the non-uniqueness of solutions: uniqueness is assured only if the data (constraints, external forces and mass matrix) satisfy some stringent conditions like analyticity [44,45,49–52,106]. Coulomb friction may also result in non-uniqueness of solutions [59,200]. We reiterate that such phenomena, although they are due to the rigid body assumption and/or the friction model, are not to be considered as artifacts. Rather, these simplified models allow one to highlight physical phenomena which otherwise would have remained hidden. For instance, the unbounded force/bounded impulse phenomenon of Painlevé's problem (see Section 8.1) can hardly be

guessed if a penalized scheme is adopted. But in the course of a simulation with a large stiffness, the contact forces may become very large as well and create subsequent numerical inaccuracies. The rigid body analysis allows one to better understand such a phenomenon, and consequently to better circumvent it.

8.1 Is Painlevé a real obstacle?

One should not think that Painlevé paradoxes (ie, non-existence or non-uniqueness of solutions to the LCP(λ) for frictional contacts) occur only for unrealistic friction coefficient values. This depends a lot on the contact geometry [201]. From the results of [33] and [200], the Painlevé paradoxes are better understood, at least in the simplest case of one contact. Time-stepping impulse-velocity schemes do not face Painlevé paradoxes since impulse remains bounded in the vicinity of critical points. Moreover the maximal dissipation principle allows one to impose a velocity jump (the so-called *impacts without collisions*) that prevents the system from penetrating into the zones of non-existence of a solution to the LCP($\mathbf{N}$), even if force-acceleration schemes are used. Notice that a penalized problem with high stiffness will necessarily yield the same problem, ie, computation of very large interaction forces. Once again the study of the rigid body case allows one to highlight some crucial properties of the models which have an important consequence for numerical applications. In other words, a compliant contact model with Coulomb (or any variant of Coulomb) friction will show instability phenomena for certain configurations and large enough coefficient of friction, see eg, [179]. However for finite stiffnesses the contact force always remains bounded. It is only the study of the limit case that reveals the underlying structure and the fundamental nature of the problem. A comparative study of compliant models behavior when the stiffness is large (via singular perturbations analysis) and the LCP($\mathbf{N}$) has been made in [125]. Since in many practical applications the contact stiffnesses are finite but quite large (gears commonly possess contact stiffnesses $k = 10^{10}$ N/m or higher) such studies are far from being of pure theoretical interest. Results in [226] and [227] show that the *impacts without collisions* are the limit behavior of solutions of penalized problems (with finite but high stiffness): there are configurations in which the velocity varies very rapidly when $k < +\infty$ and jumps in the rigid body limit as $k \to +\infty$. Though the results in [33,200] are for the moment essentially limited to the case $m = 1$, they look quite promising. Some numerical results can be found in [227] for the scheme in Eqs. (69) and (70). As we pointed out at other places of this paper, one big challenge in multibody dynamics is a better modeling and understanding of dry friction in the multi-contact case. Painlevé-like phenomena require future investigations (such as the influence of the friction models, extension to multi frictional contacts).

For the classical Painlevé example (a slender rod that slides on a rigid ground, $m = 1$ and $n = 3$), the last problem that remains to be solved is uniqueness of the solutions. In particular as shown in [200] there exists in the phase plane a critical point that some trajectories may cross with un-

bounded velocities (but bounded impulses). However, just after this point, two solutions are possible: additional *rules* are necessary to decide which one has to be chosen. It is possible that studying penalized problems is going to provide us with such rules (which could hinge for instance on a certain relationship between the friction coefficient and the tangential and normal stiffnesses). This is reinforced by the fact that mathematical works [44,45,49,51,52] conclude that uniqueness holds only under very restrictive conditions on $\phi_j(q)$ and analyticity of all data. Since analyticity is much too stringent for practical applications, adding information into the model (but keeping the rigid body approach) seems mandatory. This should be done in a way similar to what has been done for collision rules [107,108]: use a better understanding of the physical process and lump these informations in some model parameters.

8.2 Is the discontinuity with respect to initial data a real obstacle?

Such phenomenon seems unavoidable and part of the dynamics. From a mathematical point of view (well-posedness), this does not necessarily preclude to get existence and uniqueness of solutions, see [40]. For stability of trajectories this may be an obstacle (it is known that Lyapunov stability is equivalent to continuity with respect to initial data [103] (page 124) uniformly in t over $\mathbb{R}^+$). This is the reason why the available results on control of systems as in Eq. (2) remain until now restricted to codimension one or to orthogonal constraints, see [18–20]. New notions of stability have to be studied. For numerical simulations, there is no way to strike right at the singularity of ∂K due to the finite accuracy of the calculations, so one can always apply a sequence of simple impacts and treat possible accumulations as usual. Evidently when getting close to the singularity the outcome becomes random [74,75,126]. This point added to the fact that even low-dimensional systems may possess a chaotic behavior [228,229], raises the question on whether it is useful to focus on the accuracy of the numerical algorithms. In this setting, the numerical computation of Lyapunov exponents is of primary importance in many studies, since they are used to detect chaos. As shown in [155], the discretization procedure has a strong influence on the numerical result.

9 COMMERCIAL SOFTWARE PACKAGES

Commercial softwares can be classified in two categories as a function of the frictional contact model they have adopted:

- Software packages with a penalized contact model.
- Software packages with unilateral contact model.

In this section, we will briefly discuss these two categories (and the contact models).

The penalized contact models are the most widespread mainly because these contact models are very easy to implement. Contact forces (normal and tangential) are considered as external forces, so they are just added to the right hand side of the dynamic equations. The normal contact force is usually modelled by a non-linear spring (ie, the elastic part)

and a damper (ie, the dissipative part) that is added between the two contacting bodies, in order to model a visco-elastic behavior of the materials. The normal contact force is then expressed in the following form:

$$F_n = F_{elastic} + F_{damper},$$

$$\text{with} \quad \begin{cases} F_{elastic} = K y^\alpha \\ F_{damper} = C \dot{y} \end{cases} \quad \text{if} \quad y \geq 0. \quad (73)$$

Here y denotes the penetration between the contacting bodies (positive means in contact), and $\dot{y}$ its first derivative as a function of time. The parameters K and α can be estimated using Hertz theory (see, eg, [5]). It must be remarked that this theory is only valid under the following conditions:

- Central impact with the gravity centers of the contacting bodies and the contact point on the same line.
- Quasistatic phenomenon.
- Elastic impact.

Even if Hertz theory is not generally valid (due to the preceding hypothesis), it can give a rough idea of the values of parameters K and α. K is sometime considered as a constant parameter (see for example Adams, MDI software) or is calculated as a function of the geometry of the contact bodies (see for example Mechanica Motion, PTC software). The normal elastic contact force can be considered to be as a good approximation. Things are much more complicated for the normal damping contact force. In fact, each software with a penalized contact model proposes its own formula for the damping factor C. We report here some expressions given in some widespread simulation tools that are valid for $y \geq 0$.

$$\begin{cases} F_{damper} = \text{regul}(d, C_1) \dot{y} \\ F_{damper} = C_2 K y^{3/2} \dot{y} \\ F_{damper} = C_3 y^{1/2} \dot{y} \\ F_{damper} = 2 e_r m \sqrt{\dfrac{\bar{K}}{m}} \dot{y}, \quad \text{with} \quad \bar{K} = \dfrac{3}{2} \sqrt{m g K^2}. \end{cases} \quad (74)$$

The function regul is plotted in Fig. 4, C_1, C_2 and C_3 are damping factors, e_r a parameter homogeneous to a restitution coefficient, and m the equivalent mass of the contacting bodies.

