Dynamics of rigid bodies systems with unilateral or frictional constraints

Patrick Ballard

DYNAMICS OF RIGID BODIES SYSTEMS WITH UNILATERAL OR FRICTIONAL CONSTRAINTS

Formulation And Well-Posedness

Patrick Ballard
Laboratoire de Mécanique des Solides, Ecole Polytechnique,
91128 Palaiseau Cédex, France
ballard@lms.polytechnique.fr

Abstract The classical theory of rigid bodies systems dynamics is extended into two directions. First, systematic formulation of the dynamics of systems undergoing perfect unilateral constraints is derived. The general admissible form of the impact constitutive equation is obtained. Well-posedness of the evolution problem is proved under the assumption that the data are analytic. Second, systematic formulation of systems undergoing frictional bilateral constraints is discussed. Well-posedness of the associated evolution problem is also demonstrated.

Keywords: Analytical Dynamics, Non-smooth Mechanics, Impact, Friction

Introduction

The point of departure of any mechanical theory is a geometric description of the system under study and all its possible (or, more exactly, admissible) evolutions. This is always a schematization. Linear forms on the space of admissible (virtual) velocities define on turn the most general representation of internal and external forces which is consistent with the geometric description. Naturally, obtaining their precise expression for a given system remains a part of the modelling process. The mass distribution leads to the definition of the kinetic energy of the system which is a positive definite quadratic form on the space of velocities. Taking a time derivative, we obtain the expression of the virtual power of inertia forces (or acceleration) in any virtual velocity. The Fundamental Principle of Classical Mechanics asserts that the virtual power of inertia forces should equal the virtual power of external and internal forces

in any admissible virtual velocity. As a consequence, we derive the equation of motion. For some class of geometric descriptions, the equation of motion, associated with some initial conditions, determines completely the subsequent motion of the system. We shall say that the evolution problem associated with the dynamics is well-posed. On the opposite, there are many examples of mechanical theories in which initial conditions and equation of motion are not enough to determine the subsequent motion of the system. This is generally attributed to the excess of schematization of the geometric description. The missing physical information is added through a constitutive law. Actually, well-posedness of the resulting evolution problem serves generally implicitly as a guideline to identify the general form of the constitutive law, although some thermodynamical considerations can also play an important part.

In this paper, we are concerned with the dynamics of rigid bodies systems. Speaking of rigid bodies systems is, actually, the geometric description of the system. It could be said that this is the most simple geometric description of solids. Working in the framework of rigid bodies system means that we are not interested in the prediction of the deformation of the bodies. It does not mean that we do not consider physical situations in which bodies deformability play a role. Let us illustrate this by examining the impact of two billiard balls. Billiard balls are always deformable. But, generally we are not interested in the deformations of the balls but only on their 'global' motion. Thus, we shall use a geometric description based on the rigidity assumption. However, we know that impacts are governed by deformation wave propagation in each of the balls. So, we can not expect the simple theory based on the geometric assumption of rigidity to be able to predict the outcome of an impact experiment. We must expect that some indetermination will remain. To get well-posedness of the theory (this is necessary to make predictions which is the final aim of any mechanical theory), we are led to add to the theory an impact constitutive equation. This is nothing but injecting back in the theory the outcome of the impact, since the physical phenomena which governs the impact have been eliminated. Of course, in practical situations, we have to identify the impact constitutive equation. The choices are, either to make experiments or to use a refined theory (the elasticity theory which is based on a refined geometric assumption) in order to get the outcome of each situation of impacts. In some situations, identifying the impact constitutive equation can represent a huge amount of work. In such a case, depending on the desired precision of the predictions of the theory, one may be led to question the relevance of the simple geometric assumption that has been adopted. The use of one geometric description or another to model a given real situation is always a compromise between the desired precision of the predictions, the amount of computation which is possible and the physical informations on the system which are available.

Since in this case, no constitutive law has to be identified, the main field of application of rigid bodies dynamics has been for a long time, celestial mechanics where remarkable precision of the predictions was reached. Recently, some new fields of application of rigid bodies dynamics have emerged: robotics, granular dynamics, virtual reality,... All these fields have in common that determining the deformation in the bodies is of no interest. Nevertheless, in these applications, impacts are possible events that have to be incorporated in the theory. Very often, precision of the predictions is not so important and one may accept very approximate impact constitutive equations. Hence, the need has emerged to enrich the well-established theory of rigid bodies dynamics with the modelling of more complicated phenomena like impacts or friction, some of them relying physically on the deformation of the bodies. This new field is often called, after Jean Jacques Moreau, *Non-smooth mechanics.*

Actually, those more complicated phenomena are taken into account through constraints. A constraint is a kinematical specification of the motion with which some forces are associated: the reaction forces. In general, the kinematical specification in itself is not enough to determine the reaction force: a constitutive law of the constraint has to be added. It conveys some physical assumption on the way the constraint acts.

