# A time-stepping scheme for inelastic collisions Numerical handling of the nonoverlapping constraint

Bertrand Maury

**B. Maury**

# A time-stepping scheme for inelastic collisions

## Numerical handling of the nonoverlapping constraint

Laboratoire de Mathématiques, Université Paris Sud, Bâtiment 425, 91405 ORSAY, France
E-mail: Bertrand.Maury@math.u-psud.fr
Fax: 33 (0)1 69 15 67 18

**Abstract** We propose here a numerical scheme to compute the motion of rigid bodies with a non-elastic impact law. The method is based on a global computation of the reaction forces between bodies. Those forces, whose direction is known since we neglect friction effects, are identified at the discrete level with a scalar which plays the role of a Kuhn-Tucker multiplier associated to a first-order approximation of the non-overlapping constraint, expressed in terms of velocities. Since our original motivation is the handling of the non-overlapping constraint in fluid-particle direct simulations, we paid a special attention to stability and robustness. The scheme is proved to be stable and robust. As regards its asymptotic behaviour, a convergence result is established in the case of a single contact. Some numerical tests are presented to illustrate the properties of the algorithm. Firstly, we investigate its asymptotic behaviour in a situation of non-uniqueness, for a single particle. The two other sets of results show the good behaviour of the scheme for large time steps.

## 1 Introduction

The present work was initially motivated by the necessity to handle particle collisions in the direct simulation of fluid-rigid particle mixtures. By direct simulation we mean that each particle is treated individually, and the Navier-Stokes (or Stokes) equations are solved in the moving domain occupied by the fluid. Many methods have been proposed recently to compute such flows. Some are based on an embedding of the solid phase in a global domain which is covered by a cartesian

mesh (see Glowinski [9, Chapter 8] or Bertoluzza et al. [2]); the other class of methods consists in using a conforming mesh of the fluid domain (see Hu [12], Johnson et al. [13], Maury [16]). All these approaches face the problem of body overlapping, which makes it difficult to guarantee the robustness of the computation. Different strategies have been used. In Hu [12], the mesh is refined in the neighborhood of the interparticle gap, so that lubrication forces can be approximated with accuracy and consequently can be expected to prevent overlapping. Indeed, as the interparticle force between two smooth bodies separated by a viscous fluid behaves like $-\dot{\varepsilon}/\varepsilon$ (see Kim [14]), where $\varepsilon$ is the distance, there cannot be any contact in finite time for reasonably regular external forces. Another commonly used strategy consists in adding short range repulsive forces between particles, which tend to prevent particles to overlap (see Glowinski et al. [10]). Those methods have proved to behave quite satisfactorily in many situations, but they necessitate a fine tuning of some numerical parameters, and the actual minimal distance between bodies cannot be controlled *a priori*. In [16], we introduced a heuristic method to control this minimal distance, by running at each time step a minimization procedure on a global functional of the particle positions, such that the minimal distance corresponding to the resulting configuration is greater than a prescribed "safety distance" $\varepsilon > 0$. This method, although robust in practice, even in the case of multiple contacts, is not consistent from the energy point of view, and its long-term effects on the computed flow are difficult to estimate. The approach we proposed in [15] is based on a first-order approximation of the lubrication forces exerted by the fluid in the interparticle gap. It is more respectful of the underlying physics, but the singular character of those forces (proportional to the reciprocal of the distance) introduces new constraints on the time step, and we must add that this approach, despite its very respect of local phenomena, does not seem to improve the overall modeling of fluid-particle mixtures significantly. It made us look for more basic tools, yet respectful of some basic physical properties like momentum balance and decreasing character of energy.

In the present work, we propose a new algorithm to handle the nonoverlapping constraint, which fits into the general class of Contact Dynamics methods, introduced by Moreau [19]. The action of the surrounding fluid shall be simply described by a given external force field, so that from now on we consider "dry" systems of rigid bodies. Given the nature of the interaction between particles moving in a viscous fluid, which is dissipative, we chose to consider here purely inelastic collisions. The evolution problem on which our approach is based fits into the general framework of differential measure inclusions (see Schatzman [20], or Moreau [19]). From the theoretical point of view, the critical point is uniqueness, which necessitates strong regularity assumptions on the data: analyticity is required. Uniqueness is lost as soon as the external force is no longer analytic (see in Schatzman [20] or Ballard [1] counter–examples to uniqueness with a $C^\infty$ force). This question of uniqueness in regard with the numerical algorithm will be addressed at the end of the paper. We refer to Stewart [23] for a general presentation of mathematical issues raised by those models, and a wide presentation of numerical methods. The most general result concerning convergence analysis of a numerical scheme for granular flows can be found in [22], where Coulomb friction is taken into account.

The situation we consider here is less general than the aforementioned one, but the scheme we propose presents the following features:

1. Collisions are not treated as events[1]. All contacts which are likely to occur during a time step are indeed handled globally, without any prior prediction of actual violations of the non-overlapping constraint, and reactions are computed as a field of Kuhn-Tucker multipliers associated to first-order approximations of those constraints, expressed in terms of velocities.
2. As a consequence, this method can be proved to be unconditionally stable. For any choice of the time step, the kinetic energy of the approximate solution dissipates (when external forces are reduced to 0). Since there is no parameter to tune up (except for the stopping criterium associated to the Uzawa algorithm), this algorithm can be applied to very different situations with no modification.
3. Provided the quadratic minimization step is solved exactly, the scheme produces feasible configurations only (in the case of spherical bodies).
4. Non-smooth forcing terms can be taken into account, as soon as their mean values on subintervals can be computed. This makes it possible to handle multiple solutions numerically (see section 6.1).

The paper is organized as follows. In section 2 we present the rigid-body problem, and we introduce the natural functional spaces for positions, velocities and reaction forces. We then present the time-stepping scheme (section 3). In the next section we prove some general properties of the scheme, all of which are valid for any time step. Section 5 is devoted to the asymptotic behaviour of the scheme: as the time step $h$ goes to 0, a subsequence of the computed solutions is shown to converge in some sense to a solution of the original problem, in the case of a single contact. Some numerical tests are finally presented, to illustrate the behaviour of the algorithm in different contexts: non-uniqueness, sticky particle model, and many-body motion (section 6).

## 2 Model problem

### 2.1 Evolution model

We consider the mechanical system of $N$ rigid spheres (or discs, or line segments) in $\mathbb{R}^d$ (with $d = 3$, 2, or 1) of radii $(r_i)_{1\le i\le N}$ and masses $(m_i)_{1\le i\le N}$. We introduce the configuration space,

$$Q = \{\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N) \in \mathbb{R}^{dN}\}$$

and the associated feasible set

$$Q_0 = \{\mathbf{q} \in Q, D_{ij}(\mathbf{q}) \ge 0 \quad \forall i, j\ , \quad 1 \le i < j \le N\}$$

where $D_{ij}(\mathbf{q}) = |\mathbf{q}_i - \mathbf{q}_j| - (r_i + r_j)$ is the signed distance between spheres $i$ and $j$ (see Figure 2.1). Note that $D_{ij}$ is negative as soon as spheres $i$ and $j$ overlap.

As we do not consider rotations here, our configuration space is endowed with a natural Euclidean structure, so that the tangent space at any $\mathbf{q} \in Q$ can be identified with $Q$ itself. We shall nevertheless denote by $\mathrm{T}_Q = \mathbb{R}^{dN}$ this tangent space in

[1] Borrowing the terminology of numerical methods to handle discontinuity (in space) of solutions to hyperbolic PDE's, one may say that collisions (*i.e.* discontinuities in time for the velocities) are captured, whereas in most other methods, they are tracked.

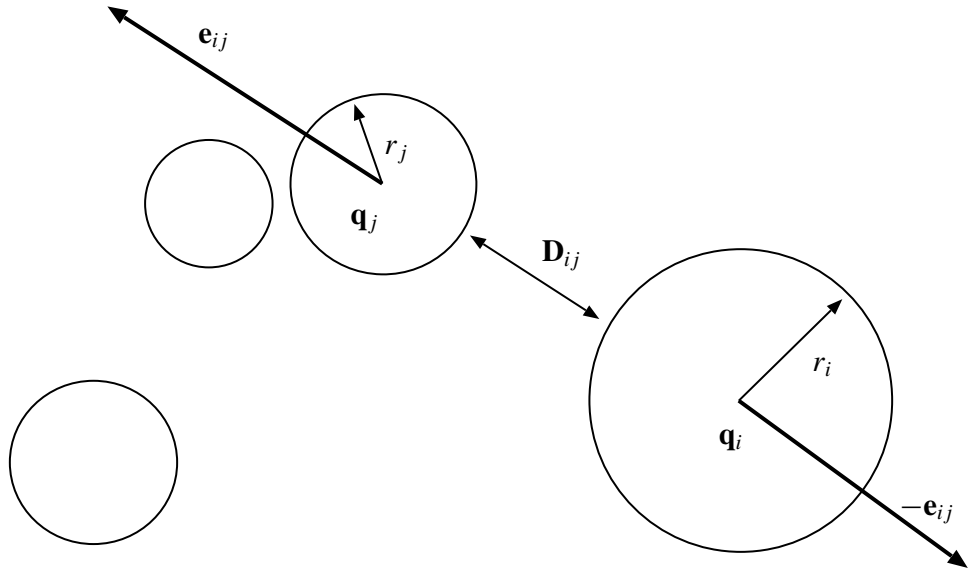


**Fig. 1** Notations

order to distinguish velocities from positions. We define $\mathbf{G}_{ij} = \nabla D_{ij}$ as the gradient of the distance between the two spheres $i$ and $j$:

$$\mathbf{G}_{ij} = (\cdots, 0, -\mathbf{e}_{ij}, 0, \ldots, 0, \mathbf{e}_{ij}, 0, \cdots) \ , \quad \mathbf{e}_{ij} = \frac{\mathbf{q}_j - \mathbf{q}_i}{\left|\mathbf{q}_j - \mathbf{q}_i\right|}, \tag{1}$$

and $\mathcal{C}_\mathbf{q}$ as the set of feasible directions at $\mathbf{q} \in Q_0$,

$$\mathcal{C}_\mathbf{q} = \left\{\mathbf{v} \in \mathrm{T}_Q \ , \quad \mathbf{G}_{ij} \cdot \mathbf{v} \geq 0 \quad \text{as soon as} \ \ D_{ij}(\mathbf{q}) = 0\right\},$$

and the outward normal cone to $Q_0$ at $\mathbf{q}$ as its polar cone:

$$\begin{aligned}\mathcal{N}_\mathbf{q} = \ & \mathcal{C}_\mathbf{q}^\circ = \{\mathbf{h} \in \mathrm{T}_Q \ , \quad \mathbf{h} \cdot \mathbf{v} \leq 0 \quad \forall \mathbf{v} \in \mathcal{C}_\mathbf{q}\} \\ = \ & \Big\{ - \sum_{i<j} \mu_{ij} \mathbf{G}_{ij}(\mathbf{q}), \\ & \qquad \mu_{ij} = 0 \text{ if } D_{ij}(\mathbf{q}) > 0, \ \mu_{ij} \in \mathbb{R}^+ \text{ if } D_{ij}(\mathbf{q}) = 0\Big\}.\end{aligned}$$

Let $\mathbf{u} = (\dot{\mathbf{q}}_1, \dot{\mathbf{q}}_2, \ldots, \dot{\mathbf{q}}_N) \in \mathrm{T}_Q$ denote the generalized velocity vector. We finally denote by $\mathbf{f}$ the force field acting on the spheres, and by M the mass matrix. The problem we are interested in may be phrased formally:

$$\mathrm{M}\ddot{\mathbf{q}} + \mathcal{N}_\mathbf{q} \ni \mathbf{f} \tag{2}$$

$$\mathbf{u}^+ = P_{\mathcal{C}_\mathbf{q}} \mathbf{u}^-, \tag{3}$$

where $\mathbf{u}^-$ (resp. $\mathbf{u}^+$) is the left (resp. right) limit of the velocity vector at time $t$, and $P_{\mathcal{C}_\mathbf{q}}$ is the Euclidean projection onto the closed convex cone $\mathcal{C}_\mathbf{q}$.

This model will be given a proper mathematical framework in the next section. As regards its mechanical sense, equation (2), which can be read formally

$$\exists \mathbf{h} \in -\mathcal{N}_\mathbf{q} \text{ such that } \mathrm{M}\ddot{\mathbf{q}} = \mathbf{f} + \mathbf{h}, \tag{4}$$

expresses the fact that overlapping is prevented by repulsive forces acting on each sphere along the normal vector at the contact point. When there is no contact, $\mathcal{N}_{\mathbf{q}}$ reduces down to $\{0\}$, so that (2) reads as the ordinary differential equation $\mathrm{M}\ddot{\mathbf{q}} = \mathbf{f}$. Equation (3) provides the collision model. This is a particular case of a more general model (which includes possibly an elastic behaviour)

$$\mathbf{u}^+ = \mathbf{u}^- - (1+e)P_{\mathcal{N}_{\mathbf{q}}}\mathbf{u}^-, \tag{5}$$

where $e$ is a restitution coefficient, set to 0 in the case of a non-elastic impact law. Indeed, as $\mathcal{N}_{\mathbf{q}}$ and $\mathcal{C}_{\mathbf{q}}$ are mutually polar, it holds $I - P_{\mathcal{N}_{\mathbf{q}}} = P_{\mathcal{C}_{\mathbf{q}}}$, where $I$ is the identity. (see Moreau [18]). Note that, in the non-elastic model we consider, some energy is lost during each collision. This kinetic energy is actually transformed into heat within the bodies. The reader may refer to Frémond [8], where this increase in temperature is explicitly integrated into a general thermomechanical model for collisions between rigid bodies.