The expressions in Eq. (74) are used respectively in Adams (MDI), Mechanica Motion (PTC), SDS (Solid Dynamics) and Dads (LMS). All these expressions verify $C = 0$ for a zero penetration $y = 0$. Moreover, the normal contact force is continuous as a function of the penetration y. The expressions in Eq. (74) can lead in some cases to energy gains. For example, in case of the simple example of a ball bouncing on a plane, the dynamic equation in the direction normal to the plane can be written (using Hertz theory for the elastic component):

$$m \ddot{y} = -F_n + m g, \quad (75)$$

where $F_n = K y^{3/2} + C \dot{y}$ is the normal contact force. When C equals a constant value or $C = C_3 y^{1/2}$, the contact force F_n
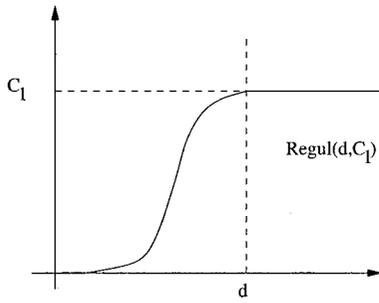
Fig. 4 The function regul

can be negative, that is to say attractive. As a consequence, the formulation of the normal contact force is changed, for example according to:

$$F_n = \mathrm{abs}(K y^{3/2} + C\dot{y}), \quad \text{or}$$
$$F_n = \max(0, K y^{3/2} + C\dot{y}).$$
(76)

It is remarked that already in the case of the bouncing ball, for the same ball movement (ie, same velocity before and after the shock) the normal contact forces predicted by the preceding models are very different. So the impact forces must be considered with care and cannot be used as a precise estimation for structural analysis. The determination of impact forces is still an active research area. To end this discussion, let us state that the parameters K, α, and C are difficult to predict and cannot be measured easily.

For the tangential contact force, penalized softwares usually adopt a regularized Coulomb law (see Fig. 5a) or the sign function (see Fig. 5b) In Fig. 5, F_t is the tangential contact force, v_t the relative tangential velocity of the contacting bodies, μ the friction coefficient, and *veltol* a parameter specified by the user.

Some recent softwares have started to implement unilateral contact models as an alternative to the penalized contact model. It is not easy to get reliable and precise informations on these contact models because vendors do not, in general, share implementation details with users. Consequently, we report here some very general informations concerning some of these softwares.

Probably the most famous software is Working Model (MSC). One of this software's particularities is its automatic

contact detection (in 2D and in 3D). It uses a Newton restitution coefficient in order to model dissipation during impacts (which may cause in certain cases energy gain), together with a fifth-order Runge-Kutta integration scheme (Kutta Merson). We assume that this program is an event-driven scheme, but are not sure that each event (like impact, lift-off or stick-slip transition) is accurately located. No precise informations could be obtained on the formulation of the frictional contact problem and on the way it is solved.

More recently, SDS (Solid Dynamics) has also adopted a 2D frictional contact model based on unilateral contact theory. Unlike Working Model, the contact model used in SDS is fairly well documented. Its general description can be found in [5]. It uses a Poisson restitution coefficient with a fifth-order Runge-Kutta integration scheme (Dormand and Prince). Each event is located accurately using a dichotomy procedure. The problem is written on the form of two coupled quadratic programming problems and is solved using the famous Moreau relaxation method (ie, Gauss-Seidel with a projection method).

The two preceding softwares are widely used in the industry. Of course other programs exist that support unilateral contact theory, like for example Simpack [192], but it seems that it is less complete than Working Model or SDS. For Simpack, only one reference could be traced on this software that only describes frictionless contacts. The problem is formulated in the form of a DAE and it is solved using a trial and error method (which may in practice be untractable).

All the commercial software packages are using an acceleration formulation (ie, event-driven formulation, with event detection). This kind of method is fast and reliable for problems with less than, say, a hundred of frictional contacts. If one plans to treat problems with thousands of contacts, one should use a time-stepping scheme (no event detection) to get an idea of the behavior of the system. (Even a rough idea because these time-stepping formulations are usually based on a first-order integration scheme.) But time-stepping schemes are research codes like LMGC of Moreau that has been extensively used for 15 years for the simulation of granular materials [13].

In conclusion, the penalized and the unilateral contact models have some advantages and some drawbacks, and the user has to be aware of them in order to choose the software
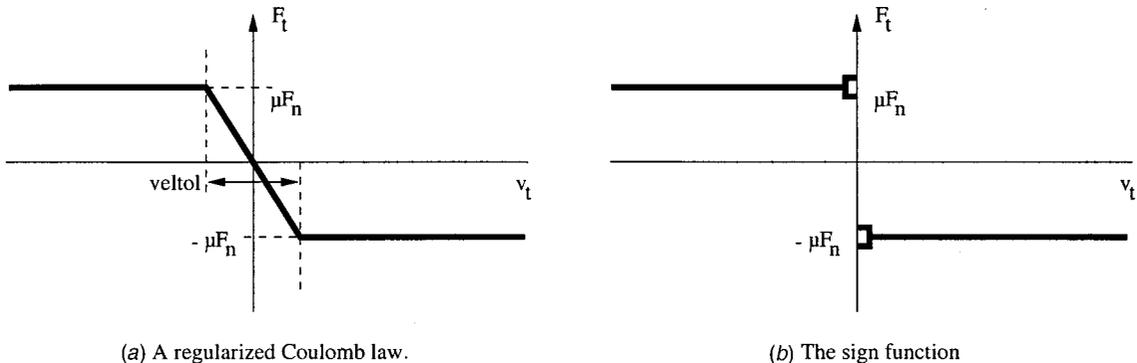

(a) A regularized Coulomb law.

(b) The sign function

Fig. 5 Examples of tangential contact force approximations

Table 1.   Advantages and drawbacks of a number of softwares

	Penalized contact model	Unilateral contact model	
		Event-driven schemes	Time-stepping schemes
Most famous softwares	• Adams (MDI) • Dads (LMS) • Mechanica Motion (PTC)	• Working Model 2D & 3D (MSC) • SDS (Solid Dynamics)	• LMGC (Moreau)
Advantages	• Model easy to implement • No problem of redundancies • No problem of impulse propagation (see also Newton cradle example)	• Restitution coefficient can be estimated • Sticking effect taken into account (real Coulomb law) • No oscillations of the contact forces • Effective for less than 100 of frictional contacts (good CPU time) • Very accurate	• Usable for thousands of frictional contacts (huge CPU time) • Can give a rough idea of the results
Drawbacks	• Contact parameters unknown (K, α, C) • Sticking effect not modeled • Stiff differential systems • Oscillations of the contact forces	• Redundancy varies during simulation • No impulse propagation (see Newton cradle example)	

to solve the problem at hand. To close this section on commercial softwares, let us summarize these advantages and drawbacks in Table I. This has to be considered as a complement of informations with respect to the material in Section 4.

Remark 18 (Vehicle crash dynamics and simulation): The general problem is quite different: find how to reconstitute the motion of accidents from some data (physical parameters, estimation of dissipated energies etc.). Details and references can be found in [93] and [230].

10   GENERAL CONCLUSIONS

"...unilateral problems are slowly noticed by the scientific community, but are migrating fast into industrial applications" [231]. It is clear that the development of a reliable and efficient software for nonsmooth multibody systems is the result of assembling various modules. We have chosen not to incorporate numerical examples in this paper. The main reason is that it is not our aim here to classify the approaches (if this is possible), but to introduce the reader to the field of numerical analysis and simulation of nonsmooth multibody systems. Despite the many open problems that remain to be solved before getting a satisfactory software, the rigid body approach is quite interesting. One of the reasons is that compliant contact models may really fail in providing reliable schemes.