At the time being, it seems that only the rigid bodies dynamics with perfect holonomic bilateral constraints has firm mathematical foundations in the sense that the theory ensures the well-posedness of the evolution problem describing the dynamics. In this paper, we are concerned by the systematic formulation and well-posedness of the evolution problem describing the dynamics of systems involving more general constraints such as unilateral or frictional ones. As seen above, this program will necessarily involve the discussion of some constitutive law. Our aim will not be to try to identify any realistic one but just to characterize the general forms of constitutive laws that are compatible with the well-posedness of the theory. My opinion is that well-posedness should be considered as a requirement of any theory in classical dynamics. With this idea in mind, the discussion of well-posedness is intimately connected with the discussion of constitutive laws. Actually, we shall consider well-posedness as the final aim of the theory. After having written the Fundamental Principle of Classical Dynamics, we shall look for the supplementary hypotheses that are necessary to get well-posedness. Each time an hypothesis will be made, we shall try to motivate it by a counter-example. These hypotheses will be classified into two categories. Those which convey physical assumption will be called 'constitutive' hypotheses and the other one whose aim is to prevent from mathematical pathologies will be called 'regularity' hypotheses. Since one aim is to obtain general forms of constitutive laws, one has to make sure that the constitutive laws do not depend on any particular parametrization of the system. For this reason, we are going to try to obtain intrinsic formulations of

dynamics, that is, formulations which do not rely on a particular choice for the parametrization of the system. This necessarily requires the use of the language of differential geometry. But, only the most elementary level of differential geometry is required.

The major enhancement of mathematical consistency which seems to be desired at the time being concerns the modelling of impacts and that of friction. These two subjects are the major concerns in this paper and I believe that a mathematically satisfactory theory is obtained on both points-of-view of general formulation as well as well-posedness. However, the task is far from being achieved. In this paper, we examine the cases of impacts and friction *separately*. There remains to mix the two theories to discuss, for example, frictional unilateral constraints, which is not done here. The result would be a general theory of the evolution of mechanisms consisting of rigid bodies.

Section 1 recalls briefly the basics of intrinsic formulation and well-posedness of the dynamics of rigid bodies systems. The aim of this section is to provide precise description of the framework and notations. Section 2 contains also only well-known material. It shows that superimposing perfect holonomic bilateral constraints does not modify the structure of the theory. In Section 3, perfect unilateral constraints are discussed. The general form for the impact constitutive equation is provided and the general formulation for the evolution problem is derived. Well-posedness is fully discussed. In Section 4, the case of general perfect non-holonomic bilateral constraints is examined. Actually, this type of constraint is a particular case of non-firm constraints which are the concern of Section 5. A complete theory of non-firm constraints is derived, including systematic formulation and well-posedness. In Section 6, the formalism of non-firm constraints is applied to the description of frictional bilateral constraints. The underlying idea is that friction should be considered as a dissipation mechanism obeying the Principle of Maximal Dissipation. In some cases (for example, systems of punctual particles), we recover standard dry friction laws such as Coulomb friction and, in some cases, we do not. Section 7 provides a brief description of the situations that are not contained in the above theories and the extensions of the content of the paper that could be done later on.

1. The dynamics of rigid bodies systems

1.1 The geometric assumption: rigidity

Classical mechanics postulates the existence of a three-dimensional oriented affine Euclidean space $\mathcal{E}$, sometimes called the (Galilean) *real world*, and an absolute chronology represented (after the choice of an origin) by a real number, generally denoted by t. The vector space associated with $\mathcal{E}$ will be denoted by E.

A solid is represented by its *real world reference configuration* which is nothing but a possible geometric locus of all the material points of the solid in $\mathcal{E}$. The geometric assumption of rigidity can be stated as follows: the only real world configuration of that solid which can be observed are obtained from the real world reference configuration by *direct isometries.* Therefore, once the real world reference configuration has been fixed, any real world conguration of the solid is represented by a direct isometry q. Considering a material point of the solid identified by its location $M \in \mathcal{E}$ in the real world reference configuration, the current location of that material point in the configuration defined by q is:

$$m(M, q) = q(M). \tag{1.1}$$

Since any direct isometry on $\mathcal{E}$ can be split into a translation and a rotation, the set of all direct isometries can be identified to $E \times \mathbb{SO}3$ (where $\mathbb{SO}3$ denotes the set of all direct orthogonal endomorphisms on E, endowed with its standard manifold structure). It is said that $E \times \mathbb{SO}3$ is the (abstract) *configuration manifold* of the rigid solid. Since its dimension is 6, we say that the rigid solid has 6 *degrees of freedom* (dof). Any (local) chart on the configuration manifold is called a *(local) parametrization.* The configuration manifold is generally denoted by Q and a configuration (an element of the configuration manifold), by q. A local chart (parametrization) will be denoted generally by ψ. Thus, for a rigid solid, the symbol $\psi(q)$ denotes an element of $\mathbb{R}^6$.