*Remark 1* For $1D$ problems (like the one which is presented in Section 6.2), it is in fact more natural to define $Q_0$ as a connected component of the set of all configurations $\mathbf{q}$ with no overlapping. Indeed, the evolution problem does not allow $\mathbf{q}$ to pass from a connected component to another (the bodies cannot leap across each other). With this convention, the feasible set

$$Q_0 = \{\mathbf{q} = (q_1, \ldots, q_N) \in \mathbb{R}^N \ , \ \ q_{i+1} - q_i \geq r_{i+1} + r_i\}$$

is closed and convex, and in this particular situation, $\mathcal{N}_{\mathbf{q}}$ identifies with the subdifferential of the indicatrix of $Q_0$:

$$\mathcal{N}_{\mathbf{q}} = \partial I_{Q_0}(\mathbf{q}) \text{ with } I_{Q_0}(\mathbf{q}) = \left| \begin{array}{rl} 0 & \text{if } \mathbf{q} \in Q_0 \\ +\infty & \text{if } \mathbf{q} \notin Q_0 \end{array} \right.$$

and hence $\mathbf{q} \longmapsto \mathcal{N}_{\mathbf{q}}$ is a maximal monotone operator. In general $Q_0$ is not convex, but it is not far from being convex: it can be proved easily to be uniformly $\eta$-prox-regular in the sense defined in Colombo [7], as soon as $1/\eta < 4\min(r_i)$.

## 2.2 Functional framework

We consider the time interval $I = (0, T)$ and we introduce the following functional spaces:

$W^{1,1}$ = set of all $\mathbb{R}^{dN}$-valued functions which are absolutely continuous over the interval $I$ ;

BV = set of all $dN$ vector-valued functions having bounded variation over $I$: BV is the set of all functions $t \mapsto \mathbf{u}(t) \in \mathrm{T}_Q$, such that each component $u$ of $\mathbf{u}$ verifies

$$\sup_{S \in \Lambda} \sum_{n=1}^{N_S} |u(t_n) - u(t_{n-1})| < \infty,$$

where $S = (t_0, t_1, \ldots, t_{N_S})$ runs over the set $\Lambda$ of increasing subdivisions of the time interval $I$ ;

$\mathcal{M}^1 =$ set of $N(N-1)/2$ vector[2]-valued bounded measures on $I$ : it contains all $\boldsymbol{\mu} = (\mu_{ij})_{1\le i<j\le N}$ such that $\mu_{ij}$ is a continuous linear functional over the set $C_0(I)$ of continuous functions over $\overline{I}$, vanishing at 0 and $T$. The set of component-wise positive measures will be denoted by

$$\mathcal{M}^1_+ = \{\boldsymbol{\mu} = (\mu_{ij})_{1\le i<j\le N} \in \mathcal{M}^1,\ \langle \mu_{ij}, \varphi\rangle \ge 0 \quad \forall \varphi \in C_0(I),\ \varphi \ge 0\}.$$

As regards the forcing term, we shall consider functions of $t$ and $\mathbf{q}$, with Carathéodory type regularity conditions (see Coddington et al. [6]) : let the function $\mathbf{f}$ be defined in $Q_0 \times I$, measurable in $t$ for fixed $\mathbf{q}$ and continuous in $\mathbf{q}$ for fixed $t$, such that

$$\exists F \in L^1(I) \text{ s.t. } |\mathbf{f}(\mathbf{q}, t)| \le F(t) \quad \forall (\mathbf{q}, t) \in Q_0 \times I, \tag{6}$$

and $\mathbf{f}$ is uniformly Lipschitz with respect to the space variable in the feasible set:

$$\exists k\ ,\ \ \left|\mathbf{f}(\mathbf{q}', t) - \mathbf{f}(\mathbf{q}, t)\right| \le k\left|\mathbf{q}' - \mathbf{q}\right| \ \ a.e. \text{ in } I \quad \forall \mathbf{q},\ \mathbf{q}' \in Q_0. \tag{7}$$

In order to avoid a special treatment of the initial time, we shall pick an initial position $\mathbf{q}_0$ in the interior of $Q_0$, so that $\mathcal{N}_{\mathbf{q}} = \{0\}$. It follows that the velocity is continuous at $0^+$, and $\mathbf{u}(0)$ is defined without ambiguity.

We may now state the problem: Given a time interval $I = (0, T)$, a force $\mathbf{f}$ (which meets regularity assumptions (6)–(7)), and initial conditions $(\mathbf{q}_0, \mathbf{u}_0) \in Q_0 \times \mathrm{T}_Q$, with $D_{ij}(\mathbf{q}_0) > 0$ for all $i \ne j$,

$$\text{Find } (\mathbf{q}, \mathbf{u}, \boldsymbol{\mu}) \in W^{1,1} \times \mathrm{BV} \times \mathcal{M}^1_+ \text{ such that}$$

$$\mathbf{u}(0) = \mathbf{u}_0, \tag{8}$$

$$\mathbf{q}(t) = \mathbf{q}_0 + \int_0^t \mathbf{u}(s)\,\mathrm{ds}\ \in Q_0 \quad \forall\ t\ \in I \tag{9}$$

$$\mathrm{M}\dot{\mathbf{u}} = \mathbf{f}(\mathbf{q}, t) + \sum_{i<j} \mu_{ij}\mathbf{G}_{ij}(\mathbf{q}(t)), \tag{10}$$

$$\forall\ i < j \quad \mathrm{supp}(\mu_{ij}) \subset \{t,\ D_{ij}(\mathbf{q}(t)) = 0\}, \tag{11}$$

$$\mathbf{u}^+(t) = \mathbf{u}^-(t) - P_{N_{\mathbf{q}}}\mathbf{u}^-(t) \quad \forall t \in I, \tag{12}$$

where $P_{\mathcal{N}_{\mathbf{q}}}$ is the euclidean projection onto the closed convex cone $\mathcal{N}_{\mathbf{q}}$. Equation (10) is to be understood in the sense of distributions (it identifies two 0-th order distributions).

---

[2] $N(N-1)/2$ is the number of possible contacts.

*Remark 2* Note that $\mathbf{f}(\cdot, t)$ is required to be in $W^{1,\infty}(Q_0)$ and not necessarily in $W^{1,\infty}(Q)$. This makes it possible to apply this approach to model interaction forces like

$$\mathbf{f} = -K\nabla V(|q_2 - q_1|),$$

where $V : (0, +\infty) \to \mathbb{R}^+$ is smooth, decreasing and convex, possibly singular at 0 : since the distance between any $\mathbf{q} \in Q_0$ and 0 is kept apart from 0 (it is always greater than twice the smallest radius), such a function is indeed lipschitzian. For instance an electrostatic potential, $V(d) = K/d$, meets the requirements.

*Remark 3* As $\mathbf{q}$ has the $W^{1,1}$ regularity in time, we can consider that $\mathbf{q}$ is continuous. Similarly, any $\mathbf{u} \in \mathrm{BV}$ (or $\mathbf{u}_h$ in the next section), shall be identified with the right-continuous element of the class. With this convention, all pointwise identities or inequalities will be meant in the classical sense (*i.e.* everywhere).

*Remark 4* Obstacles like walls or fixed bodies may be taken into account. For the sake of simplicity, we shall not introduce here extra notations for that purpose, but the set of Kuhn-Tucker multipliers $\boldsymbol{\mu}$ could be extended to include reaction forces against obstacles.

## 3 Numerical scheme

### 3.1 Approximation spaces

Let $h = T/N_h$ be the time step, and let $\mathrm{P}_h^0$ denote the set of all those functions which are piecewise constant over $I$ according to the uniform subdivision $(0, h, 2h, \ldots, (N_h - 1)h, T)$. Similarly, $\mathrm{P}_h^1$ denotes the set of continuous, piecewise affine functions according to the same subdivision. We introduce the following approximation spaces for trajectories, velocities, and reaction forces, respectively:

$$\begin{aligned}
X_h &= \{\mathbf{q}_h = (\mathbf{q}_i) : I \to Q \ , \ \ \mathbf{q}_i \in (\mathrm{P}_h^1)^d \ , \ \ 1 \le i \le N\},\\
V_h &= \{\mathbf{u}_h = (\mathbf{u}_i) : I \to \mathrm{T}_Q \ , \ \ \mathbf{u}_i \in (\mathrm{P}_h^0)^d \ , \ \ 1 \le i \le N\},\\
R_h &= \{\boldsymbol{\mu}_h = (\mu_{ij}) : I \to \mathbb{R}^r \ , \ \ \mu_{ij} \in \mathrm{P}_h^0 \ , \ \ 1 \le i < j \le N\},
\end{aligned}$$

where $r = N(N-1)/2$ is the number of constraints. For any $\mathbf{u}_h \in V_h$ (resp. $\boldsymbol{\mu}_h \in R_h$) we shall denote by $\mathbf{u}_h^n$ (resp. $\boldsymbol{\mu}_h^n$) its constant value in the subinterval $[(n-1)h, nh)$. Similarly, $\mathbf{q}_h^n$ denotes $\mathbf{q}_h(nh)$, for any $\mathbf{q}_h \in X_h$. As for external forces, we shall use the following mean value approximation $\mathbf{f}_h$ in $V_h$ of $\mathbf{f}(\mathbf{q}, \cdot)$ defined for any $\mathbf{q} \in Q$ by its $N_h$ values

$$t \in [(n-1)h, nh) \ \longmapsto \ \mathbf{f}_h^n(\mathbf{q}) = \frac{1}{h}\int_{(n-1)h}^{nh} \mathbf{f}(\mathbf{q}, s)\ \mathrm{ds}.$$

### 3.2 Time-stepping scheme

The sequence of approximated fields $(\mathbf{q}_h, \mathbf{u}_h, \boldsymbol{\mu}_h) \in X_h \times V_h \times R_h$ is built according to the following scheme, to which we shall refer as scheme $(\mathcal{S})$:

1. Initialization

$$(\mathbf{q}_h^0, \mathbf{u}_h^0) = (\mathbf{q}_0, \mathbf{u}_0). \tag{13}$$

2. Compute $\mathbf{u}_h^{n+1}$ as the solution to the constrained minimization problem

$$\min_{\mathbf{u}\in \mathrm{K}_h(\mathbf{q}_h^n)} \frac{1}{2}\left|\mathbf{u} - \mathbf{u}_h^n - h\mathrm{M}^{-1}\mathbf{f}_h^{n+1}(\mathbf{q}_h^n)\right|_M^2 \tag{14}$$

with $|\mathbf{v}|_{\mathrm{M}}^2 = \mathrm{M}\mathbf{v}\cdot\mathbf{v}$, and

$$\mathrm{K}_h(\mathbf{q}_h^n) = \{\mathbf{u} \in \mathrm{T}_Q,\, D_{ij}(\mathbf{q}_h^n) + h\mathbf{G}_{ij}(\mathbf{q}_h^n)\cdot\mathbf{u} \geq 0\}. \tag{15}$$

The approximate reaction field $\boldsymbol{\mu}_h^{n+1} = (\mu_{ij}^{n+1})$ is the dual component of a solution to the associated saddle-point problem

$$\mathcal{L}(\mathbf{u}_h^{n+1}, \boldsymbol{\lambda}) \leq \mathcal{L}(\mathbf{u}_h^{n+1}, \boldsymbol{\mu}_h^{n+1}) \leq \mathcal{L}(\mathbf{v}, \boldsymbol{\mu}_h^{n+1}), \tag{16}$$

for all $\boldsymbol{\lambda} \in \mathbb{R}_+^{N(N-1)/2}$, $\mathbf{v} \in \mathrm{T}_Q$, where $\mathcal{L}$ is the Lagrangian

$$\begin{aligned}\mathcal{L}(\mathbf{v}, \boldsymbol{\lambda}) = \frac{1}{2}\left|\mathbf{v} - \mathbf{u}_h^n - h\mathrm{M}^{-1}\mathbf{f}_h^{n+1}(\mathbf{q}_h^n)\right|_{\mathrm{M}}^2 \\ - \sum_{1\leq i<j\leq N} \lambda_{ij}\left(D_{ij}\left(\mathbf{q}_h^n\right) + h\mathbf{G}_{ij}(\mathbf{q}_h^n)\cdot\mathbf{v}\right).\end{aligned} \tag{17}$$

Note that $\mathbf{u}_h^{n+1}$ and $\boldsymbol{\mu}_h^{n+1} = (\mu_{ij}^{n+1})_{1\leq i<j\leq N}$ are related by

$$\mathrm{M}\mathbf{u}_h^{n+1} = \mathrm{M}\mathbf{u}_h^n + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n) + h\sum_{1\leq i<j\leq N} \mu_{ij}^{n+1}\mathbf{G}_{ij}(\mathbf{q}_h^n). \tag{18}$$

3. Update the positions

$$\mathbf{q}_h^{n+1} = \mathbf{q}_h^n + h\mathbf{u}_h^{n+1}. \tag{19}$$

*Remark 5* The velocity field $\mathbf{u}_h^{n+1}$ is uniquely defined (see (14): it minimizes a strictly convex fonctional over a closed convex set), whereas $\boldsymbol{\mu}_h^{n+1}$ is not. So in a strict sense scheme ($\mathcal{S}$) is an algorithm as far as $\mathbf{q}$ and $\mathbf{u}$ are concerned only.

The following remark explains why the previous scheme can be expected to approximate the original problem.

*Remark 6* Step 2, which is the essential (and time-consuming) part of the scheme, can be written

$$\frac{\mathrm{M}\mathbf{u}_h^{n+1} - \mathrm{M}\mathbf{u}_h^n}{h} + \partial I_{\mathrm{K}_h(\mathbf{q}_h^n)}\left(\mathbf{u}_h^{n+1}\right) \ni \mathbf{f}_h^{n+1}(\mathbf{q}_h^n) \tag{20}$$

where $\partial I_{\mathrm{K}_h(\mathbf{q}_h^n)}$ is the subdifferential of the indicatrix function of $\mathrm{K}_h(\mathbf{q}_h^n)$. Now observe that, for any $\mathbf{q} \in Q_0$, $\mathcal{N}_{\mathbf{q}}$ is the sum of all half lines $-\mathbb{R}_+\mathbf{G}_{ij}$ for indices verifying $D_{ij}(\mathbf{q}) = 0$, and $\partial I_{\mathrm{K}_h(\mathbf{q})}(\mathbf{u})$ can be written as the same sum for indices $i$ and $j$ such that

$$D_{ij}(\mathbf{q}) + h\mathbf{G}_{ij}\cdot\mathbf{u} = 0.$$

As the latter quantity is a Taylor expansion of $D_{ij}(\mathbf{q}+h\mathbf{u})$, the set $\partial I_{\mathrm{K}_h(\mathbf{q}_h^n)}(\mathbf{u}_h^{n+1})$ can be seen as a prediction of $\mathcal{N}_{\mathbf{q}_h^{n+1}}$. Therefore the scheme we propose can be considered, at least formally, as a semi-implicit time discretization of inclusion (2),

$$\mathrm{M}\frac{d\mathbf{u}}{dt} + \mathcal{N}_{\mathbf{q}} \ni \mathbf{f}.$$

Note that the collision law (3) does not appear explicitly in the scheme. It is actually implicitly contained in (20), which selects the velocity corresponding to the collision law we considered here. For this reason, this scheme is intrinsically dedicated to inelastic impact models, and we must say that there is no obvious way to adapt it to other impact laws.