In this article, the state of the art about numerical simulation of unilaterally constrained mechanical systems has been described. The general message is that the analysis of such systems (mathematical, numerical, system theoretical) is by far not a simple extension of that of DAEs. Also the interconnections between modeling and these topics is an important point. It is not possible to decouple the mathematical, modeling and numerical problems. For instance multiple impacts create deep modeling problems. Even the frictionless case has not yet been solved in its generality, although the work in [108] seems quite promising. As pointed out above in this article, some authors argue that statistical modeling should be investigated. In regard to this point, it could be that

a pure *continuous* discretization of the dynamics in Eqs. (2) and (3) is not sufficient in general. A possible path is to resort to *hybrid* simulators which incorporate not only the classical event-driven or time-stepping algorithms, but also some *high-level rules* that would guide the user in the choice of future events (detachment, impact, sticking etc). This can be done via the construction of a tableau containing possible choices and their probability to occur. Such data may be obtained from experiments. This may constitute a way to connect experiments, modeling and numerics via some logic included in the software [65] (page 175) The determination of the qualitative properties of a system may be sufficient in certain applications (ie, one requires only to detect some dynamical invariants in the system, like attractors). Thus those schemes with less accuracy but high robustness and consistency results may prove quite useful (like time-stepping schemes). But qualitative properties may not be sufficient: quantitative ones can be crucial in industrial applications (eg, aeronautics where long-run simulations may be needed, sometimes one wants to observe the evolution of a system on several hours or days). Notice that the length of the simulation has to be modulated by the number of events occurring during the simulation: some very short motions (like in circuit breakers) can involve a great number of events on a very small time interval. Hybrid schemes that switch between various integrators depending on the phase of motion [232] may also represent a promising path for simulation of complex multibody systems.

Let us end this article by citing some of the important problems which still deserve deep research study (this list does not pretend to be exhaustive):

• Comparison and determination of the domains of applications of complementarity problems solvers.
• More generally, determination of the domains of applications of the algorithms on suitable benchmarks allowing one to detect their capabilities with respect to various dynamical situations (periodic/non-periodic motion, sensitivity to initial data or not, large or small N, etc)
• Better understanding of Painlevé-like phenomena in higher dimensions and multiple contacts.
• Modeling of mutiple impacts with or without friction.

- Incorporation of more sophisticated dry friction models that satisfy the principle of maximum dissipation, in a mathematical programming framework.

ACKNOWLEDGMENT

REFERENCES

[1] Schiehlen W (1997), Multibody system dynamics: Roots and perspectives, *Multibody Syst. Dyn.* **1**(2), 149–188.

[2] Ambrosio JAC (2000), Rigid and flexible multibody dynamics tools for the simulation of systems subjected to contact and impacts conditions, *Eur. J. Mech. A/Solids* **19**, Special Issue, S23–S44.

[3] Euromech Colloquium 404 (1999), *Advances in Computational Multibody Dynamics*, European Mechanics Society, Lisboa, Sept 20–23, http://www.lemac.ist.utl.pt/EUROMECH

[4] Abadie M (1998), Simulation dynamique de mécanismes: Prise en compte du contact frottant, PhD Thesis, Univ of Montpellier 2, France, Sept.

[5] Abadie M (2000), Dynamic simulation of rigid bodies: Modeling of frictional contact, In: *Impacts in Mechanical Systems: Analysis and Modeling*, B Brogliato (ed). Springer Verlag, Berlin, Lecture Notes in Physics, 62–144.

[6] Babitsky VI (1998), *Theory of Vibro-Impact Systems and Applications*, Springer-Verlag, Foundations of Engineering Mechanics.

[7] Pfeiffer F and Glocker Ch (1996), *Multibody Dynamics with Unilateral Contacts*, Wiley Interscience, Series in Non-linear Dynamics.

[8] Studny D, Rittel D, and Zussman E (1999), Impact fracture of screws for disassembly, *ASME J. Manuf. Sci. Eng.* **121**, 118–125.

[9] Balandin DV, Bolotnik NN, and Pilkey WD (2000), Optimal protection from impact and shock: Theory and methods, *Appl. Mech. Rev.* **53**(9), p 237.

[10] Brilliantov NV, Spahn F, Hertzsch JM, and Pöschel T (1996), Model for collisions in granular gases, *Phys. Rev. E* **53**(5), 5382–5392.

[11] Radjai F, Jean M, Moreau JJ, and Roux S (1996), Force distributions in dense two-dimensional granular systems, *Phys. Rev. Lett.* **77**, 274.

[12] Radjai F and Wolf DE (1998), Features of static pressure in dense granular media, *Granular Matter* **1**(1), 3–8.

[13] Acary V and Jean M (1999), Numerical simulation of monuments by the contact dynamics method, available at http://chimay.esm2.imt-mrs.fr/Vince/Publi/texteHTML/texteHTML.html

[14] Brogliato B and Zavala-Rio A (2000), On the control of complementary-slackness mechanical juggling systems, *IEEE Trans. Autom. Control* **45**(2), 235–246.

[15] Hurmulzu Y, Génot F, and Brogliato B (2001), Modelling, stability and control of bipedal locomotion systems—A general framework, Internal Report, Laboratoire d'Automatique de Grenoble, INRIA Research Report No 4290.

[16] Mata Jimenez M and Brogliato B (2002), Analysis of PD and nonlinear control of mechanical systems with dynamic backlash, submitted to Journal of Control and Vibrations, in press.

[17] Awrejcewicz J, Tomczak K, and Lamarque CH (1999), Controlling systems with impacts, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **9**(36), 547–553.

[18] Brogliato B (1999), *Nonsmooth Mechanics, Second Edition*, Springer Verlag, Communications and Control Engineering Series, London (first edition in LNCIS 220, 1996), Erratum and addenda available at http://www-lag.ensieg.inpg.fr/publications.html

[19] Brogliato B, Niculescu SI, and Monteiro Marques MDP (2000), On the tracking control of a class of complementary slackness mechanical systems, *Syst. Control Lett.* **39**(4), 255–266.

[20] Brogliato B, Niculescu SI, and Orhant P (1997), On the control of finite-dimensional mechanical systems with unilateral constraints, *IEEE Trans. Autom. Control* **42**(2), 200–215.

[21] Kos J, ten Dam AA, Pruis GW, and Vankan WJ (1998), Efficient harmonisation of simulation competence in a CACE working environment, *Proc 5th Int Workshop on Simulation for European Space Programmes*, SESP'98, 3–5 Nov, ESTEC, NL.

[22] van Woerkom PThLM, de Boer A, Ellenbroek MHM, and Wijker JJ (1994), Developing algorithms for efficient simulation of flexible space manipulator operations, IAF-94-A.4.032, *45th Congress of the Int Astronautical Federation*, Oct 9–14, 1994, Israel.

[23] van Sterkenburg RP and ten Dam AA (1999), The scenario management tool SMARTFED for real-time interactive high performance networked simulations, *High-Performance Computing and Networking, Proc 7th Int Conf HPCN Europe 99*, Amsterdam, The Netherlands, Lecture Notes in Computer Science, vol 1593, Springer.

[24] Chatterjee A and Ruina AL (1998), Two interpretations of rigidity in rigid body collisions, ASME Winter Annual meeting, November, Dallas, TX, 1997. Also in *ASME J. App. Mech* **65**, 894–901, December.

[25] Hairer E and Wanner G (1996), *Solving Ordinary Differential Equations II; Stiff and Differential-Algebraic Problems, 2nd Edition*, Springer-Verlag, Series in Computational Mathematics.

[26] Garcia de Jalon J and Bayo E (1994), *Kinematic and Dynamic Simulation of Multibody Systems: The Real-time Challenge*, Springer Verlag, Mechanical Engineering Series.

[27] Lötstedt P (1984), Numerical simulation of time dependent contact and friction problems in rigid body mechanics, *SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput.* **5**, 370–393.

[28] Moreau JJ (1986), Dynamique de systèmes à liaisons unilatérales avec frottement sec éventuel; essais numériques, Tech Note 85-1, LMGC, Montpellier, France.

[29] Brenan KE, Campbell SL, and Petzold LR (1989), *Numerical Solutions of Initial-Value Problems in Differential-Algebraic Equations*, Elsevier Science Publ, NY.

[30] Rockafellar RT (1970), *Convex Analysis*, Princeton Landmarks in Mathematics, Princeton Univ Press, 10th printing 1997.

[31] Bressan A (1959), Questioni di regolarita e di unicita del moto in presenza di vincoli olonomi unilaterali, *Rend. Sem. Mat. Univ. Padova* **29**, 271–315.