Other idealizations of rigid solids can appear: the infinitely thin rigid bar whose configuration manifold is $E \times \mathbb{S}2$ ($\mathbb{S}2$ denotes the two-dimensional sphere equipped with its standard manifold structure) and the punctual particle whose configuration manifold is simply E.

A motion of a rigid solid is a curve on its configuration manifold (a mapping from a time interval I into Q). The derivative of the motion at instant t is denoted by $\dot{q}(t)$. It is called the (abstract or sometimes, generalized) velocity. It is an element of the tangent bundle TQ of the configuration manifold. One often encounters the name 'state space' for TQ, in which case $\dot{q}(t)$ is also called a state of the system. Since the mapping m defined by formula (1.1) is obviously smooth, the material velocities are expressed in terms of the (abstract) velocity by:

$$\dot{m} = \partial_q m(M, q) \cdot \dot{q}, \tag{1.2}$$

where $\partial_q m(M, q)$ is a linear operator from the tangent space $T_q Q$ into $T_m \mathcal{E} = E$.

The *mass distribution* in the rigid solid is specified on the real world reference configuration. It is a *bounded positive measure on* $\mathcal{E}$. It is denoted by μ. Considering an arbitrary motion $(I, q(t))$ of the rigid solid, the *kinetic energy* K at instant t is by definition:

$$K = \frac{1}{2} \int_{\mathcal{E}} \|\dot{m}\|_E^2 \, \mathrm{d}\mu(M). \tag{1.3}$$

Combining formulae (1.2) and (1.3), we obtain easily the expression of the kinetic energy in terms of the (abstract) velocity. Then, it is easily noticed that the kinetic energy defines a nonnegative quadratic form on each tangent space T_qQ of the configuration manifold. The mass distribution is said to be consistent with the geometric description if this quadratic form is positive definite. The following are easily proved:

- A mass distribution μ in the three-dimensional solid $E \times \mathbb{SO}3$ is consistent *if and only if* its support Supp μ contains at least three non-aligned points.
- A mass distribution μ in the infinitely thin bar $E \times \mathbb{S}2$ is consistent *if and only if* Supp μ contains at least two distinct points.
- A mass distribution μ in the punctual particle E is consistent *if and only if* Supp μ is non-void.

>From now on, we shall assume that the mass distribution is always consistent with the geometric description. As a result, the kinetic energy defines a scalar product on each tangent space of Q, endowing the configuration manifold with a Riemannian structure. This Riemannian metric is naturally called the *kinetic metric*. From now on, whenever we speak of a configuration manifold, it will always be supposed to be equipped with its Riemannian structure.

A rigid bodies system is a finite collection of rigid bodies. The configuration manifold of a rigid bodies system is the cross-product $Q_1 \times Q_2 \times \cdots \times Q_n$ of the individual configuration manifold Q_i of each rigid body of the system.

The fundamental idea which is behind these definitions is that the configuration manifold conveys all the necessary information on the system and no more. For example, we should keep aware that the kinetic metric conveys all the relevant information about the mass distribution but, one can not, generally, recover the mass distribution from the kinetic metric.

Remark 1. The reader who is not familiar with elementary differential geometry could have the feeling that we have expressed very simple (and well known) things in a complicated way. Such a reader would probably prefer a presentation where the parametrization of the system is introduced at first and each definition (the abstract configuration, the kinetic metric, ...) is made in terms of real matrices. Such a presentation should then precise what are the effects on these matrices of a change of parametrization. This leads to heavy and boring formulae and is often left aside, but this is not the main reason why I have chosen the above presentation. The possibility of defining every concept without any reference to a given parametrization ensures that all what has been defined is *intrinsic* (that is, does not depend on the particular parametrization under consideration). This fact is particularly crucial when one deals with constitutive equations and introducing constraints necessarily involves constitutive

equations. In the end, I believe that the intrinsic presentation, making apparent the structure of the theory, provides deeper understanding. However, the reader who feels more comfortable with it, might consider that the configuration manifold Q is an open subset of $\mathbb{R}^d$ equipped with a 'variable' symmetric positive definite matrix $(g_{ij}(q))$, which is nothing but considering a particular parametrization of the system. The following convention notations are made on that purpose.