### 3.3 Saddle-point problem

In the present approach we solve the saddle-point problem (16) by a Uzawa algorithm. Note that this choice is not mandatory, since the way the minimization procedure is performed is completely independant from the time-stepping scheme itself. Other algorithms could be implemented to perform this task, and the theoretical results presented in the next sections would remain unchanged.

We introduce, for this section only, the vectors and matrices which are involved in the Uzawa procedure:

$$\mathbf{u} = \mathbf{u}_h^{n+1} \in \mathbb{R}^{dN} \ , \quad \mathbf{F} = \mathrm{M}\mathbf{u}_h^n + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n) \in \mathbb{R}^{dN},$$

$$\boldsymbol{\mu} = (\mu_{ij}^{n+1})_{1\le i<j\le N} \in \mathbb{R}^r \ , \quad \mathbf{D} = (D_{ij}(\mathbf{q}_h^n))_{1\le i<j\le N},$$

$$C \in \mathcal{M}_{r,dN}(\mathbb{R}) \ \text{ such that } \ -C^T\boldsymbol{\mu} = h\sum_{1\le i<j\le N} \mu_{ij}^{n+1}\mathbf{G}_{ij}(\mathbf{q}_h^n),$$

where $r = N(N-1)/2$ is the number of constraints. The problem can be put in the classical saddle-point form

$$\begin{aligned} \text{Minimize } J(\mathbf{v}) &= \frac{1}{2}\mathrm{M}\mathbf{v}\cdot\mathbf{v} - \mathbf{F}\cdot\mathbf{v}, \\ \text{over } K &= \{\mathbf{v}\,,\ C\mathbf{v} \le \mathbf{D}\}. \end{aligned}$$

The Lagrangian of the problem is

$$\mathcal{L}(\mathbf{v},\boldsymbol{\lambda}) = \frac{1}{2}\mathrm{M}\mathbf{v}\cdot\mathbf{v} - \mathbf{F}\cdot\mathbf{v} + \boldsymbol{\lambda}\cdot(C\mathbf{v}-\mathbf{D}),$$

and any saddle-point of $\mathcal{L}$, *i.e.* any couple $(\mathbf{u},\boldsymbol{\mu})$ verifying

$$\sup_{\boldsymbol{\lambda}\in\mathbb{R}^r_+} \mathcal{L}(\mathbf{u},\boldsymbol{\lambda}) = \mathcal{L}(\mathbf{u},\boldsymbol{\mu}) = \inf_{\mathbf{v}\in\mathbb{R}^{dN}} \mathcal{L}(\mathbf{v},\boldsymbol{\mu}),$$

is such that $\mathbf{u}$ minimizes $J$ over $K$.

The Uzawa algorithm (see Ciarlet [5]) consists in approximating a saddle–point $(\mathbf{u},\boldsymbol{\mu})$ of $\mathcal{L}$ by sequences $(\mathbf{u}_k)$ and $(\boldsymbol{\mu}_k)$. A step decomposes into two substeps:

1. Solve the primal problem

$$\mathrm{M}\mathbf{u}_{k+1} = \mathbf{F} - C^T \boldsymbol{\mu}_k.$$

2. Update the Kuhn-Tucker multipliers

$$\boldsymbol{\mu}_{k+1} = \Pi_+ \left( \boldsymbol{\mu}_k + \rho(C\mathbf{u}_{k+1} - \mathbf{D}) \right),$$

where $\Pi_+$ is the orthogonal projection onto $\mathbb{R}^r_+$:

$$\boldsymbol{\mu} \in \mathbb{R}^r \longmapsto \Pi_+(\boldsymbol{\mu}) = \left( \max(\mu_{ij}, 0) \right)_{1 \le i < j \le N}.$$

**Proposition 1** *We suppose that*

$$0 < \rho < \rho_{\max} = \frac{2\alpha}{\|C^T C\|_2},$$

*where $\alpha$ is the smallest eigenvalue of the matrix M. Then the sequence $(\mathbf{u}_k)$ converges to the solution $\mathbf{u}$ to the constrained minimization problem. The sequence $(\boldsymbol{\mu}_k)$ converges toward some $\boldsymbol{\mu}$ such that $(\mathbf{u}, \boldsymbol{\mu})$ is a saddle-point of $\mathcal{L}$.*

*Proof* The convergence of the primal sequence $(\mathbf{u}_k)$ is a classical result (see Ciarlet [5]). As for the Kuhn-Tucker multipliers, the sequence $(\boldsymbol{\mu}_k)$ can be shown to meet the assumptions of Opial's Lemma (see Haraux [11]), and thus it converges (weakly) to some $\boldsymbol{\mu}$ in the dual solution set. As the problem is finite-dimensional, the convergence is strong. □

We conclude this section by two remarks on the actual implementation of this algorithm.

*Remark 7* It is not necessary to assemble the whole matrix $C$. At a given time step, positions and velocities are known, so that most constraints (which correspond to particles far away from each other) can be eliminated *a priori*, in the spirit of what is presented in Sigurgeirsson [21]. For monodisperse situations (all radii are identical), the total number of potentially active constraints (*i.e.* number of rows of $C$) is $\kappa N/2$, with $\kappa = 6$ in two dimensions (each disc is close to 6 other discs, at most), and $\kappa = 12$ in three dimensions.

*Remark 8* A sufficient condition on $\rho$ for the algorithm to converge can be obtained explicitly in standard situations. We give here a rough lower bound for $\rho_{\max}$ in the monodisperse case. In this situation, the maximal number of likely contacts per body is smaller than the aforementioned parameter $\kappa$. The ellipticity constant $\alpha$ is simply $\min(m_i)_{1 \le i \le N}$. Using expression (1) of $\mathbf{G}_{ij}$, we can write $C^T C \in \mathcal{M}_{dN}(\mathbb{R})$ by blocks:

$$C^T C = h^2 \, (C_{ij})_{1 \le i,\, j \le N},$$

where $C_{ij}$ is a $d \times d$ matrix defined by

$$C_{ii} = \sum_{j \in J_i} \mathbf{e}_{ij} \otimes \mathbf{e}_{ij} \ , \quad C_{ij} = -\mathbf{e}_{ij} \otimes \mathbf{e}_{ij} \text{ for } i \neq j,$$

with $J_i$ denoting the set of indices of the spheres which may get into contact with sphere $i$. By Gerschgorin circle theorem, the spectral radius of $C^TC$ (*i.e.* $||C^TC||_2$) is smaller than the maximum among the 1-norms of its rows. Since $\left|\mathbf{e}_{ij}\right| = 1$, it follows that $||C^TC||_2 \leq 2\kappa h^2\sqrt{d}$, where $\kappa$ is the upper bound of $\sharp(J_i)$ (cardinal number of $J_i$). Finally,

$$\rho_{\max} \geq \frac{\min(m_i)}{2\kappa h^2\sqrt{d}},$$

where $\kappa$ is 12 for $3D$ problems, as mentioned in the previous remark. It is to be noted that this lower bound is independent of $N$, the number of bodies.

## 4 Properties of the scheme

This section is devoted to some general properties of the algorithm in the general case (multiple contacts are allowed). Some of those properties will be used in the next section to prove convergence in the case of a single contact.

**Proposition 2** *(Feasibility) If the minimization step is solved exactly, then the scheme* $(\mathcal{S})$ *produces feasible configurations only:*

$$\mathbf{q}_h(t) \in Q_0 \quad \forall h > 0, \forall t.$$

*Proof* As Step 2 of Scheme $(\mathcal{S})$ is supposed to be solved exactly, then for any $h$, $n$, $i \neq j$,

$$\mathbf{u}_h^{n+1} \in \mathrm{K}_h(\mathbf{q}_h^n) \Longrightarrow D_{ij}(\mathbf{q}_h^n) + h\mathbf{G}_{ij}(\mathbf{q}_h^n) \cdot \mathbf{u}_h^{n+1} \geq 0,$$

therefore

$$\begin{aligned} D_{ij}(\mathbf{q}_h^{n+1}) &= D_{ij}(\mathbf{q}_h^n + h\mathbf{u}_h^{n+1}) \\ &\geq D_{ij}(\mathbf{q}_h^n) + h\mathbf{G}_{ij}(\mathbf{q}_h^n) \cdot \mathbf{u}_h^{n+1} \quad \text{(since } D_{ij} \text{ is convex)} \\ &\geq 0. \end{aligned} \tag{21}$$

We established that $\mathbf{q}_h^n \in Q_0$ for any $h, n$. Now for any $t = nh + \theta h$, with $\theta \in [0, 1]$, one has

$$\begin{aligned} D_{ij}(\mathbf{q}_h(t)) &= D_{ij}(\mathbf{q}_h^n + \theta h\mathbf{u}_h^{n+1}) \\ &\geq D_{ij}(\mathbf{q}_h^n) + \theta h\mathbf{G}_{ij}(\mathbf{q}_h^n) \cdot \mathbf{u}_h^{n+1} \quad \text{(since } D_{ij} \text{ is convex)} \\ &= (1-\theta)D_{ij}(\mathbf{q}_h^n) + \theta\left(D_{ij}(\mathbf{q}_h^n) + h\mathbf{G}_{ij}(\mathbf{q}_h^n) \cdot \mathbf{u}_h^{n+1}\right) \qquad (22) \\ &\geq 0. \qquad (23) \end{aligned}$$

Note that this property is no longer true if the bodies are not spherical, because the functions $\mathbf{q} \longmapsto D_{ij}(\mathbf{q})$ are not convex in general. □

**Proposition 3** *(Stability) Let* $(\mathbf{q}_h, \mathbf{u}_h)_{h>0}$ *be a sequence built according to Scheme* $(\mathcal{S})$*, for a given* $\mathbf{f}$*. Assuming again that the minimization procedure (Step 2) is performed exactly, then* $\mathbf{u}_h$ *verifies the following estimate:*

$$|\mathbf{u}_h(t)|_{\mathrm{M}} \leq |\mathbf{u}_0|_{\mathrm{M}} + \frac{1}{\sqrt{\alpha}} \int_0^{t+h} F(s)\, ds, \tag{24}$$

*where* $|\cdot|_{\mathrm{M}}$ *is the euclidean norm associated to the symmetric positive definite matrix* M*:*

$$|\mathbf{u}|_{\mathrm{M}} = \sqrt{\mathrm{M}\mathbf{u} \cdot \mathbf{u}},$$

$F$ *is the function which dominates* $\mathbf{f}$ *(see condition (6)), and* $\alpha$ *is the smallest eigenvalue of* M *(*$\alpha = \min(m_i)$*).*

*Proof* Let us first establish that, for every $n$, $i \neq j$,

$$\mu_{ij}^{n+1} \mathbf{G}_{ij}(\mathbf{q}_h^n) \cdot \mathbf{u}_h^{n+1} \leq 0. \tag{25}$$

It is a direct consequence of the fact that a Kuhn-Tucker multiplier may be different from zero only if the corresponding constraint is active. Indeed, either $\mu_{ij}^{n+1} = 0$ (the constraint is non-active), or, if $\mu_{ij}^{n+1} > 0$, then necessarily (the constraint is active)

$$D_{ij}(\mathbf{q}_h^n) + h\mathbf{G}_{ij}(\mathbf{q}_h^n) \cdot \mathbf{u}_h^{n+1} = 0,$$

so that $\mathbf{G}_{ij}(\mathbf{q}_h^n) \cdot \mathbf{u}_h^{n+1} = -D_{ij}(\mathbf{q}_h^n)/h \leq 0$ by Proposition 2.

We perform the scalar product of the relation (18) with the velocity field $\mathbf{u}_h^{n+1}$:

$$\begin{aligned} \left|\mathbf{u}_h^{n+1}\right|_{\mathrm{M}}^2 &= \mathrm{M}\mathbf{u}_h^n \cdot \mathbf{u}_h^{n+1} + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n) \cdot \mathbf{u}_h^{n+1} \\ &\quad + h \sum_{1 \leq i < j \leq N} \underbrace{\mu_{ij} \mathbf{G}_{ij} \cdot \mathbf{u}_h^{n+1}}_{\leq 0} \\ &\leq \left|\mathbf{u}_h^n\right|_{\mathrm{M}} \left|\mathbf{u}_h^{n+1}\right|_{\mathrm{M}} + h \left|\mathbf{f}_h^{n+1}(\mathbf{q}_h^n) \cdot \mathbf{u}_h^{n+1}\right| \end{aligned}$$

with

$$\begin{aligned} h \left|\mathbf{f}_h^{n+1}(\mathbf{q}_h^n) \cdot \mathbf{u}_h^{n+1}\right| &= \left| \int_{nh}^{(n+1)h} \mathbf{f}(\mathbf{q}_h^n, s) \cdot \mathbf{u}_h^{n+1}\, \mathrm{ds} \right| \\ &= \left| \int_{nh}^{(n+1)h} \mathrm{M}^{-1}\mathbf{f}(\mathbf{q}_h^n, s) \cdot \mathrm{M}\mathbf{u}_h^{n+1}\, \mathrm{ds} \right| \\ &\leq \frac{1}{\sqrt{\alpha}} \left|\mathbf{u}_h^{n+1}\right|_{\mathrm{M}} \int_{nh}^{(n+1)h} F(s)\, \mathrm{ds}. \end{aligned}$$

We finally get, by summing up over steps $0, 1, \ldots, n$,

$$\left|\mathbf{u}_h^{n+1}\right|_{\mathrm{M}} \leq |\mathbf{u}_0|_{\mathrm{M}} + \frac{1}{\sqrt{\alpha}} \int_0^{(n+1)h} F(s)\, \mathrm{ds},$$

which gives inequality (24). □

*Remark 9* In the case $\mathbf{f} \equiv \mathbf{0}$, the previous proposition expresses the fact that kinetic energy decreases at the discrete level, for any time step.