[32] Monteiro-Marques MDP (1993), *Differential Inclusions in Nonsmooth Mechanical Problems: Shocks and Dry Friction*, Birkhauser, Boston, PNLDE 9.

[33] Stewart DE (1998), Convergence of a time-stepping scheme for rigid-body dynamics and resolution of Painlevé's problem, *Arch. Ration. Mech. Anal.* **145**, 215–260.

[34] Stewart DE (1997), Existence of solutions to rigid body dynamics and the Painlevé paradoxes, *C. R. Acad. Sci. Paris* **325**, Série I, 689–693.

[35] Mabrouk M (1998), A unified variational model for the dynamics of perfect unilateral constraints, *Eur. J. Mech. A/Solids,* **17**(5), 819–842.

[36] Mabrouk M (1998), Liaisons unilatérales et chocs élastiques quelconques: Un résultat d'existence, *C. R. Acad. Sci Paris* **326**, Série I, 1353–1357.

[37] Carriero M and Pascali E (1982), Uniqueness of the one-dimensional bounce problem as a generic property in $L_1([0,T],\mathbf{R})$, *Bolletino U.M.I.* **6**(1-A), 87–91.

[38] Carriero M and Pascali E (1985), Uniqueness for the elastic one-dimensional bounce problem, *Proc of Nonlinear Variational Problems*, *Res. Notes Math.,* **127***, 88–91.

[39] Heemels WPMH (1999), Linear Complementarity Systems: A study in hybrid dynamics, PhD Thesis, Electrical Eng Dept, Eindhoven Univ of Technology, The Netherlands, ISBN 90-386-1690-2.

[40] Heemels WPMH, Schumacher JM, and Weiland S (2000), Linear complementarity systems, *SIAM (Soc. Ind. Appl. Math.) J. Appl. Math.* **60**(4), 1234–1269.

[41] van der Schaft A and Schumacher JM (1998), Complementarity modeling of hybrid systems CWI Report, Dept Operations Research, Statistics and Systems Theory, BS-R9611, August, NL. Also in *IEEE Trans. Autom. Control* **43**(4), 483–490.

[42] van der Schaft A and Schumacher JM (2000), *An Introduction to Hybrid Dynamical Systems*, Springer Verlag, London, LNCIS 251.

[43] Lötstedt P (1982), Mechanical systems of rigid bodies subject to unilateral constraints, *SIAM (Soc. Ind. Appl. Math.) J. Appl. Math.* **42**(2), 281–296.

[44] Percivale D (1985), Uniqueness in the elastic bounce problem, *J. Diff. Eqns.* **56**, 206–215.

[45] Percivale D (1991), Uniqueness in the elastic bounce problem II, *J. Diff. Eqns.* **90**, 304–315.

[46] Percivale D (1986), Bounce problem with weak hypotheses of regularity, *Ann. Mat. Pura. Appl.* **143**(4), 259–274.

[47] Buttazo G and Percivale D (1981), The bounce problem on n-dimensional Riemannian manifolds, *Atti. Accad. Naz. Lincei, Cl. Sci. Fis., Mat. Nat., Rend.* **70**, 246–250.

[48] Buttazo G and Percivale D (1983), On the approximation of the elas-

tic bounce problem on Riemannian manifolds, *J. Diff. Eqns.* **47**(2), 227–245.

[49] Ballard P (2000), The dynamics of discrete mechanical systems with perfect unilateral constraints, *Arch. Ration. Mech. Anal.* **154**(3), 199–274.

[50] Ballard P (1999), Dynamique des systèmes mécaniques avec liaisons unilatérales parfaites, *C. R. Acad. Sci., Ser. IIb: Mec., Phys., Chim., Astron.* **327**, 953–958.

[51] Schatzman M (1978), A class of nonlinear differential equations of second order in time, *Nonlinear Analysis, Theory, Methods and Applications* **2**(3), 355–373.

[52] Schatzman M (1998), Uniqueness and continuous dependence on data for one-dimensional impact problem, *Math. Comput. Modell.* **28**(4–8), 1–18.

[53] Paoli L and Schatzman M (1993), Mouvement à un nombre fini de degrés de liberté avec contraintes unilatérales: cas avec perte d'énergie, *Modèl. Math. Anal. Numér.* **27**(6), 673–717.

[54] Paoli L (1993), Analyse numérique de vibrations avec contraintes unilatérales, PhD Thesis, Univ Claude Bernard-Lyon 1, Laboratoire d'Analyse Numérique, France.

[55] Paoli L and Schatzman M (2002), A numerical scheme for impact problems, preprint de l'Equipe d'Analyse Numérique, UMR 5585, June, SIAM on Numerical Analysis, in press. (Available at http://numerix.univ-lyon1.fr/publis/publiv/1999/publis.html)

[56] Paoli L and Schatzman M (1999), Approximation et existence en vibro-impact, *C. R. Acad. Sci., Ser. I: Math.* **329**, Série I, 1103–1107.

[57] Paoli L (2000), An existence result for vibrations with unilateral constraints: case of a non-smooth set of constraints, *Math. Methods Appl. Sci.* **10**(10), 815–832.

[58] Ivanov AP (1995), On multiple impacts, *Prikl. Mat. Mekh.* **59**(6), 930–946; (Transl. in *ASME J. Appl. Mech.* **59**(6), 887–902).

[59] Schatzman M, Lamarque CH, and Bastien J (1999), An ill-posed mechanical problem with friction, *Eur. J. Mech. A/Solids* **18**, 415–420.

[60] Hilding D, Klarbring A, and Peterson J (1999), Optimization of structures in unilateral contact, *Appl. Mech. Rev.* **52**(4), 139–160.

[61] Featherstone R (1987), Robot Dynamics Algorithms, Kluwer Academic Publ, Boston.

[62] Wehage RA and Haug EJ (1982), Generalized coordinate partitioning for dimension reduction in analysis of constrained dynamic systems, *J. Mech. Des.* **104**, 247–255.

[63] Baraff D (1996), Linear-time dynamics using Lagrange multipliers, *Computer Graphics Proc*, Annual Conference Series, 137–146.

[64] Mattsson SE, Otter M, and Elmqvist H (1999), Modelica hybrid modeling and efficient simulation, *Proc of 38th IEEE Conf on Decision and Control*, Phoenix AZ, Dec, 3502–3507.

[65] ten Dam AA (1997), Unilaterally constrained dynamical systems, PhD Thesis, Rijksuniversiteit Groningen, NL, http://www.ub.rug.nl/eldoc/dis/science/a.a.ten.dam/

[66] ten Dam AA (1992), Stable numerical integration of dynamical systems subject to equality state-space constraints, *J. Eng. Math.* **26**, 315–337.

[67] Gear CW, Leimkuhler B, and Gupta GK (1985), Automatic integration of Euler-Lagrange equations with constraints, *J. Comput. Appl. Math.* **12-13**, 77-90.

[68] ten Dam AA, Dwarshuis KF, and Willems JC (1997), The contact problem for linear continuous-time dynamical systems, a geometric approach, *IEEE Trans. Autom. Control* **42**(4), 458–472.

[69] Delassus E (1923), Sur les lois du frottement de glissement, *Bull. Soc. Math. France* **51**, 22–33.

[70] Moreau JJ (1963), Les liaisons unilatérales et le principe de Gauss, *C. R. Acad. Sciences Paris* **256**, 871–874.

[71] Moreau JJ (1966), Quadratic programming in mechanics: Dynamics of one sided constraints, *SIAM J. Control* **4**(1), 153–158.

[72] Gear CW and Petzold LR (1984), ODE methods for the solution of differential/algebraic systems, *SIAM J. Numer. Anal.* **21**(4), 716–728.

[73] Ivanov AP (1987), Impacts in a system with certain unilateral couplings, *Prikl. Mat. Mekh.* **51**(4), 559–566 (Transl. in *J. Appl. Math. Mech.* **51**(4), 436–442 (1998)).