Notations. For Q being a smooth Riemannian manifold of dimension d, we shall denote by:

- TQ and T^*Q, the tangent and cotangent bundles,
- Π_Q and Π_Q^*, the natural projection mappings of TQ and T^*Q,
- $\langle \cdot, \cdot \rangle_q$, the local duality product between tangent space T_qQ and cotangent space T_q^*Q,
- $(\cdot, \cdot)_q$ and $\|\cdot\|_q$, the local scalar product and norm on T_qQ (a * will be added when referring to the scalar product and norm on T^*Q),
- $\flat$ (and $\sharp = \flat^{-1}$, its inverse), the isomorphism of vector bundles from TQ onto T^*Q naturally associated with the Riemannian metric of Q.

For $q(t)$ being a curve on Q, we have decided above to denote the derivative at t by $\dot{q}(t) \in TQ$. In order to be consistent with the suggestion made in remark 1; we shall alternatively use the notation $(q(t), \dot{q}(t))$ as often as it will not be too heavy or confusing. This is clearly a redundant notation since the base-point $q = \Pi_Q(\dot{q})$ is contained in the derivative, but I believe that this notation may help the understanding. More generally, an element v of TQ will also be denoted by (q, v) with q being the base-point of v. For ψ being a local chart on Q, $\psi(q)$ is an element of $\mathbb{R}^d$ that we denote by $(q^1, q^2, \ldots, q^d)$. Still to be consistent with the suggestion of remark 1, we shall sometimes keep the notation q to refer to $\psi(q)$. Thus, for q being an abstract configuration, we might write $q = (q^1, q^2, \ldots, q^d)$. More generally, each time it will not be confusing, we will keep the same notation for an object and its representative in a chart. As usual, the natural basis of T_qQ (resp. T_q^*Q) naturally associated with the chart ψ is denoted by $(e_1(q), e_2(q), \ldots, e_d(q))$ (resp. $(e^1(q), e^2(q), \ldots, e^d(q))$). For (q, v) belonging to TQ, we denote by v^i $(i = 1, 2, \ldots d)$ its components in the natural basis and we shall write:

$$v = v^i e_i(q).$$

Einstein's summation convention will always apply unless explicitly stated. For $q(t)$ being a curve, we shall write:

$$\dot{q}(t) = \dot{q}^i(t) e_i(q(t)),$$

and $\dot{q}^i(t)$ is the derivative at time t of the real-valued function $q^i(t)$. As usual, $g_{ij}(q)$ will be the covariant components of the metric in the considered chart and $g^{ij}(q)$ its contravariant components; $\Gamma^i_{jk}(q)$ will be the associated Christoffel symbols:

$$\Gamma^i_{jk}(q) = \frac{1}{2} g^{ih}(q) \left(\frac{\partial g_{hk}}{\partial q^j}(q) + \frac{\partial g_{jh}}{\partial q^k}(q) - \frac{\partial g_{jk}}{\partial q^h}(q) \right).$$

For $q(t)$ being a curve on Q and v a vector field on that curve, the covariant derivative of v along $q(t)$ is denoted by:

$$\frac{\mathrm{D}}{\mathrm{d}t} v(t) = \left(\frac{\mathrm{d}}{\mathrm{d}t} v^i(t) + \Gamma^i_{jk}(q(t)) v^j(t) \dot{q}^k(t) \right) e_i(q(t)).$$

1.2 Formulation of the dynamics

Consider a rigid bodies system of configuration manifold Q and a motion $q(t)$ of that system. The *power of inertial forces* at instant t is, by definition, the time derivative at t of the kinetic energy:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} K(q, \dot{q}) &= \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \left(\dot{q}(t), \dot{q}(t) \right)_{q(t)}, \\ &= \left(\frac{\mathrm{D}}{\mathrm{d}t} \dot{q}(t), \dot{q}(t) \right)_{q(t)}, \\ &= \left\langle \flat \frac{\mathrm{D}}{\mathrm{d}t} \dot{q}(t), \dot{q}(t) \right\rangle_{q(t)}. \end{aligned}$$

Hence, it is seen that the power of inertial forces at time t defines the cotangent vector $\flat \mathrm{D}\dot{q}(t)/\mathrm{d}t \in T^*_{q(t)}Q$. An arbitrary element T_qQ is often called a *virtual velocity* of the system in the configuration q. Then, the linear form $\flat \mathrm{D}\dot{q}(t)/\mathrm{d}t$ is called virtual power of inertial forces.