**Proposition 4** *(Momentum balance) If the forcing term does not depend on* $\mathbf{q}$*, then the total momentum balance is exactly verified at the right end of each subinterval by the approximated solution.*

*Proof* Let us denote by $\mathbf{P}(\mathbf{u})$ the total momentum associated to the velocity field $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_N)$:

$$\mathbf{P}(\mathbf{u}) = \sum_{i=1}^{N} m_i \mathbf{u}_i,$$

and by $\mathbf{f}_i$ the force acting on sphere $i$. As the approximated reaction forces verify the law of action and reaction, it follows from (18) that

$$\mathbf{P}(\mathbf{u}_h^{n+1}) = \mathbf{P}(\mathbf{u}_h^n) + \int_{nh}^{(n+1)h} \bar{\mathbf{f}}(s)\, \mathrm{d}s,$$

where $\bar{\mathbf{f}} = \sum \mathbf{f}_i$ is the resultant force, so that

$$\lim_{t \to (n+1)h^-} \mathbf{P}(\mathbf{u}_h(t)) = \mathbf{P}(\mathbf{u}_h^{n+1}) = \mathbf{P}(\mathbf{u}_0) + \int_0^{(n+1)h} \bar{\mathbf{f}}(s)\, \mathrm{d}s,$$

which is exactly the momentum balance. □

*Remark 10* Note that the previous property remains valid even if the saddle-point problem is not solved exactly.

*Remark 11* As a direct consequence of the previous property, the computed trajectory $\bar{\mathbf{q}}_h$ of the center of mass of the system turns out to be an interpolation of its exact trajectory (which is uniquely defined, even if particle trajectories are not), which can be expressed with obvious notations

$$\bar{\mathbf{q}} = \bar{\mathbf{q}}_0 + \bar{\mathbf{u}}_0\, t + \int_0^t \int_0^\tau M^{-1} \bar{\mathbf{f}}(s)\, \mathrm{d}s\, \mathrm{d}\tau.$$

We shall complete this section by a proposition concerning the Kuhn-Tucker multipliers. As $\mathbf{u}_h^{n+1}$ is uniquely defined, so is the global reaction term

$$\mathcal{G}_{\mathbf{q}_h^n}(\boldsymbol{\mu}_h^{n+1}) = \sum_{1 \le i < j \le N} \mu_{ij}^{n+1} \mathbf{G}_{ij}(\mathbf{q}_h^n).$$

As for the vector $\boldsymbol{\mu}_h^{n+1}$ itself, uniqueness does not hold in general, as shown by the following example: considering a two-dimensional collection of $N$ identical discs in a cristal-like configuration such that most discs are in contact with 6 others, the number of active constraints is asymptotically $3N$, whereas the number of degrees of freedom is $2N$. This non-uniqueness of the vector of Kuhn-Tucker multipliers is known to lead to numerical instabilities in some situations. We show next that the risk of instabilities is actually limited, because the solution set is bounded, as will be deduced from the following lemma.

**Lemma 1** *We consider* $\mathbf{q} \in Q_0$, $\boldsymbol{\mu} = (\mu_{ij}) \in \mathbb{R}^r_+$, *and we define*

$$\mathbf{F} = \mathcal{G}_{\mathbf{q}}(\boldsymbol{\mu}) = \sum_{1 \le i < j \le N} \mu_{ij} \mathbf{G}_{ij}(\mathbf{q}) \in \mathbb{R}^{dN}.$$

*The set* $\Lambda_{\mathbf{F}} = \left\{ \boldsymbol{\lambda} = (\lambda_{ij}) \in \mathbb{R}^r_+ \ , \ \ \mathcal{G}_{\mathbf{q}}(\boldsymbol{\lambda}) = \mathbf{F} \right\}$ *is bounded.*

*Proof* Let us first establish the uniqueness for the homogeneous problem. For $\mathbf{q} \in Q_0$, let $\boldsymbol{\lambda} = (\lambda_{ij}) \in \mathbb{R}^r_+$ be such that

$$\mathcal{G}_{\mathbf{q}}(\boldsymbol{\lambda}) = \sum_{1 \le i < j \le N} \lambda_{ij} \mathbf{G}_{ij}(\mathbf{q}) = 0.$$

Let $i_0$ denote the index of an extremal vertex of the convex hull $\mathrm{conv}\{\mathbf{q}_i, \ 1 \le i \le N\}$. By Hahn-Banach's theorem, the compact $\{\mathbf{q}_{i_0}\}$ and the closed convex set $\mathrm{conv}\{\mathbf{q}_i, \ 1 \le i \le N, \ i \ne i_0\}$ can be separated in a strict sense by a plane in $\mathbb{R}^d$. We denote by $\mathbf{x}$ an element of this plane, and by $\mathbf{v}$ a normal vector to it. One has

$$\left(\mathbf{q}_{i_0} - \mathbf{x}\right) \cdot \mathbf{v} > 0 \ , \ \ \left(\mathbf{q}_j - \mathbf{x}\right) \cdot \mathbf{v} < 0 \quad \forall j = 1, \dots, N \ , \ \ j \ne i_0,$$

so that $(\mathbf{q}_{i_0} - \mathbf{q}_j) \cdot \mathbf{v} > 0$ for $j \ne i_0$. Now the balance of contact forces exerted upon sphere $i_0$ in the direction $\mathbf{v}$ reads

$$\sum_{j \ne i_0} \lambda_{j i_0} \mathbf{e}_{j i_0}(\mathbf{q}) \cdot \mathbf{v} = \sum_{j \ne i_0} \lambda_{j i_0} \frac{\mathbf{q}_{i_0} - \mathbf{q}_j}{\left|\mathbf{q}_{i_0} - \mathbf{q}_j\right|} \cdot \mathbf{v}$$

which is positive unless $\lambda_{j i_0} = 0$ for all $j \ne i_0$. Therefore all multipliers associated to a contact with sphere $i_0$ are equal to 0, and this approach can be iterated for the reduced family $(\mathbf{q}_j, \ j \ne i_0)$. By downward induction on the number of active spheres, we establish that $\Lambda_0$ is actually reduced to $\{0\}$.

As for the non-homogeneous problem, $\Lambda_{\mathbf{F}}$ is obviously convex. The asymptotic cone of $\Lambda_{\mathbf{F}}$ is defined as (see e.g. Bourbaki [3])

$$\mathcal{C}_{\mathbf{F}} = \bigcap_{s > 0} s(\Lambda_{\mathbf{F}} - \boldsymbol{\lambda}),$$

where $\boldsymbol{\lambda}$ is any element of $\Lambda_{\mathbf{F}}$ (for example $\boldsymbol{\mu}$). It can be checked that (the proof, which is elementary, can be found in Maury [17]), in a finite-dimensional space, the asymptotic cone of a convex set is reduced to $\{0\}$ if and only if the set is bounded. If $\Lambda_{\mathbf{F}}$ is not bounded, then its asymptotic cone $\mathcal{C}_{\mathbf{F}}$ contains a half line $\mathbb{R}_+ \boldsymbol{\xi}$, with $\boldsymbol{\xi} \in \mathbb{R}^r_+$, which yields $\boldsymbol{\mu} + \mathbb{R}_+ \boldsymbol{\xi} \subset \Lambda_{\mathbf{F}}$, and consequently $\boldsymbol{\xi}$ is a solution to the homogeneous problem, so that $\boldsymbol{\xi} = 0$. Therefore, $\Lambda_{\mathbf{F}}$ is bounded. □

**Proposition 5** *At each time step of Scheme* $(\mathcal{S})$, *the solution set for* $\boldsymbol{\mu}_h^{n+1}$ *is bounded.*

*Proof* This is a direct consequence of lemma 1, with $\mathbf{q} = \mathbf{q}_h^n$ and

$$h\mathbf{F} = \mathrm{M}(\mathbf{u}_h^{n+1} - \mathbf{u}_h^n) - h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n).$$

□

## 5 Convergence theorem (case of a single contact)

We establish in this section a convergence result for the proposed time-stepping scheme, in the case of a single constraint (two spheres) in $\mathbb{R}^3$. As uniqueness does not hold, neither does continuity with respect to initial conditions. Thus we cannot expect any error estimate for the scheme $(\mathcal{S})$. Moreover, the necessity to extract a subsequence to establish convergence is not purely technical: different subsequences may actually converge to distinct solutions to the same problem (see Section 6.1).

In order to alleviate notations, we consider the system of two spheres with unit mass and same radius $r$, but this assumption is not needed in the proof. The result remains valid for the case of two different spheres, or a sphere and an obstacle. The constrained minimization problem is again supposed to be solved exactly at each time step.

### 5.1 Two-body system

We consider here the mechanical sytem of two identical, hard spheres, described by

$$Q = \{\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2) \in \mathbb{R}^3 \times \mathbb{R}^3\},$$

and the associated feasible set

$$Q_0 = \{\mathbf{q} \in Q \ , \quad D(\mathbf{q}) \geq 0\},$$

where $D(\mathbf{q}) = |\mathbf{q}_2 - \mathbf{q}_1| - 2r$ is the distance between the two bodies. As previously, the tangent space at any $\mathbf{q} \in Q$ is denoted by $\mathrm{T}_Q = \mathbb{R}^6$. The convex cone $\mathcal{N}_\mathbf{q}$ is now

$$\mathcal{N}_\mathbf{q} = \begin{cases} \{0\} & \text{if } D(\mathbf{q}) > 0, \\ -\mathbb{R}^+\mathbf{G} = \{-\lambda\mathbf{G} \ , \quad \lambda \in \mathbb{R}^+\} & \text{if } D(\mathbf{q}) = 0 \end{cases}$$

where $\mathbf{G} = \nabla D$ is the gradient of $D$. In order to avoid special care of the initial instant, we shall pick an initial configuration $\mathbf{q}_0$ such that $D(\mathbf{q}_0) > 0$.

We may now state the single-contact problem: given a time interval $I = (0, T)$, a force field $(\mathbf{q}, t) \mapsto \mathbf{f}(\mathbf{q}, t) \in \mathbb{R}^6$ verifying the regularity conditions (6)–(7) and the initial conditions $(\mathbf{q}_0, \mathbf{u}_0) \in Q_0 \times \mathrm{T}_Q$ with $D(\mathbf{q}_0) > 0$, find $(\mathbf{q}, \mathbf{u}, \mu) \in W^{1,1} \times \mathrm{BV} \times \mathcal{M}^1_+$ such that

$$\mathbf{u}(0) = \mathbf{u}_0, \tag{26}$$

$$\mathbf{q}(t) = \mathbf{q}_0 + \textstyle\int_0^t \mathbf{u}(s)\ \mathrm{ds} \in Q_0 \quad \forall t \in I, \tag{27}$$

$$\dot{\mathbf{u}}(t) = \mathbf{f}(\mathbf{q}(t), t) + \mu\, \mathbf{G}(\mathbf{q}(t)), \tag{28}$$

$$\mathrm{supp}(\mu) \subset \{t \ , \quad D(\mathbf{q}(t)) = 0\}, \tag{29}$$

$$\mathbf{u}^+ = \mathbf{u}^- - P_{\mathcal{N}_\mathbf{q}}\mathbf{u}^- \quad \forall t \in I, \tag{30}$$

where (28) is understood in the sense of distributions, and $P_{\mathcal{N}_\mathbf{q}}$ is the Euclidean projection onto the closed convex cone $\mathcal{N}_\mathbf{q}$.

Let us now rewrite the overall time stepping-scheme in the present case of a single contact. The discretization spaces for pathlines ($X_h$), velocities ($V_h$), and reactions ($R_h$), are

$$X_h = \{\mathbf{q}_h = (\mathbf{q}_i) : I \to Q \ , \quad \mathbf{q}_i \in (\mathrm{P}_h^1)^3 \ , \quad i = 1, 2\},$$

$$V_h = \{\mathbf{u}_h = (\mathbf{u}_i) : I \to \mathrm{T}_Q \ , \quad \mathbf{u}_i \in (\mathrm{P}_h^0)^3 \ , \quad i = 1, 2\},$$

$$R_h = \left\{\mu_h : I \to \mathbb{R} \ , \quad \mu_h \in \mathrm{P}_h^0\right\}$$

The approximated force field $\mathbf{f}_h(\mathbf{q}) \in V_h$ is defined for any $\mathbf{q} \in Q$ by

$$t \in [(n-1)h, nh) \ \longmapsto \ \mathbf{f}_h^n(\mathbf{q}) = \frac{1}{h} \int_{(n-1)h}^{nh} \mathbf{f}(\mathbf{q}, s) \ \mathrm{ds}.$$

For a given $h > 0$, we shall consider

$$(\mathbf{q}_h, \mathbf{u}_h, \mu_h) \in X_h \times V_h \times R_h$$

built according to Scheme ($\mathcal{S}$):

1. Initialization

$$(\mathbf{q}_h^0, \mathbf{u}_h^0) = (\mathbf{q}_0, \mathbf{u}_0). \tag{31}$$

2. Time step

$$\mathbf{u}_h^{n+1} = \mathbf{u}_h^n + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n) + h\mu_h^{n+1}\mathbf{G}(\mathbf{q}_h^n) \tag{32}$$

where $\mathbf{u}_h^{n+1}$ is the solution to the constrained minimization problem

$$\min_{\mathbf{u} \in \mathrm{K}_h(\mathbf{q}_h^n)} |\mathbf{u} - \mathbf{u}_h^n - h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n)|^2 \tag{33}$$

with

$$\mathrm{K}_h(\mathbf{q}_h^n) = \{\mathbf{u} \in \mathrm{T}_Q \ , \quad D(\mathbf{q}_h^n) + h\mathbf{G}(\mathbf{q}_h^n) \cdot \mathbf{u} \geq 0\}, \tag{34}$$

and $\mu_h^{n+1} \in \mathbb{R}^+$ is the associated Kuhn-Tucker multiplier.