[74] Paoli L and Schatzman M (1999), Moreau's rule and the penalty method, *Proc of IUTAM Symp on Unilateral Multibody Contacts*, Kluwer, F Pfeiffer and Ch Glocker (eds), 35–44.

[75] Paoli L and Schatzmann M (2001), Penalty approximation for non smooth constraints in vibroimpact, to *J. Diff. Eqs.* **177**, 375–418. (Preprint de l'Equipe d'Analyse Numèrique, UMR 5585 Maply, available at http://numerix.univ-lyonl.fr/publis/publiv/2000/publis.html)

[76] Carnot L (1803), *Principes Fondamentaux de l'Equilibre et du Mouvement*, Paris, Bachelier.

[77] Chatterjee A (1999), On the realism of complementarity conditions in rigid body collisions, *Nonlinear Dyn.,* **20**, 159–168.

[78] Moreau JJ (1994), Some numerical methods in multibody dynamics: Application to granular materials, *Eur. J. Mech. A/Solids* **13**(4), 93–114.

[79] Klarbring A (1986), A mathematical programming approach to three dimensional contact problems with friction, *Comput. Methods Appl. Mech. Eng.* **58**, 175–200.

[80] Klarbring A and Björkman G (1988), A mathematical programming approach to contact problems with friction and varying contact surface, *Comput. Struct.* **30**(5), 1185–1198.

[81] Lootsma YJ, van der Schaft AJ, and Camlibe MK (1999), Uniqueness of solutions of relay systems, *Automatica* **35**(3), 467–478.

[82] Pang JS and Trinkle JC (1996), Complementarity formulations and existence of solutions of dynamic multi-rigid-body contact problems with Coulomb friction, *Math. Program.* **73**, 199–226.

[83] Glocker Ch (1999), Formulation of spatial contact situations in rigid multibody systems, *Comput. Methods Appl. Mech. Eng.* **177**, 199–214.

[84] Anitescu M, Potra FA, and Stewart DE (1999), Time-stepping for three-dimensional rigid body dynamics, *Comput. Methods Appl. Mech. Eng.* **177**, 183–197.

[85] Stewart DE and Trinkle JC (1996), An implicit time-stepping scheme for rigid body dynamics with inelastic collisions and Coulomb friction, *Int. J. Numer. Methods Eng.* **39**(5), 2673–2691.

[86] Anitescu M, Cremer JF, and Potra FA (1996), Formulating three-dimensional contact dynamics problems, Mech. Struct. Mach. **86**(4).

[87] Stewart DE (1999), Time stepping methods and the mathematics of rigid body dynamics, in *Impact and Friction* A Guran (ed), Birkhauser (available at http://www.math.uiowa.edu/dstewart/).

[88] Stewart DE (1998), Convergence of a time-stepping scheme for rigid body dynamics and resolution of Painlevé's problem; Summary, (available at http://www.math.uiowa.edu/dstewart/).

[89] Trinkle JC, Pang JS, Sudarsky S, and Lo G (1997), On dynamic multi-rigid-body contact problems with Coulomb friction, *Z. Angew. Math. Mech.* **77**, 267–279.

[90] Pang JS and Stewart D (1999), A unified approach to discrete frictional contact problems, *Int. J. Eng. Sci.* **37**, 1747–1768.

[91] Stewart DE (2000), Rigid-body dynamics with friction and impacts, *SIAM Rev.* **42**(1), 3–39.

[92] Wösle M and Pfeiffer F (1996), Dynamics of multibody systems containing dependent unilateral constraints with friction, *J. Vib. Control* **2**, 161–192.

[93] Brach RM (1991), *Mechanical Impact Dynamics*, John Wiley, New York.

[94] Jean M (1999), The non-smooth contact dynamics approach, *Comput. Methods Appl. Mech. Eng.* **177**, 235–277.

[95] Moreau JJ (1966–1967) *Fonctionnelles Convéxes*, Séminaire sur les Equations aux Dérivées Partielles, Collège de France, Paris.

[96] Moreau JJ (1988), Unilateral contact and dry friction in finite freedom dynamics, in [233], 1–82.

[97] Jean M (1993), Simulation numérique des problèmes de contact avec frottement, *Matériaux et Techniques* **1,2,3**, 22–32.

[98] Moreau JJ (1983), Liaisons unilatérales sans frottement et chocs inélastiques, *C. R. Acad. Sci., Ser. II: Mec., Phys., Chim., Sci. Terre Univers* **296**(30 mai), série II, 1473–1476.

[99] Cholet C (1998), Chocs de solides rigides, PhD Thesis, Laboratoire d'Analyse Numérique–Université Paris 6 and Laboratoire des Matériaux et des Structures du Génie Civil–LCPC-CNRS, March.

[100] Moreau JJ (1985), Standard inelastic shocks and the dynamics of unilateral constraints, CISM Courses and Lectures, no 288, Springer-Verlag, 173–221.

[101] Moreau JJ (1999), Numerical aspects of the sweeping process, *Comput. Methods Appl. Mech. Eng.* **177**, 329–349.

[102] Kunze M and Monteiro Marques MDP (2000), An introduction to Moreau's sweeping process, In *Impacts in Mechanical Systems: Analysis and Modeling*, Springer Verlag, Berlin, Lecture Notes in Physics, B. Brogliato (ed), 1–60.

[103] Cartan H (1977), *Cours de Calcul Différentiel*, Hermann, Paris.

[104] Moreau JJ (1979), Application of convex analysis to some problems of dry friction, in *Trends in Applications of Pure Mathematics to Mechanics, Vol 2*, H Zorski (ed), Pitman Publ, London, 263–280.

[105] Panagiotopoulos PD (1995), Modeling of nonconvex nonsmooth energy problems. Dynamics hemivariational inequalities with impact effects, *J. Comput. Appl. Math.* **63**, 123–138.

[106] Bressan A (1959), Incompatibilità dei teoremi di esistenza e di unicità del moto per un tipo molto comune e regolare di sistemi meccanici, *Rend. Scu. Norm. di Pisa*, 333–348.

[107] Hurmuzlu Y (1998), An energy based coefficient of restitution for

planar impacts of slender bars with massive external surfaces, *ASME J. Appl. Mech.* **65**(4), 952–962.

[108] Hurmuzlu Y and Ceanga V (2000), Impulse correlation ratio in solving multiple impact problems, *Impacts in Mechanical Systems: Analysis and Modeling*, Springer Verlag, Berlin, Lecture Notes in Physics, B Brogliato (ed), 235–273.

[109] Paoli L and Schatzman M (1999), Dynamics of an impacting bar: A model and numerics, *ECCM'99, European Congress on Computational Mechanics*, W Wunderlich (ed), CD–Rom.

[110] Cottle RW, Pang JS, and Stone RE (1992), *The Linear Complementarity Problem*, Academic Press, Boston.

[111] Ferris MC and Pang JS (1997), Engineering and economic applications of complementarity problems, *SIAM Rev.* **39**(4), 669–713.

[112] Leine RI, Glocker Ch, and van Campen DH (2001), Nonlinear dynamics and modeling of some wooden toys with impact and friction, submitted to J. Vib. Control, in press.

[113] Jourdan F, Alart P, and Jean M (1998), A Gauss Seidellike algorithm to solve frictional contact problems, *Comput. Methods Appl. Mech. Eng.* **155**, 31–47.

[114] Raous M, Chabrand P, and Lebon F (1988), Numerical methods for frictional contact problems and applications, *J. Theo. Appl. Mech.* **7**, 111–128.

[115] Raous M (1999), Quasistatic Signorini problem with Coulomb friction and coupling to adhesion, *New Developments in Contact Problems*, CISM 384, P Wriggers and P Panaglotiopoulos (eds) Springer Wien New York.

[116] Vola D, Pratt E, Jean M, and Raous M (1998), Consistent time discretization for a dynamical frictional contact problem and complementarity techniques, Revue Européenne des Eléments Finis **7**(1–3).