The analysis of the dynamics has to take into account external and internal forces. They are usually given as a force distribution on the current real world configuration. This is an E-valued measure which may depend on the current state $(q, \dot{q})$ and on time t. We shall denote it by $\phi(q, \dot{q}; t)$ The power of the internal and external forces at time t in the motion $q(t)$ is:

$$\begin{aligned} &\int_{\mathcal{E}} \left(\dot{m}, \mathrm{d}\phi(q, \dot{q}; t)(m(M, q)) \right)_E \\ &\qquad = \int_{\mathcal{E}} \left(\partial_q m(M, q) \cdot \dot{q}, \mathrm{d}\phi(q, \dot{q}; t)(m(M, q)) \right)_E, \end{aligned}$$

which also defines a linear form $f(q, \dot{q}; t)$ on T_qQ by:

$$\langle f(q, \dot{q}; t), v \rangle_q \stackrel{\text{def}}{=} \int_{\mathcal{E}} \left(\partial_q m(M, q) \cdot v, d\phi(q, \dot{q}; t)(m(M, q)) \right)_E,$$

for any virtual velocity $v \in T_qQ$. This linear form $f(q, \dot{q}, t) \in T_q^*Q$ is called *virtual power of external and internal forces.* The reason for such a modelling of forces by duality is that it ensures the consistency of the forces modelling with the geometrical description of the system. The virtual power mapping $f(q, \dot{q}, t)$ extracts from the force field ϕ only the information which is relevant to the dynamics analysis in the framework of the geometrical assumption of rigidity.

The fundamental principle of classical mechanics asserts that the virtual power of inertial forces should equal at every instant the virtual power of external and internal forces:

$$\forall t, \quad \flat \frac{\mathrm{D}}{\mathrm{d}t}\dot{q}(t) = f(q(t), \dot{q}(t), t). \tag{1.4}$$

Equation (1.4) is referred to as the *equation of motion.* It is a second-order differential equation on the configuration manifold. To express it in a particular parametrization of the system, the following is useful.

Proposition 1 (Lagrange) *Let ψ be a local chart and $q(t)$ a C^2 motion on Q. One has:*

$$\flat \frac{\mathrm{D}}{\mathrm{d}t}\dot{q}(t) = \left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial}{\partial \dot{q}^i} K(q(t), \dot{q}(t)) - \frac{\partial}{\partial q^i} K(q(t), \dot{q}(t)) \right) e^i(q(t)).$$

Proof. It is straightforward since:

$$\begin{aligned} \flat \frac{\mathrm{D}}{\mathrm{d}t}\dot{q} &= g_{ij}\left(\frac{\mathrm{d}}{\mathrm{d}t}\dot{q}^j + \Gamma^j_{kl}\dot{q}^k\dot{q}^l \right) e^i, \\ &= g_{ij}\left(\frac{\mathrm{d}}{\mathrm{d}t}\dot{q}^j + \frac{1}{2}g^{jh}\left(\frac{\partial g_{hl}}{\partial q^k} + \frac{\partial g_{hk}}{\partial q^l} - \frac{\partial g_{kl}}{\partial q^h} \right)\dot{q}^k\dot{q}^l \right) e^i, \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial}{\partial \dot{q}^i}\left(\frac{1}{2}\dot{q}^j g_{jk}\dot{q}^k \right) - \frac{\partial}{\partial q^i}\left(\frac{1}{2}\dot{q}^j g_{jk}\dot{q}^k \right) \right) e^i. \end{aligned}$$

□

We are given an initial instant t_0 and an initial state $(q_0, v_0) \in TQ$. Then, the evolution problem associated with the dynamics of rigid bodies system is the Cauchy problem:

Problem I. Find $T > t_0$ and $q \in C^2([t_0, T[; Q)$ such that:

- $(q(t_0), \dot{q}(t_0)) = (q_0, v_0)$,
- $\forall t \in [t_0, T[, \quad \flat \frac{\mathrm{D}}{\mathrm{d}t}\dot{q}(t) = f(q(t), \dot{q}(t), t).$

1.3 Well-posedness of the dynamics

To study the well-posedness (existence and uniqueness of solution) of problem I, we have to specify regularity assumptions on Q and f.

Counter-example 1. Consider the evolution equation

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}q(t) = 6\,|q(t)|^{\frac{1}{3}}$$

($q \in \mathbb{R}$) with initial condition $(q(0), \dot{q}(0)) = (0,0)$. It is readily checked that the two motions defined on $\mathbb{R}^+$ $q(t) = 0$ and $q(t) = t^3$ provide two distinct solutions.