3. Update the positions

$$\mathbf{q}_h^{n+1} = \mathbf{q}_h^n + h\mathbf{u}_h^{n+1}. \tag{35}$$

## 5.2 Convergence

**Theorem 1** *Let* $(\mathbf{q}_h, \mathbf{u}_h, \mu_h)_h$ *be a sequence of solutions obtained by Scheme* $(\mathcal{S})$ *(equations (31) to (35)), with* $h \to 0$*. Then there exists a subsequence of time steps (still denoted by* $h$*), and*

$$(\mathbf{q}, \mathbf{u}, \mu) \in W^{1,1} \times \mathrm{BV} \times \mathcal{M}^1_+$$

*such that*

$$\mathbf{q}_h \longrightarrow \mathbf{q} \quad \text{in } W^{1,1},$$

$$\mu_h \overset{\star}{\rightharpoonup} \mu \quad \text{in } \mathcal{M}^1,$$

*and* $(\mathbf{q}, \mathbf{u}, \mu)$ *is a solution to problem (26)–(30).*

*Proof* The proof will be decomposed into 9 steps, which we shall briefly describe below. The critical ones are steps 3 and 9.

1. The scheme produces feasible configurations only, *i.e.*,
$$\mathbf{q}_h(t) \in Q_0 \quad \forall h, t.$$
2. The family $(\mathbf{u}_h)$ is uniformly bounded, *i.e.*,
$$\exists C_\infty \ , \quad |\mathbf{u}_h(t)| \leq C_\infty \quad \forall t \in [0, T] \ , \quad \forall h > 0.$$
3. The fields $\mathbf{u}_h$ have uniform bounded variation, *i.e.*,
$$\exists C_{\mathrm{var}} \ , \quad \mathrm{var}(\mathbf{u}_h) = \sum_{n=1}^{N} |\mathbf{u}_h^n - \mathbf{u}_h^{n-1}| \leq C_{\mathrm{var}} \quad \forall h.$$
4. The familly $(\mathbf{u}_h)$ is relatively compact in $L^1(I)$. One can extract a subsequence (still denoted $\mathbf{q}_h$) such that $\mathbf{q}_h \longmapsto \mathbf{q}$ in $W^{1,1}$. The limit velocity $\mathbf{u} = \dot{\mathbf{q}} = \lim \dot{\mathbf{q}}_h$ is in BV, and the limit motion $\mathbf{q}$ is feasible.
5. The sequence $(\mu_h)_h$ is bounded in $L^1$, so that one can extract a subsequence which converges weak-$\star$ to a vector-valued bounded measure $\mu \in \mathcal{M}^1$.
6. The pair $(\mathbf{u}, \mu)$ verifies (28).
7. The complementarity slackness condition (29) holds true.
8. The initial condition (26) is verified.
9. The jump equation (30) is verified.

□

**Step 1** (*Feasibility*)

This is exactly Proposition 2, which was established for any number of contacts:

$$D(\mathbf{q}_h(t)) \geq 0 \quad \forall h, \ t \in [0, T].$$

**Step 2** ($\mathbf{u}_h$ *is uniformly bounded*)

By Proposition 3,

$$|\mathbf{u}_h(t)|_{\mathrm{M}} \leq |\mathbf{u}_0|_{\mathrm{M}} + \frac{1}{\sqrt{\alpha}} \int_0^T F(t)\,\mathrm{dt} \leq C_\infty.$$

**Step 3** ($\mathbf{u}_h$ *has uniformly bounded variation*)

We shall first establish a lemma, which can be seen as a mono-dimensional version of the property.

**Lemma 2** *We consider sequences* $(u_h)_h$, $(\rho_h)_h$, *and* $(g_h)_h$ *in* $\mathrm{P}_h^0$ *such that*

$$u_h^{n+1} = hg_h^n + \rho_h^n u_h^n \ , \quad \rho_h^n \in [0,1] \quad \forall n \ , \quad \forall h,$$

*with* $u_h^0 = u_0$. *We suppose furthermore that there exists a constant* $C \geq 0$ *such that*

$$h \sum_{n=0}^{N-1} \left| g_h^n \right| \leq C \quad \forall h. \tag{36}$$

*Then the familly* $(u_h)$ *has uniform bounded variation:*

$$\exists C_{\mathrm{var}} \ , \quad \mathrm{var}(u_h) = \sum_{n=0}^{N-1} \left| u_h^{n+1} - u_h^n \right| \leq C_{\mathrm{var}} \quad \forall h.$$

*Proof* A straightforward calculation leads to

$$\begin{aligned} u_h^{n+1} &= \rho_h^0 \rho_h^1 \dots \rho_h^n \, u_0 \\ &\quad + hg_h^n + h\rho_h^n g_h^{n-1} + h\rho_h^n \rho_h^{n-1} g_h^{n-2} \\ &\quad + \dots + h\rho_h^n \rho_h^{n-1} \dots \rho_h^1 g_h^0 \\ &= \rho_h^0 \rho_h^1 \dots \rho_h^n \, u_0 + \sum_{k=0}^{n} h\lambda_h^k g_h^k, \end{aligned}$$

with $\lambda_h^k \in [0,1]$ for any $h, k$. Therefore function $u_h$ can be written as a sum of the homogeneous solution $v_h$ defined by

$$v_h^{n+1} = \rho_h^n v_h^n = \rho_h^n \rho_h^{n-1} \dots \rho_h^0 u_0$$

and a function $w_h$ determined by

$$w_h^{n+1} = \sum_{k=0}^{n} h\lambda_h^k g_h^k.$$

The first one $v_h$ is monotonous and keeps a constant sign, so that $\mathrm{var}(v_h) \leq |u_0|$. The second one verifies

$$\mathrm{var}(w_h) = \sum_{n=0}^{N-1} \left| w_h^{n+1} - w_h^n \right| = \sum_{n=0}^{N-1} h\lambda_h^n \left| g_h^n \right| \leq C,$$

so that finally

$$\mathrm{var}(u_h) \leq |u_0| + C,$$

which ends the proof of Lemma 2. □

Remark now that, as $\mathbf{u}_h^{n+1}$ is defined as the Euclidean projection of

$$\mathbf{u}_h^n + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n)$$

onto the half space of $\mathrm{T}_Q$

$$\mathrm{K}_h(\mathbf{q}_h^n) = \{\mathbf{u} \in \mathrm{T}_Q \ , \quad D(\mathbf{q}_h^n) + h\,\mathbf{G}(\mathbf{q}_h^n) \cdot \mathbf{u} \geq 0\},$$

whose boundary is a hyperplane of $\mathrm{T}_Q$ with normal $\tilde{\mathbf{G}} = \mathbf{G}/\left|\mathbf{G}\right|$, one has

$$\mathbf{u}_h^{n+1} - (\mathbf{u}_h^n + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n)) = \alpha\,\tilde{\mathbf{G}}$$

where $\alpha \geq 0$ is some real parameter. If the constraint is non-active (*i.e.* $\mathbf{u}_h^n + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n) \in \mathrm{K}_h(\mathbf{q}_h^n)$), then $\alpha = 0$. Otherwise, as $D(\mathbf{q}_h^n) \geq 0$,

$$\left(\mathbf{u}_h^n + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n)\right) \cdot \tilde{\mathbf{G}} \leq 0,$$

so that $\alpha\tilde{\mathbf{G}}$ can be written

$$\alpha\tilde{\mathbf{G}} = -\lambda_h^n((\mathbf{u}_h^n + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n)) \cdot \tilde{\mathbf{G}})\,\tilde{\mathbf{G}},$$

where $\lambda_h^n$ is a non-negative real number. Let us show that $\lambda_h^n \leq 1$. Again, as $D(\mathbf{q}_h^n)$ is non-negative, 0 lies in $\mathrm{K}_h(\mathbf{q}_h^n)$, therefore

$$\left(\mathbf{u}_h^{n+1} - (\mathbf{u}_h^n + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n)), \mathbf{u}_h^{n+1} - 0\right) \leq 0,$$

which yields

$$\lambda_h^n(\lambda_h^n - 1)\left((\mathbf{u}_h^n + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n)) \cdot \tilde{\mathbf{G}}\right)^2 \leq 0,$$

and therefore $\lambda_h^n \leq 1$. We established that

$$\mathbf{u}_h^{n+1} - (\mathbf{u}_h^n + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n)) = -\lambda_h^n((\mathbf{u}_h^n + h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n)) \cdot \tilde{\mathbf{G}})\,\tilde{\mathbf{G}},$$

with $\lambda_h^n \in [0, 1]$.

Let us now introduce the Euclidean decomposition

$$\mathbf{u}_h^n = U_h^n\tilde{\mathbf{G}}(\mathbf{q}_h^n) + \mathbf{v}_h^n,$$

where $\mathbf{v}_h^n$ is the projection of $\mathbf{u}_h^n$ onto $\tilde{\mathbf{G}}^\perp$. We verify now that $U_h^n$ meets the assumptions of Lemma 2 :

$$\begin{aligned}
U_h^{n+1} &= \mathbf{u}_h^{n+1} \cdot \tilde{\mathbf{G}}\left(\mathbf{q}_h^{n+1}\right) \\
&= \mathbf{u}_h^{n+1} \cdot \tilde{\mathbf{G}}\left(\mathbf{q}_h^n\right) + \mathbf{u}_h^{n+1} \cdot \left(\tilde{\mathbf{G}}\left(\mathbf{q}_h^{n+1}\right) - \tilde{\mathbf{G}}\left(\mathbf{q}_h^n\right)\right) \\
&= (1 - \lambda_h^n)\,\mathbf{u}_h^n \cdot \tilde{\mathbf{G}}\left(\mathbf{q}_h^n\right) + (1 - \lambda_h^n)\,h\,\mathbf{f}_h^{n+1}(\mathbf{q}_h^n) \cdot \tilde{\mathbf{G}}\left(\mathbf{q}_h^n\right) \\
&\quad + \mathbf{u}_h^{n+1} \cdot \left(\tilde{\mathbf{G}}\left(\mathbf{q}_h^{n+1}\right) - \tilde{\mathbf{G}}\left(\mathbf{q}_h^n\right)\right) \\
&= \rho_h^n U_h^n + \rho_h^n h\,\mathbf{f}_h^{n+1}\left(\mathbf{q}_h^n\right) \cdot \tilde{\mathbf{G}}\left(\mathbf{q}_h^n\right) \\
&\quad + \mathbf{u}_h^{n+1} \cdot \left(\tilde{\mathbf{G}}\left(\mathbf{q}_h^{n+1}\right) - \tilde{\mathbf{G}}\left(\mathbf{q}_h^n\right)\right),
\end{aligned}$$

with $\rho_h^n = 1 - \lambda_h^n$. As $\mathbf{q}_h^{n+1} = \mathbf{q}_h^n + h\mathbf{u}_h^{n+1}$, with $\left|\mathbf{u}_h^{n+1}\right| \leq C_\infty$, and as $\tilde{\mathbf{G}}$ is $\tilde{k}$-Lipschitz in $Q_0$, one has

$$\left|\mathbf{u}_h^{n+1} \cdot \left(\tilde{\mathbf{G}}(\mathbf{q}_h^{n+1}) - \tilde{\mathbf{G}}(\mathbf{q}_h^n)\right)\right| \leq \left|\mathbf{u}_h^{n+1}\right| \tilde{k}\, C_\infty h \leq \tilde{k}\, C_\infty^2 h.$$

Finally, $h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n) \cdot \tilde{\mathbf{G}}(\mathbf{q}_h^n) + \mathbf{u}_h^{n+1} \cdot \left(\tilde{\mathbf{G}}(\mathbf{q}_h^{n+1}) - \tilde{\mathbf{G}}(\mathbf{q}_h^n)\right)$ can be written as $hg_h^n$, and the corresponding family of piecewise constant functions $(g_h)$ verifies (36). As $\rho_h^n \in [0, 1]$, all assumptions of Lemma 2 are met, which implies that $\mathrm{var}(U_h)$ is bounded uniformly in $h$.

We denote by $P_h^n$ the projection onto $\tilde{\mathbf{G}}(\mathbf{q}_h^n)^\perp$, so that $\mathbf{v}_h^n = P_h^n \mathbf{u}_h^n$. We conclude this step by writing

$$\begin{aligned}
\mathrm{var}(\mathbf{u}_h) &= \sum_{n=0}^{N-1} \left|\mathbf{u}_h^{n+1} - \mathbf{u}_h^n\right| \\
&\leq \sum_{n=0}^{N-1} \left|(\mathbf{u}_h^{n+1} - \mathbf{u}_h^n) \cdot \tilde{\mathbf{G}}(\mathbf{q}_h^n)\right| + \sum_{n=0}^{N-1} \left|P_h^n(\mathbf{u}_h^{n+1} - \mathbf{u}_h^n)\right| \\
&\leq \sum_{n=0}^{N-1} \left|U_h^{n+1} - U_h^n\right| + \sum_{n=0}^{N-1} \left|\mathbf{u}_h^{n+1} \cdot (\tilde{\mathbf{G}}(\mathbf{q}_h^{n+1}) - \tilde{\mathbf{G}}(\mathbf{q}_h^n))\right| \\
&\quad + \sum_{n=0}^{N-1} \left|h\, P_h^n \mathbf{f}_h^{n+1}(\mathbf{q}_h^n)\right| \quad \left(\text{because } P_h^n \tilde{\mathbf{G}}(\mathbf{q}_h^n) = 0\right) \\
&\leq \mathrm{var}(U_h) + T\tilde{k}\, C_\infty^2 + ||\mathrm{F}||_1.
\end{aligned}$$

**Step 4** ($\mathbf{q}_h$ *converges to a feasible trajectory*)

As $(\mathbf{u}_h)$, which is bounded in $L^\infty$, is now proved to have uniform bounded variations over $I$, it is relatively compact in $L^1$, so that one can extract a subsequence, still denoted by $(\mathbf{u}_h)$, which converges in $L^1$ to $\mathbf{u}$. As $\mathbf{u}_h$ is the Sobolev derivative of $\mathbf{q}_h$ (in $W^{1,1}$), then $\mathbf{q}_h$ converges to a $\mathbf{q}$ in $W^{1,1}$, and $\dot{\mathbf{q}} = \mathbf{u}$. The limit velocity $\mathbf{u}$ is in BV. Indeed, for any $\boldsymbol{\varphi} \in (C_c^1(I))^6$,

$$\left|\int_I \mathbf{u} \cdot \boldsymbol{\varphi}'\right| = \lim_{h\to 0} \left|\int_I \mathbf{u}_h \cdot \boldsymbol{\varphi}'\right| \leq \sup_h \mathrm{var}(\mathbf{u}_h)||\boldsymbol{\varphi}||_\infty < +\infty.$$

Finally, as the convergence $\mathbf{q}_h \longrightarrow \mathbf{q}$ is uniform ($W^{1,1} \hookrightarrow L^\infty$), and $D(\mathbf{q}_h(t)) \geq 0$ for every $h$, $t$, with $D$ continuous with respect to $\mathbf{q}$, then $D(\mathbf{q}(t)) \geq 0$ for every $t \in I$.