[117] Lozano, R, Brogliato B, Egeland O, and Maschke B (2000), *Dissipative Systems Analysis and Control: Theory and Applications*, Springer Verlag, CCE Series, London.

[118] Schiehlen W (1999), Unilateral contacts in machine dynamics, *Proc of IUTAM Symp on Unilateral Multibody Contacts*, F Pfeiffer and Ch Glocker (eds), Kluwer, 287–298.

[119] Rückgauer A and Schiehlen W (1997), Simulation of modular mechatronic systems with application to vehicle dynamics, *Acta Mech.* **175**, 183–195.

[120] Eberhard, P, (1999) Collision detection and contact approaches for a hybrid multibody system/finite element simulation, *Proc of IUTAM Symp on Unilateral Multibody Contacts*, F Pfeiffer and Ch Glocker (eds), Kluwer, 193–202.

[121] Ravn P (1998), A continuous analysis method for planar multibody systems with joint clearance, *Multibody Syst. Dyn.* **2**, 1–24.

[122] Bliman PA and Sorine M (1996), Dry friction models for automatic control, *Proc of Euromech Colloquium 351: Systems with Coulomb Friction*, Vadstena (Sweden), August 5–7.

[123] Sorine M and Szymanski J (2000), A new dynamic multi dof tiremodel, *IFAC Symp on Transportation Systems*, Braunschweig, Germany, June 13–15.

[124] Paoli L and Schatzman M (2000), Non convex aspects of dynamics with impacts, *From Convexity to Non Convexity*, R Gilbert and P Pardalos (eds), Kluwer Academic Publ, Boston.

[125] Dupont PE and Yamajako SP (1997), Stability of rigid body dynamics with sliding frictional contacts, *IEEE Trans. Rob. Autom.* **13**(2), 230–236.

[126] Paoli L and Schatzman M (2000), Ill-posedness in vibro-impact and its numerical consequences, *European Congress on Computational Methods in Applied Sciences and Engineering, ECCOMAS*, Barcelona, 11–14 Sept.

[127] Baraff D (1993), Non-penetrating rigid body simulation, *Eurographics'93 State of the Art Reports*, Barcelona, Sept 6–10.

[128] Hori N, Morl T, and Nikiforuk PN (1994), Discrete-time models of continuous-time systems, Control and Dynamic Systems, CT Leondes (ed), **66**, 1–45.

[129] Atkinson KE (1978), *An Introduction to Numerical Analysis*, John Wiley & Sons.

[130] Stoer J and Burlisch R (1979), *Introduction to Numerical Analysis*, Springer Verlag.

[131] Baumgarte J (1972), Stabilization of constraints and integrals of motion in dynamical systems, *Comput. Methods Appl. Mech. Eng,* **1**, 1–16.

[132] Bayo E, Garcia de Jalon J, and Serna MA (1988), A modified Lagrangian formulation for the dynamic analysis of constrained mechanical systems, *Comput. Methods Appl. Mech. Eng.* **71**, 183–195.

[133] Park KC and Chiou JC (1988), Stabilization of computational procedures for constrained dynamical systems, *Journal of Guidance* **11**(4), 365–370.

[134] Sciacovelli D (1985), Flexible spacecraft: Computer oriented control analysis, *Systems and Control Encyclopedia*, Pergamon Press, 1649–1665.

[135] Eich E (1993), Convergence results for a coordinate projection method applied to mechanical systems with algebraic constraints, *SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal.* **30**(5), 1467–1482.

[136] Campbell SL (1990), Descriptor systems in the 90's, *Proc of 29th Conf on Decision and Control*, December, Honolulu, Hawaii, 442–447.

[137] Petzold L (1982), A description of DASSL: A differential/algebraic system solver, *Proc of 10th IMACS World Congress*, August 8–13, Montreal.

[138] Führer C and Leimkuhler B (1989), Formulation and numerical solution of the equations of constrained mechanical motion, DVFLR Forschungsbericht 89-08, Institut für Dynamic der Flugsysteme, Oberpfaffenhofen, Germany.

[139] Baumgarte J (1983), A new method of stabilization for holonomic constraints, *ASME J. Appl. Mech.* **50**, 869–870.

[140] Chang CO and Nikravesh PE (1985), An adaptive constraint violation stabilization method for dynamic analysis of mechanical systems, *Journal of Mechanisms, Transmissions, and Automation in Design* **107**, 488–492.

[141] Asher UM, Chin H, and Reich S (1994), Stabilization of DAEs and invariant manifolds, *Numer. Math.* **67**, 131–149.

[142] Lin ST and Hong MH (1998), Stabilization method for numerical integration of multibody mechanical systems, *ASME J. Mech. Des.* **120**, 565–572.

[143] McClamroch NH and Wang D (1990), Linear feedback control of position and contact force of a nonlinear constrained mechanism, *J. Dyn. Syst., Meas., Control* **122**, 640–645.

[144] Veldman AEP (1990), Missing boundary conditions? Discretise first, substitute next, combine later, *SIAM Journal on Stat. Comp.* **11**, 82–91.

[145] Hart CW and Harlow FH (1967), A general corrective procedure for the numerical solution of initial-value problems, *J. Comput. Phys.* **2**, 114–119.

[146] Gustafsson K, Lundh M, and Oderlind S (1988), A PI step-size control for the numerical solution of ordinary differential equations, *BIT* **28**, 270–287.

[147] Gustafsson K (1990), Using control theory to improve step-size selection in numerical integration of ODE, *11th IFAC* (preprint), 139–144.

[148] von Herzen B, Barr AH, and Zatz HR (1990), Geometric collisions for time-dependent parametric surfaces, *SIGGRAPH Conf Proc*, ACM Press 39–48.

[149] ten Dam AA and Willems JC (1997), A system theoretical framework to study unilaterally constrained dynamical systems, *Proc of European Control Conf ECC'97*, June 30-July 4, Brussels, Belgium.

[150] Wang Y (1993), Global analysis and simulation of mechanical systems with time-varying topologies, *ASME J. Mech. Des.* **115**, 817–821.

[151] Lamarque CH and Janin O (1999), Mechanical systems with impacts: comparison of several numerical methods, *Euromech Colloquium 397 Impact in Mechanical Systems*, Grenoble, LAG-CNRS, June-July.

[152] Janin O and Lamarque CH (2001), Comparison of several numerical methods for mechanical systems with impacts, *Int. J. Numer. Methods Eng.,* **51**(9), 1101.

[153] Janin O (2001), Contribution à l'identification, la modèlisation et l'analyse de systèmes dynamiques à non-linéarites irrégulieres, PhD Thesis, Univ of Lyon 1 C Bernard, Juin.

[154] Towne DH and Hadlock CR (1977), One-dimensional collisions and Chebyshev polynomials, *Am. J. Phys.* **45**(3), 255–259.

[155] Nqi FZ (1997), Etude numérique de divers problémes dynamiques avec impact et de leurs propriétés qualitatives, PhD Thesis, Univ Claude Bernard Lyon 1, June.

[156] Bainov DD and Simeonov PS (1989), *Systems with Impulse Effects; Stability, Theory and Applications*, Wiley.

[157] Kunze M and Neumann J (1997), Linear complementarity problems and the simulation of the motion of rigid body systems subject to Coulomb friction, *Z. Angew. Math. Mech.* **77**(11), 833–838.

[158] Schmaedeke WW (1965), Optimal control theory for nonlinear vector differential equations containing measures, *SIAM J. Control* **3**(2), 231–280.

[159] Camlibel MK, Heemels WPMH, and Schumacher JM (2000), Dynamical analysis of linear passive networks with ideal diodes. Part II: Consistency of a time-stepping method, Tech Report 00 I/03, Eindhoven Univ of Tech, Dept of Electrical Eng, Measurement, and Control Systems, Eindhoven, The Netherlands.

[160] Leenaerts DMW (1999), On linear dynamic complementarity sys-

tems, *IEEE Trans. Circuits Syst., I: Fundam. Theory Appl.* **46**(8), 1022–1026.