To get well-posedness, we have to make further hypotheses. Throughout this paper, we shall distinguish two classes of hypotheses: the *constitutive* hypotheses and the *regularity* hypotheses. A constitutive hypothesis is an hypothesis which conveys physical meaning. A regularity hypothesis conveys no physical meaning and is stated to eliminate mathematical pathologies. The following regularity hypothesis is slightly stronger than necessary.

Regularity hypothesis. The Riemannian configuration manifold is of class C^2 and the mapping $f : TQ \times \mathbb{R} \to T^*Q$ is of class C^1.

It should be pointed out that the first part of this hypothesis is actually no hypothesis at all. The configuration manifold of the three-dimensional rigid solid, of the infinitely thin rigid bar or of the punctual particle, with arbitrary consistent mass distribution are C^∞ (or, even more, analytic) Riemannian manifolds. The configuration manifold of a rigid bodies system (with no constraint), being a cross-product of such manifolds, can be assumed to have arbitrarily regularity. This is a restriction neither on the geometry nor on the mass distribution of the system, but on the class of admissible parametrizations.

Under this regularity hypothesis, we have the following well-posedness result.

Theorem 2 (Cauchy) *There exists a unique maximal solution for problem I.*

More precisely, theorem 2 states that there exists $T_m > t_0$ ($T_m \in \mathbb{R} \cup \{+\infty\}$) and $q_m \in C^2([t_0, T_m[, Q)$ being a solution of problem I such that any other solution of problem I is a restriction of q_m. Of course, we expect that $T_m = +\infty$, in which case the dynamics is said to be *eternal*. This situation can not be taken for granted, in general.

Counter-example 2. Consider the evolution equation

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}q(t) = (\dot{q}(t))^2$$

($q \in \mathbb{R}$) with initial condition $(q(0), \dot{q}(0)) = (0, 1)$. It is readily checked that the maximal solution is defined on the interval $[0, 1[$.

In the usual cases encountered in mechanics, eternal dynamics is ensured by the following general sufficient condition.

Theorem 3 *The configuration manifold Q is assumed to be a complete Riemannian manifold (this is no hypothesis in the case of rigid bodies system with no constraints). The mapping f is supposed to admit the following estimate:*

$$\forall\, (q, v) \in TQ, \quad \textit{for almost all } t \in [t_0, +\infty[, \qquad \|f(q, v; t)\|_q^* \leq l(t) \left(1 + d(q, q_0) + \|v\|_q\right),$$

where $d(\cdot, \cdot)$ is the Riemannian distance and $l(t)$, a (necessarily nonnegative) function of $L^1_{loc}(\mathbb{R}; \mathbb{R})$.

Then, the dynamics is eternal: $T_m = +\infty$.

The proof of theorem 3 relies on the Gronwall-Bellman lemma which is now recalled.

Lemma 4 (Gronwall-Bellman) *Let $m_1 \in BV([t_0, T]; \mathbb{R})$ and $m_2 \in L^1(t_0, T; \mathbb{R})$ be two functions such that:*

$$\textit{for almost all } t \in]t_0, T[, \quad m_2(t) \geq 0.$$

Let $\phi \in BV([t_0, T]; \mathbb{R})$ be such that:

$$\forall\, t \in [t_0, T], \quad \phi(t) \leq m_1(t) + \int_{t_0}^{t} m_2(s)\phi(s)\, \mathrm{d}s.$$

Then,

$$\forall\, t \in [t_0, T], \quad \phi(t) \leq m_1(t) + \int_{t_0}^{t} m_1(s)m_2(s)e^{\int_s^t m_2(\sigma)\, \mathrm{d}\sigma}\, \mathrm{d}s.$$

Lemma 5 *Let m be in $L^1(t_0, T; \mathbb{R})$ such that $m(t) \geq 0$ for almost all t in $]t_0, T[$ and a be a real nonnegative constant. Consider $\phi \in BV([t_0, T]; \mathbb{R})$ such that:*

$$\forall\, t \in [t_0, T], \quad \frac{1}{2}\phi^2(t) \leq \frac{1}{2}a^2 + \int_{t_0}^{t} m(s)\phi(s)\, \mathrm{d}s,$$

then:

$$\forall\, t \in [t_0, T], \quad |\phi(t)| \leq a + \int_{t_0}^{t} m(s)\, \mathrm{d}s.$$

Elementary proofs of lemmas 4 and 5 can be found in BREZIS (1973), p. 156.