**Step 5** ($\mu_h$ *converges weak-*$\star$ *to* $\mu$)

For any $h$,

$$\begin{aligned}
|\mathbf{G}| \int_I |\mu_h| &\leq \sum_{n=0}^{N-1} \left|\mathbf{u}_h^{n+1} - \mathbf{u}_h^n\right| + \sum_{n=0}^{N-1} h \left|\mathbf{f}_h^{n+1}(\mathbf{q}_h^n)\right| \\
&\leq \mathrm{var}(\mathbf{u}_h) + ||\mathrm{F}||_1,
\end{aligned}$$

so that $(\mu_h)$ is bounded in $L^1$ ($|\mathbf{G}|$ is the modulus of $\mathbf{G}(\mathbf{q}_h)$, which is a positive constant). Therefore, by the De La Vallée Poussin compactness criterion, one can extract a subsequence which converges weakly-$\star$ to $\mu$ in $\mathcal{M}^1$.

**Step 6** (*The balance of momentum is verified by the limits* $\mathbf{u}$ *and* $\mu$)

Let us first remark that the derivative of $\mathbf{u}_h$ can be defined[3] in the sense of distributions as

$$\dot{\mathbf{u}}_h = \sum_{n=1}^{N-1} (\mathbf{u}_h^{n+1} - \mathbf{u}_h^n) \cdot \delta_n,$$

where $\delta^n$ is the Dirac mass at $t^n = nh$. For any $\boldsymbol{\varphi} \in (D(I))^6$, it comes

$$\begin{aligned} \int_I \dot{\mathbf{u}}_h \cdot \boldsymbol{\varphi} &= \sum_{n=1}^{N-1} (\mathbf{u}_h^{n+1} - \mathbf{u}_h^n) \cdot \boldsymbol{\varphi}(t^n) = -\sum_{n=1}^{N} \mathbf{u}_h^n \cdot (\boldsymbol{\varphi}(t^n) - \boldsymbol{\varphi}(t^{n-1})) \\ &= -\sum_{n=1}^{N} h\mathbf{u}_h^n \cdot (\boldsymbol{\varphi}'(t^n) + o(1)) \longrightarrow -\int_I \mathbf{u}(t) \cdot \boldsymbol{\varphi}'(t)\ \mathrm{dt} \end{aligned} \tag{37}$$

when $h$ goes to 0. The integral $\int_I \dot{\mathbf{u}}_h \cdot \boldsymbol{\varphi}$ may also be expressed

$$\begin{aligned} \int_I \dot{\mathbf{u}}_h \cdot \boldsymbol{\varphi} &= \sum_{n=1}^{N-1} (h\mathbf{f}_h^{n+1}(\mathbf{q}_h^n) + h\mu_h^{n+1}\mathbf{G}(\mathbf{q}_h^n)) \cdot \boldsymbol{\varphi}(t^n) \\ &= \int_I (\mathbf{f}_h(\overline{\mathbf{q}}_h) + \mu_h \mathbf{G}_h) \cdot \boldsymbol{\varphi}_h\ \mathrm{dt} \end{aligned}$$

where $\boldsymbol{\varphi}_h \in V_h$ is defined by taking the left value of $\boldsymbol{\varphi}$ on each subinterval $[t^n, t^{n+1})$, and similarly $\overline{\mathbf{q}}_h$ stands for the function of $V_h$ which is equal to $\mathbf{q}_h^n$ on $[t^n, t^{n+1})$. As $\mathbf{f}$ is uniformly Lipschitz with respect to $\mathbf{q}$, and as $\overline{\mathbf{q}}_h$ converges to $\mathbf{q}$ in $L^\infty$, $\mathbf{f}_h(\overline{\mathbf{q}})$ converges to $t \mapsto \mathbf{f}(t, \mathbf{q}(t))$ in $L^1$. Similarly $\boldsymbol{\varphi}_h$ converges to $\boldsymbol{\varphi}$ in $L^\infty$, so that

$$\int_I \mathbf{f}_h(\overline{\mathbf{q}}_h) \cdot \boldsymbol{\varphi}_h \overset{h\to 0}{\longrightarrow} \int_I \mathbf{f}(t, \mathbf{q}(t)) \cdot \boldsymbol{\varphi}(t)\ \mathrm{dt}.$$

For the second term, we remark that $(\mu_h)$ can be seen as a bounded sequence of linear functionals on $L^\infty$. According to this remark, one may write ($\mathbf{G} \circ \mathbf{q}$ designs $t \longmapsto \mathbf{G}(\mathbf{q}(t))$, which is the uniform limit of $\mathbf{G} \circ \mathbf{q}_h$)

$$\begin{aligned} \int_I \mu_h \mathbf{G}_h \cdot \boldsymbol{\varphi}_h\ \mathrm{dt} &= < \mu_h, \mathbf{G}_h \cdot \boldsymbol{\varphi}_h > \\ &= < \mu_h, \mathbf{G} \circ \mathbf{q} \cdot \boldsymbol{\varphi} >_{\mathcal{M}^1, C_0(I)} \\ &\quad + < \mu_h, \mathbf{G} \circ \mathbf{q} \cdot \boldsymbol{\varphi} - \mathbf{G}_h \cdot \boldsymbol{\varphi}_h >_{(L^\infty)', L^\infty} \\ &\longrightarrow < \mu, \mathbf{G} \circ \mathbf{q} \cdot \boldsymbol{\varphi} > . \end{aligned}$$

[3] This notation means: for any $\mathbf{v} \in \mathrm{T}_Q$, $\boldsymbol{\varphi} \in (D(I))^6$, $< \mathbf{v} \cdot \delta_n, \boldsymbol{\varphi} >= \mathbf{v} \cdot \boldsymbol{\varphi}(t^n)$.

Finally we have

$$\int_I \dot{\mathbf{u}}_h \cdot \boldsymbol{\varphi} \longrightarrow \int_I \mathbf{f}(t, \mathbf{q}(t)) \cdot \boldsymbol{\varphi}(t) \ \mathrm{dt} + < \mu, \mathbf{G} \circ \mathbf{q} \cdot \boldsymbol{\varphi} > . \tag{38}$$

From (37) and (38), it follows

$$-\int_I \mathbf{u}(t) \cdot \boldsymbol{\varphi}'(t) \ \mathrm{dt} = \int_I \mathbf{f}(t, \mathbf{q}(t)) \cdot \boldsymbol{\varphi}(t) \ \mathrm{dt} + < \mu, \mathbf{G} \circ \mathbf{q} \cdot \boldsymbol{\varphi} >$$

for all $\boldsymbol{\varphi} \in D(I)$: this is equation (28) in the sense of distributions.

**Step 7** (*The limit reaction field $\mu$ is non-active when there is no contact*)

For any $t_0 \in I$ such that $D(\mathbf{q}(t_0)) > 0$, there exists $\eta$ such that

$$D(\mathbf{q}(t)) \geq a > 0 \quad \forall t \in (t_0 - \eta, t_0 + \eta),$$

so that, by uniform convergence of $\mathbf{q}_h$ to $\mathbf{q}$, there exists $h_1$ such that

$$D(\mathbf{q}_h(t)) \geq a/2 > 0 \quad \forall t \in ]t_0 - \eta, t_0 + \eta[ \quad \forall h \leq h_1.$$

Since the velocities $\mathbf{u}_h$ are uniformly bounded in $L^\infty$, there exists $h_2 \leq h_1$ such that, for any $h \leq h_2$,

$$D(\mathbf{q}_h) + h\mathbf{G}(\mathbf{q}_h) \cdot \mathbf{u}_h \geq a/4 > 0,$$

so that the numerical constraint $D + h\mathbf{G} \cdot \mathbf{u} \geq 0$ is not activated in $(t_0 - \eta, t_0 + \eta)$. Consequently, for $h \leq h_2$, $\mu_h \equiv 0$ in $(t_0 - \eta, t_0 + \eta)$, so that

$$\mathrm{supp}(\mu) \subset (t_0 - \eta, t_0 + \eta)^C,$$

which yields the property.

**Step 8** (*The initial condition is verified*)

As $\mathbf{q}_0$ is chosen apart from the boundary of $Q_0$ and thanks again to the uniform boundedness of $\mathbf{u}_h$ in $L^\infty$, there exists $\eta$ such that $\mu_h$ vanishes in $(0, \eta)$ for all values of $h$. It holds then

$$|\mathbf{u}_h(t) - \mathbf{u}_0| \leq \int_0^{t+h} \mathrm{F}(s) \, \mathrm{ds},$$

so that

$$|\mathbf{u}(t) - \mathbf{u}_0| \leq \int_0^t \mathrm{F}(s) \, \mathrm{ds} \quad a.e. \text{ in } (0, \eta),$$

which implies $\mathbf{u}(0) = \mathbf{u}_0$.

**Step 9** (*The collision law is verified*)

For $t_0$ such that $D(\mathbf{q}(t_0)) > 0$, we already established (see Step 7) that $\mu \equiv 0$ in a neighbourhood of $t_0$, so that $\mathbf{u}$ is absolutely continuous in this neighbourhood (as it solves (28)), hence continuous at $t_0$:

$$\mathbf{u}^+ = \mathbf{u}^- = \mathbf{u}^- - P_{\{0\}}\mathbf{u}^-.$$

The interesting case is of course $D(\mathbf{q}(t_0)) = 0$. We first establish from (28) that the discontinuity occurs along direction $\mathbf{G}$ :

**Lemma 3** *At any contact time $t_0$ there exists $\lambda \in \mathbb{R}$ such that*

$$\mathbf{u}^+(t_0) - \mathbf{u}^-(t_0) = \lambda \tilde{\mathbf{G}}(\mathbf{q}(t_0)),$$

*where $\tilde{\mathbf{G}}$ is the normalized gradient $\mathbf{G}/|\mathbf{G}|$ introduced in step 3.*

*Proof* For $\eta > 0$, and $\mathbf{H} \in \mathrm{T}_Q$ such that $\mathbf{H} \cdot \mathbf{G} = 0$, we consider the test function $\psi_\eta \mathbf{H}$ where $\psi_\eta$ is continuous, piecewise affine, zero outside $[t_0 - \eta, t_0 + \eta]$, taking values 0, 1, 0 at $t_0 - \eta$, $t_0$, $t_0 + \eta$, respectively[4]. Now writing (28) against $\psi_\eta \mathbf{H}$ yields

$$-\int_I \mathbf{u} \cdot \mathbf{H}\psi_\eta' = \int_I \psi_\eta \mathbf{H} \cdot \mathbf{f} + \langle \mu, \mathbf{G} \circ \mathbf{q} \cdot \mathbf{H}\psi_\eta \rangle .$$

The left-hand side can be expressed

$$\begin{aligned} -\int_I \mathbf{u} \cdot \mathbf{H}\psi_\eta' &= \frac{1}{\eta}\left(\int_{t_0}^{t_0+\eta} \mathbf{u} \cdot \mathbf{H} - \int_{t_0-\eta}^{t_0} \mathbf{u} \cdot \mathbf{H}\right) \\ &= \left(\mathbf{u}^+(t_0) - \mathbf{u}^-(t_0)\right) \cdot \mathbf{H} + o(\eta). \end{aligned}$$

The force integral $\int_I \psi_\eta \mathbf{H} \cdot \mathbf{f}$ goes to zero with $\eta$. The last term can be bounded

$$\left|\langle \mu, \mathbf{G} \circ \mathbf{q} \cdot \mathbf{H}\psi_\eta \rangle\right| \leq \left(\sup_{t \in [t_0-\eta, t_0+\eta]} |\mathbf{G} \circ \mathbf{q}(t) \cdot \mathbf{H}|\right) \left|\langle \mu, \psi_\eta \rangle\right|,$$

where the sup goes to zero because $\mathbf{G} \circ \mathbf{q}(t)$ tends to $\mathbf{G}(\mathbf{q}(t_0))$ as $\eta$ tends to 0, with $\mathbf{H} \cdot \mathbf{G}(\mathbf{q}(t_0)) = 0$. So finally the jump $(\mathbf{u}^+ - \mathbf{u}^-) \cdot \mathbf{H}$ is zero for any $\mathbf{H}$ orthogonal to $\mathbf{G}$, which ends the proof of Lemma 3. □

Unless explicitly mentioned, fields are now taken at $t_0$, where $t_0$ is a time at which contact occurs, and $\tilde{\mathbf{G}}$ designs $\tilde{\mathbf{G}}(\mathbf{q}(t_0))$. We first check that $\mathbf{u}^- \cdot \tilde{\mathbf{G}} \leq 0$. If otherwise, then $D(\mathbf{q}) = 0$ implies that

$$D(\mathbf{q}(t_0 - \varepsilon)) = -\varepsilon \mathbf{G} \cdot \mathbf{u}^- + o(\varepsilon)$$

is negative for some $\varepsilon > 0$, meaning that $\mathbf{q}(t_0 - \varepsilon) \notin Q_0$, which is impossible. So

$$\mathbf{u}^- \cdot \tilde{\mathbf{G}} \leq 0.$$

[4] This "hat" function is obviously not in $D(I)$, but it can be approximated by regular functions in norm $W^{1,1}$, which is enough here as $\mathbf{u} \in L^\infty$. Therefore it can be used as a test function.