[161] Petzold L (1982), Differential/Algebraic equations are not ODEs, *SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput.* **3**(3), 367–384.

[162] Wu SC, Yang SM, and Haug EJ (1986), Dynamics of mechanical systems with Coulomb friction, stiction, impact and constraint addition-deletion; II-planar systems, *Mech. Mach. Theory* **21**(5), 407–416.

[163] Wang D, Conti C, and Beale D (1999), Interference impact analysis of multibody systems, *ASME J. Mech. Des.* **121**, 128–135.

[164] Tzitzouris JA and Pang J-S (2001), A time-stepping complementarity approach for frictionless systems of rigid bodies, (Draft Tech Report 600, John Hopkins Univ, Math Sci Dept), SIAM J. Optim., in press.

[165] Mirtich BV (1997), Impulse-based dynamic simulation of rigid body systems, PhD Thesis, Univ of California at Berkeley.

[166] Hubbard PM (1996), Approximating polyhedra with spheres for time-critical collision detection, *ACM Trans. Graphics* **15**(3), 181–210.

[167] Ponamgi MK, Manosha D, and Lin MC (1995), Incremental algorithm for collision detection between solid models, *Proc 3rd ACM Symp on Solid Modeling and Applications*, ACM Press, NY, 293–304.

[168] Conti C, Corron P, and Michotte P (1992), A computer-aided kinematic analysis system for mechanism design and computer simulation, *Mech. Mach. Theory* **27**(5), 563–574.

[169] Goyal S, Pinson EN, and Sinden FW (1994), Simulation of dynamics of interacting rigid bodies including friction I: General problem and contact model, *Eng. Comput.* **10**, 162–174.

[170] Goyal S, Pinson EN, and Sinden FW (1994), Simulation of dynamics of interactiong rigid bodies including friction II: Software system design and implementation, *Eng. Comput.* **10**, 175–195.

[171] Han I, Gilmore BJ, and Ogot MM (1993), The incorporation of arc boundaries and stick/slip friction in a rule-based simulation algorithm for dynamic mechanical systems with changing topologies, *ASME J. Mech. Des.* **115**, 423–434.

[172] Wang D, Conti C, Dehombreux P, and Verlinden O (1997), A computer-aided simulation approach for mechanisms with time-varying topology, *Comput. Struct.* **64**(14), 519–530.

[173] Filip D, Magedson R, and Marhot R (1986), Surface algorithms using bound derivatives, *Computer Aided Geometric Design* **3**, 255–311.

[174] Lin MC and Canny JF (1991), A fast algorithm for incremental distance calculation, *7th IEEE Conf on Robotics and Automaton*, Sacramento, 1008–1014.

[175] Cohen J, Lin M, Manocah D, and Ponamgi K (1995), I-Collide—An interactive and exact collision detection system for large scale environments, *ACM Int 3D Graphics Conf*, Monterey, 112–120.

[176] Baraff D (1990), Curved surfaces and coherence for non-penetrating rigid body simulation, *Computer Graphics (Proc SIGGRAPH)* **24**, 19–28.

[177] Allgower E and Georg K (1990), *Numerical Continuation Methods: An Introduction*, Springer Verlag, Berlin.

[178] Chabrand P, Dubois F, and Raous M (1998), Various numerical methods for solving unilateral contact problems with friction, *Math. Comput. Modell.* **28**(4-8), 97–108.

[179] Baraff D (1993), Issues in computing contact forces for non-penetrating rigid bodies, *Algorithmica* **10**, 292–352.

[180] Mangasarian OL (1976), Equivalence of the complementarity problem to a system of nonlinear equations, *SIAM (Soc. Ind. Appl. Math.) J. Appl. Math.* **31**(1), 89–92.

[181] Leenaerts DMW and van Bokhoven WMG (1998), *Piecewise Linear Modeling and Analysis*, Kluwer Academic Publ.

[182] Wösle M and Pfeiffer F (1999), Formulations and solution methods of the dynamical spatial contact problem, *Proc of IUTAM Symp on Unilateral Multibody Contacts*, F Pfeiffer and Ch Glocker (eds), Kluwer, 45–56.

[183] Kojima M, Megiddo N, Noma T, and Yoshise A (1991), *A unified Approach to Interior Point Algorithms for Linear Complementarity Problems*, Lecture Notes in Computer Science 538, Springer.

[184] Chabrand P, Dubois F, and Raous M (1995), Programmation mathématique pour le contact avec frottement et comparison avec d'autres méthodes, Deuxième Colloque National en Calcul des Structures, 16–20 mai, Giens (France), Hermes.

[185] Baraff D (1994), Fast contact force computation for nonpentrating rigid bodies, *Proc SIG-GRAPH'94*, July 24–29 Orlando, Computer Graphics Proc, 23–34.

[186] Christensen PW, Klarbring A, Pang JS, and Strömberg N (1998), Formulation and comparison of algorithms for frictional contact problems, *Int. J. Numer. Methods Eng.* **42**, 145–173.

[187] Pang JS and Qi LQ (1993), Nonsmooth equations: Motivations and algorithms, *SIAM J. Optim.*, **3**, 443–465.

[188] Yamashita N (1998), Properties of restricted NCP functions for nonlinear complementarity problems, *J. Optim. Theory Appl.* **98**(3), 701–717.

[189] Facchine F and Kanzow C (1997), On unconstrained and constrained stationary points of the implicit Lagrangian, *J. Optim. Theory Appl.* **92**(1), 99–115.

[190] Press WH, Teukolsky SA, Vetterling WT, and Flannery BP (1992), *Numerical Recipies in Fortran*, 2nd Edition, Cambridge Univ Press, NY.

[191] Klisch T (1999), Contact mechanics in multibody systems, *Mech. Mach. Theory* **34**, 665–675.

[192] Simpack (analysis and design of general mechanical systems), http://www.simpack.de

[193] Haug EJ, Wu SC, and Yang SM (1986), Dynamics of mechanical systems with Coulomb friction, stiction, impact and constraint addition-deletion; I-theory, *Mech. Mach. Theory* **21**(5), 401–406.

[194] Wu SC, Yang SM, Haug EJ (1986), Dynamics of mechanical systems with Coulomb friction, stiction, impact and constraint addition-deletion; III-spatial systems, *Mech. Mach. Theory* **21**(5), 417–425.

[195] Han I and Gilmore BJ (1993), Multi-body impact motion with friction—Analysis, simulation, and experimental validation, *ASME J. Mech. Des.* **115**, 412–422.

[196] Pandy MG and Bermé N (1998), A numerical method for simulating the dynamics of human walking, *J. Biomech.* **21**(12), 1043–1051.

[197] Joli P, Pascal M, and Gibert JR (1993), Systéme avec changement de topolgie dû à des liaisons de contact, *Proc of 11th Congress Français de Mécanique* Vol 5, pp. 429–432.

[198] Escalona JL, Mayo J, and Dominguez J (1999), A new numerical method for the dynamic analysis of impact loads in flexible beams, *Mech. Mach. Theory* **34**, 765–780.

[199] Ferretti G, Maffezzoni C, Magnani G, and Rocco P (1996), Simulating discontinuous phenomena affecting robot motion, *J. Intell. & Robotic Syst.* **15**, 53–65.

[200] Génot F and Brogliato B (1999), New results on Painlevé paradoxes, *Eur. J. Mech. A/Solids* **18**, 653–677.

[201] Génot F (1998), Contribution à la modélisation et à la commande des systèmes mécaniques de corps rigides avec contraintes unilatérales, PhD Thesis, Inst Nat Polytech de Grenoble, Jan.

[202] Lötstedt P (1981), Coulomb friction in two-dimensional rigid body systems, *Z. Angew. Math. Mech.* **61**, 605–615.

[203] Moreau JJ (1968), La notion de sur-potentiel et les liaisons unilatérales en élastostatique, *CRAS Paris* **271A**, 954–957.