Proof of theorem 3. Suppose T_m is finite. From the equation of motion (1.4), we have, for all $t \in [t_0, T_m[$,

$$\frac{1}{2}\|\dot{q}_m(t)\|^2_{q_m(t)} - \frac{1}{2}\|v_0\|^2_{q_0} \leq \int_{t_0}^{t} \langle f(q_m(s), \dot{q}_m(s); s), \dot{q}_m(s)\rangle_{q_m(s)}\, \mathrm{d}s,$$

$$\leq \int_{t_0}^{t} l(s)\left(1 + d(q_m(s), q_0) + \|\dot{q}_m(s)\|_{q_m(s)}\right)\|\dot{q}_m(s)\|_{q_m(s)}\, \mathrm{d}s,$$

which gives, by lemma 5,

$$\|\dot{q}_m(t)\|_{q_m(t)} \leq \|v_0\|_{q_0} + \int_{t_0}^{t} l(s)\left(1 + d(q_m(s), q_0) + \|\dot{q}_m(s)\|_{q_m(s)}\right)\, \mathrm{d}s.$$

But, by definition of the Riemannian distance,

$$\forall\, t \in [t_0, T_m[, \quad d(q_m(t), q_0) \leq \int_{t_0}^{t} \|\dot{q}_m(s)\|_{q_m(s)}\, \mathrm{d}s,$$

therefore,

$$\forall\, t \in [t_0, T_m[, \quad d(q_m(t), q_0) + \|\dot{q}_m(t)\|_{q_m(t)} \leq$$

$$\|v_0\|_{q_0} + \int_{t_0}^{t} l(s)\, ds + \int_{t_0}^{t} (1 + l(s))\left(d(q_m(s), q_0) + \|\dot{q}_m(s)\|_{q_m(s)}\right)\, \mathrm{d}s.$$

By lemma 4, one gets:

$$d(q_m(t), q_0) + \|\dot{q}_m(t)\|_{q_m(t)} \leq \left(\|v_0\|_{q_0} + \int_{t_0}^{t} l(s)\, ds\right) e^{\int_{t_0}^{t}(1+l(s))\, \mathrm{d}s},$$

which shows that the function $t \mapsto \|\dot{q}_m(t)\|_{q(t)}$ is bounded over $[t_0, T_m[$. By the completeness of Q, we deduce that

$$q_T = \lim_{t \to T_m^-} q_m(t)$$

exists in Q. Then, it is an easy matter to deduce that

$$(q_T, v_T) = \lim_{t \to T_m^-} (q_m(t), \dot{q}_m(t)) \quad \text{exists in } TQ,$$

and that the function q_m, extended by continuity at T_m satisfies the equation of motion on $[t_0, T_m]$. Then, theorem 2 furnishes $T'_m > T_m$ and an extension of q_m, belonging to $C^2([t_0, T'_m[; Q)$ and being a solution of problem I. But, this contradicts the definition of T_m. □

2. Perfect holonomic bilateral constraints

A constraint describes a type of forces which are not taken into account by the forces mapping f. Indeed, it is possible to specify (partially) some forces by their kinematical effects. These kinematical effects leave in general the associated forces partially undetermined and we have to add phenomenological assumptions on the way the constraint acts, through a *constitutive law* of the constraint.

2.1 The geometric description

A *holonomic bilateral constraint* is a restriction on the admissible motions of the system which is expressed by means of a finite number n of smooth real-valued functions φ_i defined on the configuration manifold Q:

$$\forall\, i \in \{1, 2, \cdots, n\}, \quad \varphi_i(q) = 0. \tag{1.5}$$

The word constraint in the singular will be used indifferently to speak either of a constraint specifically associated with a single function φ_i or of the constraint associated with all the functions φ_i. In this terminology, a finite collection of constraints is still a constraint. We denote by S the set of all admissible configurations:

$$S = \big\{ q \in Q \,;\, \forall\, i \in \{1, 2, \cdots, n\}, \quad \varphi_i(q) = 0 \big\}.$$

The following hypothesis is usual in this framework.

Regularity hypothesis I. The functions φ_i are *functionally independent*, that is, for all $q \in S$, the $\mathrm{d}\varphi_i(q)$ ($i \in \{1, 2, \cdots, n\}$) are linearly independent in T^*Q.

A straightforward consequence of this hypothesis is that S is a submanifold of Q of dimension $d - n$. As a result, S inherits a Riemannian structure from Q. We shall say that S is the configuration manifold of the constrained system.

2.2 Formulation of the dynamics

The realization of the constraint (1.5) necessarily involves a modification of the equation of motion (1.4). This is done by adding to the virtual power of forces $f(q, \dot{q}; t)$ a corrective unknown term R called the *virtual power of reaction forces*:

$$\forall\, t, \quad \flat \frac{\mathrm{D}}{\mathrm{d}t}\dot{q}(t) = f(q(t), \dot{q}(t), t) + R(t).$$

We might expect R to be determined by the geometric constraint (1.5). It does not work in general. We have to add phenomenological assumptions on the

way the constraint acts. This is the *constitutive law* of the constraint. At this point, we restrict ourselves to the following.