Let us show that $\lambda \geq -\mathbf{u}^- \cdot \tilde{\mathbf{G}}$. If not, then

$$\mathbf{u}^+ \cdot \tilde{\mathbf{G}} = \mathbf{u}^- \cdot \tilde{\mathbf{G}} + \lambda \tilde{\mathbf{G}} \cdot \tilde{\mathbf{G}} < 0,$$

which implies similarly that $\mathbf{q}$ enters the forbidden domain $Q_0^C$ right after $t_0$. Therefore

$$\lambda \geq -\mathbf{u}^- \cdot \tilde{\mathbf{G}}. \tag{39}$$

Let us now notice that the condition (30) characterizes the velocity after the collision as the solution to a constrained minimization problem : it minimizes $\left|\mathbf{u} - \mathbf{u}^-\right|$ over $\mathcal{C}_{\mathbf{q}}$, the polar cone of $\mathcal{N}_{\mathbf{q}}$, which is the closed half space

$$\{\mathbf{u} \in \mathrm{T}_Q \ , \quad \mathbf{u} \cdot \mathbf{G} \geq 0\}.$$

As $\mathbf{u}^- \cdot \mathbf{G} \leq 0$, the solution to that problem is clearly

$$\tilde{\mathbf{u}} = \mathbf{u}^- - (\tilde{\mathbf{G}} \cdot \mathbf{u}^-)\tilde{\mathbf{G}},$$

which we have to identify with $\mathbf{u}^+$ to prove that the collision law (30) holds at time $t_0$. As we know that $\mathbf{u}^+$ is $\mathbf{u}^- + \lambda \tilde{\mathbf{G}}$ with condition (39) on $\lambda$, we just have to check that

$$\mathbf{u}^+ \cdot \mathbf{G} \leq \tilde{\mathbf{u}} \cdot \mathbf{G} = 0. \tag{40}$$

We extend notations of Step 3 by introducing $U \in L^1$ (which can be seen as the velocity in the moving $\mathbf{G}$ direction) as the limit of $\mathbf{u}_h \cdot \tilde{\mathbf{G}}(\mathbf{q}_h)$. Now we recall the property we want to establish : at any $t_0$ such that $D(\mathbf{q}(t_0)) = 0$, equation (40) is verified. As $t \mapsto \mathbf{G}(\mathbf{q}(t))$ is continuous, it can be formulated in terms of $U$ : at any contact time, $U^+ \leq 0$. We propose actually to establish the property (which is stronger as $U^- \leq 0$ at any contact time) :

$$U^- \leq 0 \Longrightarrow U^+ \leq 0. \tag{41}$$

*Proof of (41) :* We first recall an important property of $U_h = \mathbf{u}_h \cdot \tilde{\mathbf{G}}(\mathbf{q}_h)$ which was established in Step 3 : there exists sequences $g_h^n$ and $\rho_h^n$ such that

$$U_h^{n+1} = h g_h^n + \rho_h^n U_h^n,$$

with $\rho_h^n \in [0, 1]$ and $|g_h| \leq g$ for some $g \in L^1$. By summing up over time steps between $t$ and $t + \tau$, it comes

$$U_h(t + \tau) = \varepsilon(h, t, \tau) + \rho(h, t, \tau) U_h(t), \tag{42}$$

with

$$|\varepsilon(h, t, \tau)| \leq \int_{t-h}^{t+\tau} g(s) \ \mathrm{ds} \tag{43}$$

and $\rho(h, t, \tau)$ is in $[0, 1]$.

Let $t_0$ be such that $U^-(t_0) \leq 0$. Then, for any $\varepsilon > 0$, there exists $\eta_1$ such that

$$U(t) \leq \varepsilon \quad \forall t \in (t_0 - \eta_1, t_0).$$

Now, as $g \in L^1(I)$, from (43) there exists $\eta_2 \leq \eta_1$ such that, for any $t \in (t_0 - \eta_2, t_0)$,

$$|\varepsilon(h, t, \tau)| \leq \varepsilon \qquad \forall \tau \in (0, \eta_2] \quad \forall h \in (0, \eta_2).$$

As $U_h$ converges in $L^1(t_0 - \eta_2, t_0)$ to $U$,

$$U_h(t) \leq U(t) + \varepsilon \leq 2\varepsilon \quad \forall t \in (t_0 - \eta_2, t_0) \setminus A_h, \tag{44}$$

for some measurable set $A_h \subset (t_0 - \eta_2, t_0)$ with $|A_h| \to 0$. Using (42) to "translate" approximately inequality (44), it comes

$$U_h(t) \leq 2\rho\varepsilon + \varepsilon \quad \forall t \in (t_0, t_0 + \eta_2) \setminus A'_h \quad \forall h \in (0, \eta_2),$$

with $\left|A'_h\right| = |A_h|$. The last inequality and $L^1$ convergence of $U_h$ to $U$ (now considered in the translated interval $(t_0, t_0 + \eta_2)$) imply

$$U(t) \leq 2\rho\varepsilon + \varepsilon \qquad \text{a.e. in } (t_0, t_0 + \eta_2),$$

so that $U^+(t_0) \leq 3\varepsilon$, for any $\varepsilon > 0$. Therefore we have

$$U^+(t_0) \leq 0.$$

This completes the proof. □

## 6 Numerical experiments

### 6.1 Numerical non-uniqueness

This first set of results is somewhat distant from the original purpose of this work, but it illustrates an interesting property of Scheme $(\mathcal{S})$, which is the capability to compute multiple solutions. We consider the situation of a single material point moving on the real line, subject to the constraint $q(t) \geq 0$, with an inelastic impact law at $0^+$. The system reads :

$$u(0) = 0, \tag{45}$$

$$q(t) = \int_0^t u(s)\ \mathrm{ds} \quad \forall t \in I, \tag{46}$$

$$\dot{u}(t) = f(t) + \mu\, G(q(t)), \tag{47}$$

$$\mathrm{supp}(\mu) \subset \{t\ ,\ \ q(t) = 0\}, \tag{48}$$

$$u^+ = u^- - P_{\mathcal{N}_q} u^- \quad \forall t \in I, \tag{49}$$

where $\mathcal{N}_q$ is $\{0\}$ whenever $q > 0$, and $\mathbb{R}^-$ as soon as $q = 0$. We shall consider a force field $f$ which does not depend explicitly on $q$, and which is defined in the time interval $I = (0, 4)$ by (see Figure 2)

$$f(t) = \left| \begin{array}{l} 1 \quad \text{for } t \in \left( \dfrac{1}{2^{k+1}}, \dfrac{1}{2^{k+1}} + \dfrac{1}{2^{k+2}} \right) \\ -\alpha \text{ for } t \in \left( \dfrac{1}{2^{k+1}} + \dfrac{1}{2^{k+2}}, \dfrac{1}{2^k} \right) \end{array} \right\} k \in \mathbb{Z}\ ,\ \ k \geq -4.$$

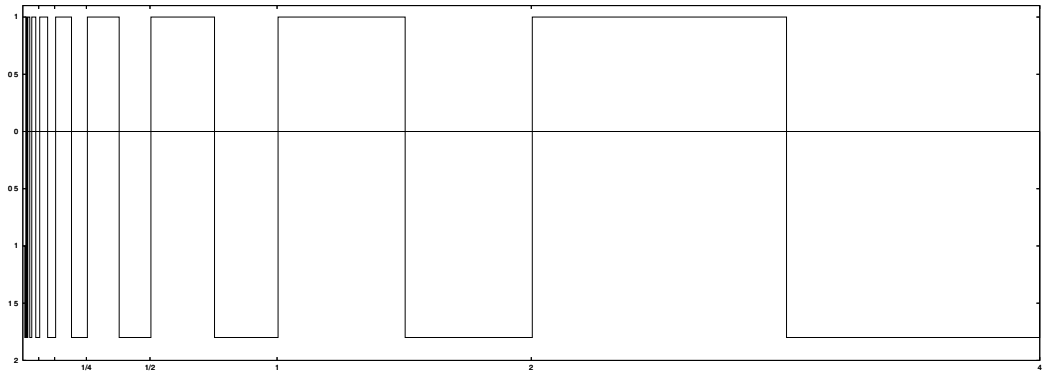


**Fig. 2** Force field $t \mapsto f(t)$

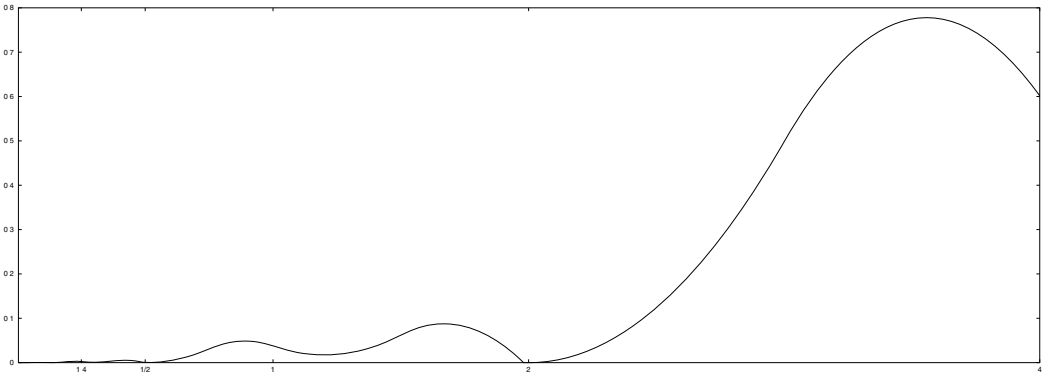


**Fig. 3** Time step $h = 2^{-8}$

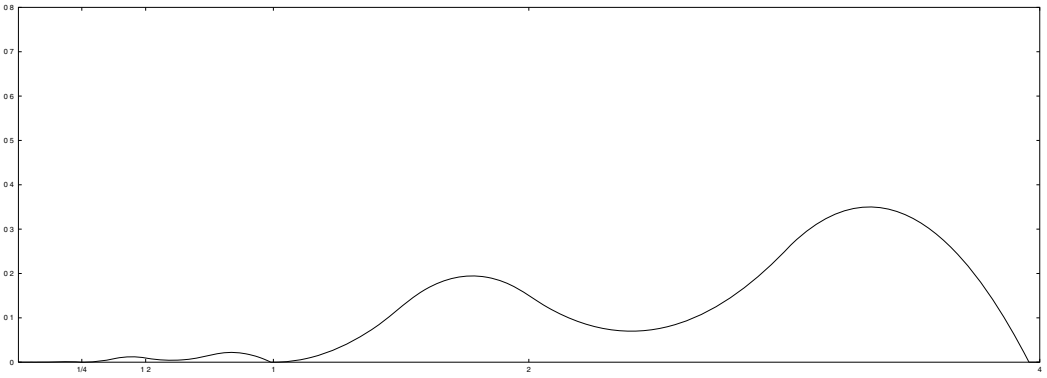


**Fig. 4** Time step $h = 2^{-9}$

We chose $\alpha = 1.8$. When the time step $h$ is a negative power of 2, the right–hand side involved in the time-stepping process

$$f_h^{n+1} = \int_{nh}^{(n+1)h} f(t)\ \mathrm{dt}$$

can be computed exactly. We present two solutions which have been computed with $h = 2^{-8}$ and $h = 2^{-9}$, respectively. Those approximations correspond to distinct piecewise polynomial functions, which can be shown to be both solutions of the problem. One can observe that for time steps of the type $2^{-2n}$, we have convergence to a solution (Figure 3), and for $h = 2^{-2(n+1)}$, we have convergence to the other

solution (which corresponds to Figure 4). Other sequences of time steps were not investigated ($f$-integral can no longer be computed exactly).

## 6.2 Sticky particles

This second set of computations illustrates the good behaviour of the scheme when the time step is large. The bodies we consider are punctual masses moving on a line, with no external force, so that when two particles meet, they collide and behave like a single particle. We consider the following situation: the initial condition is

$$\mathbf{q}_0 = (0, a, 2a, \ldots, 1-a, 1) \text{ with } a = \frac{1}{N-1},$$

the initial velocity vector $\mathbf{u}_0$ is chosen arbitrarily in $[-1, 1]^N$, and particles move according to system (8)–(12).