[204] Moreau JJ (1974), On unilateral constraints, friction and plasticity, *New Variational Techniques in Mathematical Physics*, CIME II ciclo 1973, G Capriz and G Stampacchia (eds), Edizioni Cremonese, 175–322.

[205] Moreau JJ (1994), Numerical experiments in granular dynamics: Vibration-induced size segregation, *Proc of 2nd Contact Mechanics Int Symp*, Carry-le-Rouet, France, M Raous, M Jean, and JJ Moreau (eds), Plenum Press, NY.

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

[207] Chertier O and Chabrand P (1998), Etude du frottement pour des problèmes de contact en grandes dèformations, *Revue Européenne des Elements Finis* **7**(1-2-3), 163–176.

[208] Pierrot F, Jean M, and Dauchez P (1994), Nonsmooth mechanics approach for robot simulations, *IFAC SYROCO*, Sept 19–21, Capri, Italy, preprints, 577–582.

[209] Jean M and Moreau JJ (1987), Dynamics in the presence of unilateral contacts and dry friction: A numerical approach, in *Unilateral Problems in Structural Analysis 2*, G Del Piero and F Maceri, (ed), CISM courses and lectures 304, Springer Verlag, 151–196.

[210] Galka A and Telega JJ (1998), Dynamic contact problems for systems with finite degrees of freedom: The case of piecewise regular constraints, *Math. Comput. Modell.* **28**(4-8), 479–495.

[211] Kim J and Kwak B (1996), Dynamic analysis of two-dimensional frictional contact by linear complementarity problem formulation, *Int. J. Solids Struct.* **33**(30), 4605–4624.

[212] Johansson L (1999), A linear complementarity algorithm for rigid body impact with friction, *Eur. J. Mech. A/Solids* **18**(4), 703–717.

[213] Johansson L and Klarbring A (2000), Study of frictional impact using a nonsmooth equations solver, *ASME J. Appl. Mech.* **67**, 267–273.

[214] Alart P and Curnier A (1991), A mixed formulation for frictional contact problems prone to Newton like solution methods, *Comput. Methods Appl. Mech. Eng.* **92**, 353–375.

[215] Pang JS (1990), Newton's method for B-differentiable equations, *Math. Op. Res.* **15**, 311–341.

[216] Paoli L and Schatzman M (1993), Schéma numérique pour un modéle

de vibrations avec contraintes unilatérales et perte d'énergie aux impacts, en dimension finie, *C. R. Acad. Sci., Ser. I: Math.* **317**, 211–215.

[217] Paoli L and Schatzman M (1998), Theoretical and numerical study for vibrations with unilateral constraints: Case of a non-convex set of constraints, *Proc 4th World Congress on Computational Mechanics, •New Trends, and Applications*, SR Idelsohn, E Onate, and EN Dvorkin (eds), CD-Rom.

[218] Hubbard JH and West BH (1991), *Differential Equations: A Dynamical Systems Approach, Ordinary Differential Equations*, Springer Verlag TAM 5, NY.

[219] Paoli L and Schatzman M (1995), Dynamical impact problem with loss of energy: Numerical study for large times. Application to a model of tight joints, *Proc of Contact Mechanics* 95, MH Aliabadi and C Alessandri (eds), Computational Mechanics Publications, Southampton, Boston, 233–240.

[220] Panet M, Paoli L, and Schatzman M (1995), Theoretical and numerical study for a model of vibrations with unilateral constraints, *Contact Mechanics Int Symp* **2**, 19–23 Sept 1994, M Raous, M Jean, JJ Moreau (eds), Plenum Press, NY London, 457–464.

[221] Bokhoven WMG (1981), *Piecewise Linear Modelling and Analysis*, Kluwer, Deventer NL.

[222] Sandberg IW (1970), Theorems on the computation of the transient response of nonlinear networks containing transistors and diodes, *Bell Syst. Tech. J.* **49**, 1739–1776.

[223] Massarini A, Reggiani U, and Kazimierczuk K (1997), Analysis of networks with ideal switches by state equations, *IEEE Trans. Circuits Syst., I: Fundam. Theory Appl.* **44**(8), 692–697.

[224] Bedrosian D and Vlach J (1992), Time-domain analysis of networks with internally controlled switches, *IEEE Trans. Circuits Syst., I: Fundam. Theory Appl.* **39**(3), 199–212.

[225] Lin M and Canny F (1992), Efficient collision detection for animation, *Third Eurographics Workshop on Animation and Simulation*, Sept.

[226] An Le Suan (1990), The Painlevé paradoxes and the law of motion of mechanical systems with Coulomb friction, *Prikl. Math. Mekh.* **54**(4), 520-529 (Transl. in *J. Appl. Math. Mech.* **54**(4), 430–438.

[227] Stewart DE (1999), Painlevé's problem and its implication for rigid-body dynamics, *Euromech Colloquium 397 Impact in Mechanical Systems*, Grenoble, 30 June-02 July.

[228] Wood LA and Byrne KP (1981), Analysis of a random repeated impact process, *J. Sound Vib.* **82**, 329–345.

[229] Holmes PJ (1982), The dynamics of repeated impacts with a sinusoidally vibrating table, *J. Sound Vib.* **84**, 173–189.

[230] Brach RM and Brach RM (1987), A review of impact models for vehicle collision, SAE Paper No. 870048.

[231] Pfeiffer F and Glocker Ch (eds) (1999), Unilateral multibody contacts, *Proc of IUTAM Symp*, Munich, Aug 1998, Kluwer, Dordrecht.

[232] Son W, Trinkle JC, and Amato NM (2001), Hybrid dynamic simulation of rigid-body contact with Coulomb friction, *Proc of IEEE Int Conf on Robotics and Automation*, May 21–26, COEX Seoul, Korea, CD-Rom.

[233] Moreau JJ and Panagiotopoulos PD (eds) (1988), *Nonsmooth Mechanics and Applications*, CISM Courses and Lectures no 302, Int Centre for Mechanical Sciences, Springer-Verlag.

Bernard Brogliato *graduated from the Ecole Normale Supèrieure de Cachan in 1987, and got his PhD degree from the Institut National Polytechnique de Grenoble in 1991. He was a CNRS investigator from September 1991 to September 2001 and has been at INRIA Rhone-Alpes since September 2001. His main research interests are in dissipative systems and in nonsmooth dynamical systems analysis and control.*

AA (Tonny) ten Dam *obtained his masters degree in Mathematics (specializing in Systems and Control) in 1986 and his doctoral degree in Mathematics in 1997, both from the University Groningen. In 1988, he joined the National Aerospace Laboratory NLR, Amsterdam, The Netherlands. He acts as consultant for simulation solutions and is involved in NLR's real-time high-performance simulation activities. His main research interests are representation, simulation, and control issues of constrained dynamical systems, with emphasis on aerospace applications.*

Laetitia Paoli *obtained a civil engineer degree from ENTPE (Ecole Nationale des Travaux Publics de l'Etat, Vaulx-en-Velin, France) in 1989 and got her PhD degree from the University of Lyon 1 in 1993 in Applied Mathematics. She is an Associate Professor in the University Jean Monnet-Saint Etienne, and her main research interests are in the mathematical and numerical analysis of nonsmooth dynamical systems.*

Frank Génot *graduated from ENSIMAG (Ecole Nationale Supèrieure d'Informatique et de Mathématiques Appliquées de Grenoble) in 1993 and received his PhD degree from the INPG (Institut National Polytechnique de Grenoble) in 1998. He spent one year at the Mechanical Engineering Dept, MIT, and then joined the INRIA Rocquencourt, Paris. His main research interests are in modeling and control of non-smooth mechanical systems and structural control.*

Michel Abadie *graduated from the Ecole Nationale Supèrieure de Micro-Mécanique (ENSMM, Besançon, France) in 1993, and obtained his PhD degree from the University of Montpellier in 1998. Since 1995, he has been with Schneider Electric, working on the development of Mechanism Simulation. He is now in charge of the Mechanical Simulation group for the Low Voltage activity.*