*Remark 12* This one-dimensional model is now commonly referred to as *sticky particle model* (although particles do not stick in a strict sense, as they could be pulled apart with infinitesimal forces). It is used in Brenier [4] to establish existence of solutions to the pressureless gas model

$$\partial_t \rho + \partial_x(\rho u) = 0,$$
$$\partial_t(\rho u) + \partial_x(\rho u^2) = 0.$$

Figures 5, 6 and 7 represent the computed pathlines of the particles in the time interval [0, 3], for a "small" time step $h = 0.01$, and two larger time steps $h = 0.5$ and $h = 1.5$. For $h = 0.01$, all collisions are distinctly captured by the scheme. For larger time steps, several collisions are taken into account by the scheme at the

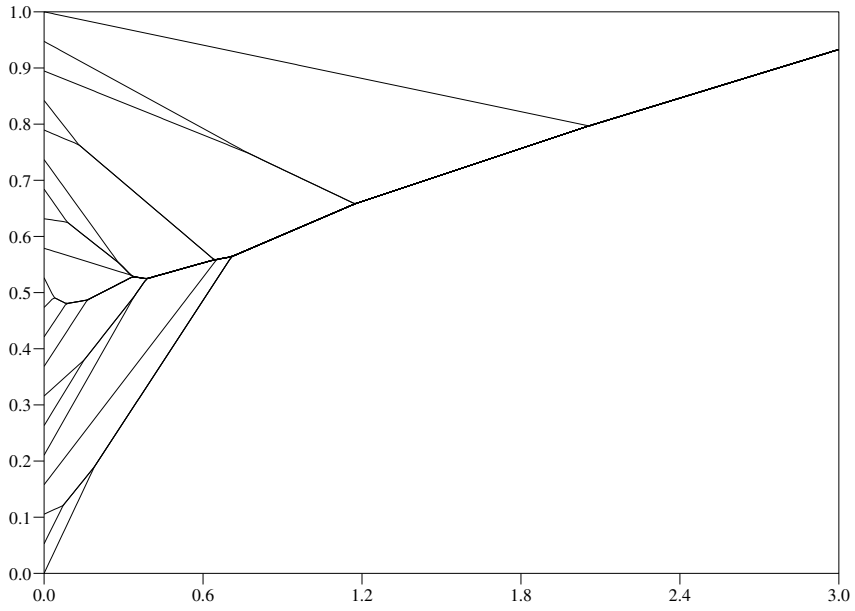


**Fig. 5** Time step $h = 0.01$

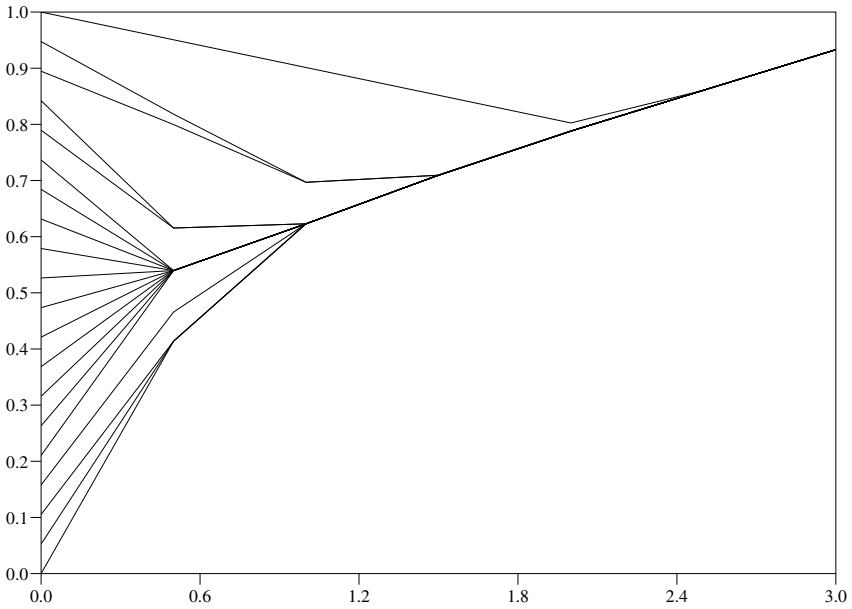


**Fig. 6** Time step $h = 0.5$

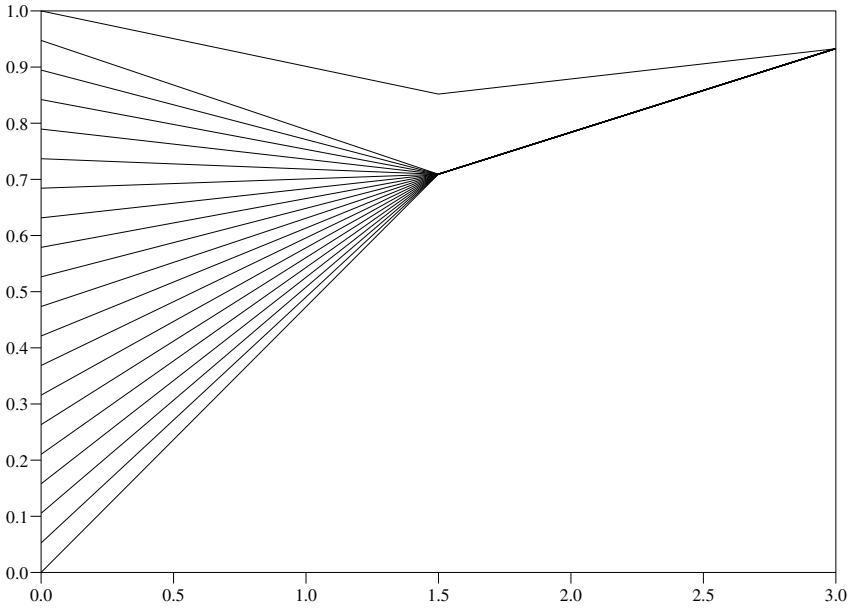


**Fig. 7** Time step $h = 1.5$

same time, so that the individual behaviour of particles is poorly described. Nevertheless, all these computations give exactly the same final state (up to errors due to the constrained minimization procedure). This behaviour is actually a direct consequence of Proposition 4 and Remark 11. Indeed, in this one-dimensional model, the final state (when all the masses are stuck together) is completely determined by the position of the aggregate and its velocity. As the center of mass moves with a constant velocity, the scheme is exact for its trajectory so that, no matter how large the time step is (as soon as all collisions have been taken into account), the approximated final state is exact.

## 6.3 Many-body computation

Numerical tests have been performed satisfactorily for large numbers of spheres (up to 50 000). We present here its application to a polydisperse collection of 2000 spheres. For readability reasons, we chose a situation such that the spheres remain

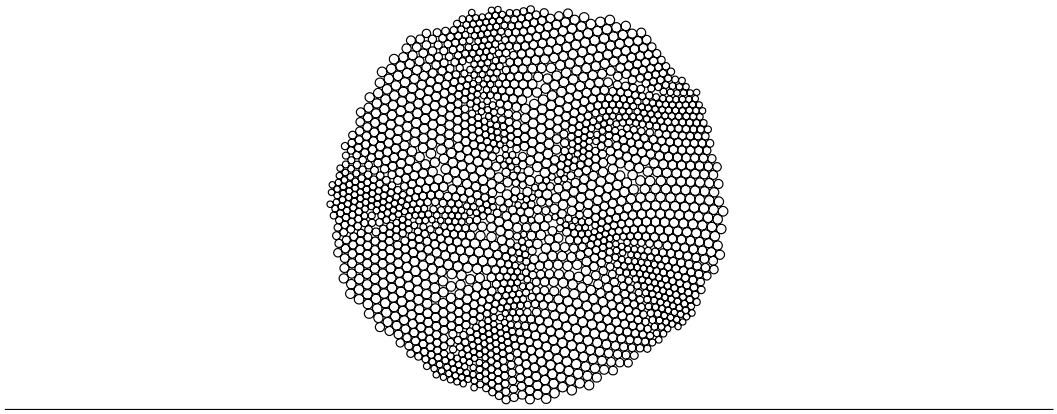

**Fig. 8** Step 0

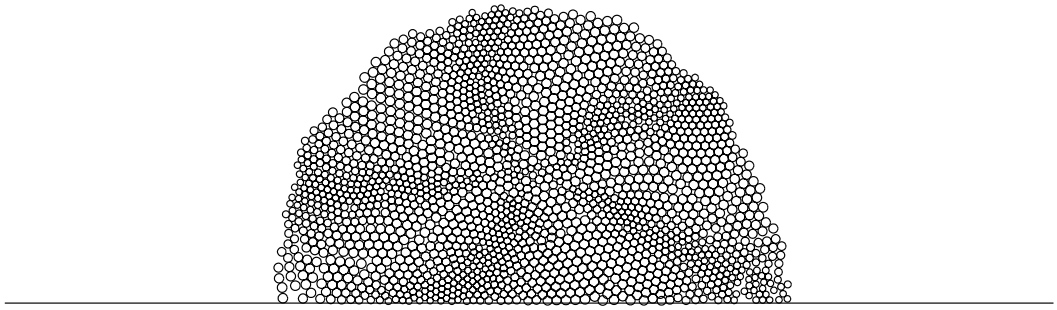

**Fig. 9** Step 3

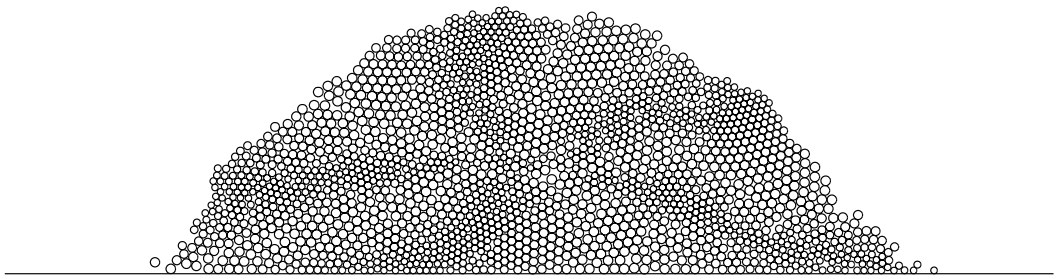

**Fig. 10** Step 6

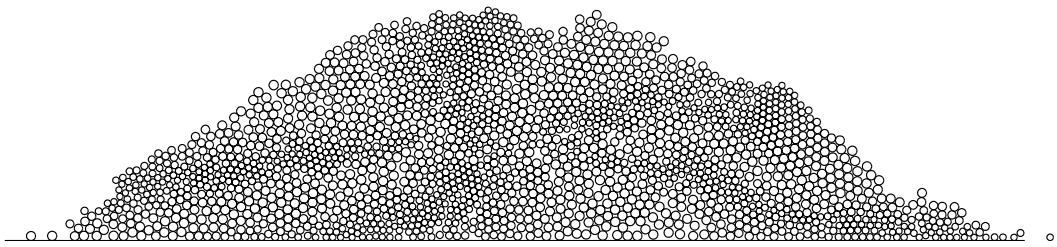

**Fig. 11** Step 9

in a fixed plane of $\mathbb{R}^3$, so that they behave like discs in $\mathbb{R}^2$, but the algorithm can be applied indifferently to genuine $3D$ problems. Radii are distributed in the interval $[0.8r_0, 1.2r_0]$, with $r_0 = 1.5 \times 10^{-2}$, and the mass $m_i$ of sphere $i$ is proportional to its volume.

As initial condition, we consider a cluster of spheres with uniform downward velocity $\mathbf{U}_0 = -\mathbf{e}_z$. We simulate the impact of this cluster with a rigid obstacle defined as the half space $\mathbb{R} \times \mathbb{R} \times \mathbb{R}_-$. To illustrate the stability of the scheme, we run a case where the time step $h = 0.1$ is such that the displacement of a sphere during one time step is typically 3 times its own size. Figures 8, 9, 10, and 11 represent the 2000 spheres at steps 0, 3, 6, and 9.

## 7 Conclusion

We presented a scheme to compute the motion of rigid bodies with nonelastic impact law. Because of its stability properties, its robustness (it produces feasible configurations only, even for large time steps), this scheme is particularly suitable to control the minimal distance between rigid particles in the context of fluid-particle simulations, with a controlled influence of the perturbation on the energy.

Beside this particular purpose, the fact that it can be used to handle situations where uniqueness does not hold, and its special behaviour in the case of simultaneous or quasi-simultaneous collisions even for large time steps, should make it an efficient tool to model more general granular flows.

## References


1. Ballard, P.: The dynamics of discrete mechanical systems with perfect unilateral constraints. Arch. Rational Mech. Anal. **154**, 199–274 (2000)
2. Bertoluzza, S., Ismail, M., Maury, B.: The fbm method: Semi-discrete scheme and some numerical experiments. Springer-Verlag, 2004
3. Bourbaki, N.: Topological vector spaces. Chapters 1–5. In: Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 1987, Translated from the French by H. G. Eggleston and S. Madan
4. Brenier, Y., Grenier, E.: Sticky particles and scalar conservation laws. SIAM J. Numer. Anal. **35**(6), 2317–2328 (1998) (electronic)
5. Ciarlet, P.G.: Introduction à l'analyse numérique matricielle et à l'optimisation. Masson, Paris, 1990
6. Coddington, E.A., Levinson, N.: Theory of ordinary differential equations. McGraw-Hill Education, Europe, 1984
7. Colombo, G., Monteiro Marques, M.D.P.: Sweeping by a continuous prox-regular set. J. Differential Equations **187**(1), 46–62 (2003)
8. Frémond, M.: Non-smooth thermomechanics. Springer-Verlag, Berlin, 2002
9. Glowinski, R.: Finite element methods for incompressible viscous flow. Handb. Numer. Anal. IX, North-Holland, Amsterdam, 2003
10. Glowinski, R., Pan, T.W.: Direct simulation of the motion of neutrally buoyant circular cylinders in plane Poiseuille flow. J. Comput. Phys. **181**(1), 260–279 (2002)
11. Haraux, A.: Nonlinear evolution equations – global behaviour of solutions. Springer-Verlag, Berlin Haidelberg New York, 1981
12. Hu, H.H.: Direct simulation of flows of solid-liquid mixtures. Int. J. of Multiphase Flow **22**(2), 335–352 (1996)
13. Johnson, A.A., Tezduyar, T.E.: Simulation of multiple spheres falling in a liquid-filled tube. Comput. Methods Appl. Mech. Engrg. **134**(3–4), 351–373 (1996)

14. Kim, S., Karrila, S.J.: Microhydrodynamics: principles and selected applications. Butterworth-Heinemann, Boston, 1991
15. Maury, B.: A many-body lubrication model. C. R. Acad. Sci. Paris Sér. I Math. **325**(9), 1053–1058 (1997)
16. Maury, B.: Direct simulations of 2D fluid-particle flows in biperiodic domains. J. Comput. Phys. **156**(2), 325–351 (1999)
17. Maury, B.: Analyse fonctionnelle, exercices et problèmes corrigés. Ellipses, Paris, 2004
18. Moreau, J.J.: Décomposition orthogonale d'un espace Hilbertien selon deux cônes mutuellement polaires. C. R. Acad. Sci. Série I **255**, 199–274 (1962)
19. Moreau, J.J.: Some numerical methods in multibody dynamics: Application to granular materials. Eur. J. Mech. A/Solids **13**, 93–114 (1994)
20. Schatzman, M.: A class of nonlinear differential equations of second order in time. Nonlinear Analysis, Theory, Methods & Applications **2**, 355–373 (1978)
21. Sigurgeirssons, H., Stuart, A., Wan, W.-L.: Algorithms for particle-field simulations with collisions. J. Comput. Phys. **172**, 766–807 (2001)
22. Stewart, D.E.: Convergence of a time-stepping scheme for rigid-body dynamics and resolution of Painlevé's problem. Arch. Rational Mech. Anal. **145**, 215–260 (1998)
23. Stewart, D.E.: Rigid-body dynamics with friction and impact. SIAM Review **42**(1), 3–39 (2000)