Constitutive hypothesis II. The holonomic bilateral constraint (1.5) is supposed to be *perfect* (one also says synonymousl y *ideal*), that is, the virtual power of the reaction forces R vanishes in any virtual velocity compatible with the bilateral constraint:

$$\forall v \in \left\{ v \in T_q Q \; ; \; \forall i \in \{1, 2, \cdots, n\}, \quad \langle \mathrm{d}\varphi_i(q), v \rangle_q = 0 \right\} \simeq TS, \qquad \langle R, v \rangle_q = 0.$$

Hypotheses I and II imply that there exists n real-valued functions λ_i, unique, such that:

$$R(t) = \sum_{i=1}^{n} \lambda_i(t) \, \mathrm{d}\varphi_i(q).$$

Now, we formulate the evolution problem associated with the dynamics of rigid bodies systems with perfect bilateral constraints. The initial condition is assumed to be compatible with the realization of the constraint: $(q_0, v_0) \in TS$.

Problem II. Find $T > t_0$, $q \in C^2([t_0, T[; Q)$ and n functions $\lambda_i \in C^0([t_0, T[; \mathbb{R})$ such that:

- $(q(t_0), \dot{q}(t_0)) = (q_0, v_0)$,
- $\forall t \in [t_0, T[, \quad q(t) \in S$,
- $\forall t \in [t_0, T[, \quad \flat \dfrac{\mathrm{D}_Q}{\mathrm{d}t} \dot{q}(t) = f(q(t), \dot{q}(t), t) + \displaystyle\sum_{i=1}^{n} \lambda_i(t) \, \mathrm{d}\varphi_i(q(t))$.

Here, we used the notation $\mathrm{D}_Q/\mathrm{d}t$ for the covariant derivative to underline the fact the covariant derivative is understood with respect to the Riemannian structure of Q (and not to that of S).

Let q be a point of Q, v a vector in $T_q Q$, and E a subspace of $T_q Q$. The orthogonal projection of v on E for the scalar product of $T_q Q$ induced by the Riemannian structure of Q is denoted by $\mathrm{Proj}_q [v; E]$. Similarly, $\mathrm{Proj}_q^* [v^*; E^*]$ denotes the orthogonal projection of the cotangent vector v^* on the subspace E^* of $T_q^* Q$. If $q(t)$ is a curve on the Riemannian submanifold S of Q and v a vector field on that curve, then we have (CHAVEL (1993), p. 54):

$$\frac{\mathrm{D}_S v}{\mathrm{d}t} = \mathrm{Proj}_q \left[\frac{\mathrm{D}_Q v}{\mathrm{d}t}; T_q S \right].$$

Therefore, any solution of problem II is seen to be a solution of

Problem II′. Find $T > t_0$ and $q \in C2([t_0, T[; S)$ such that:

- $(q(t_0), \dot{q}(t_0)) = (q_0, v_0)$,
- $\forall t \in [t_0, T[, \quad \flat \dfrac{\mathrm{D}_S}{\mathrm{d}t}\dot{q}(t) = \mathrm{Proj}^*_{q(t)} \left[f(q(t), \dot{q}(t); t); T^*_{q(t)} S \right]$.

Reciprocally, any solution of problem II′ is readily seen to generate a solution of problem II: the two evolution problems are equivalent.

The linear form (cotangent vector) $\mathrm{Proj}^*_q \left[f(q, \dot{q}; t); T^*_q S \right]$ equals the restriction of the linear form $f(q, \dot{q}; t)$ on the space $T_q S$ of virtual velocities compatible with the bilateral constraint. Therefore, it is the virtual power of external and internal forces in any virtual velocity compatible with the constraint.

2.3 Well-posedness of the dynamics

Problem II′ has formally the same structure of problem I. Since problems II′ and II are equivalent, the results of Section 1(1.3) give the well-posedness of the dynamics of rigid bodies systems with perfect bilateral constraints.

Regularity hypothesis III. The configuration manifold Q and the functions φ_i are of class C^2 and the mapping $f : TQ \times \mathbb{R} \to T^*Q$ is of class C^1.

Proposition 6 *Problems II and II′ have a unique maximal solution q_m. Moreover, if Q and the functions φ_i are of class C^p ($p \geq 2$), and f of class C^{p-1} then q_m is of class C^p. If Q, f and the φ_i are analytic functions then so is q_m.*

The second part of proposition 6 follows from standard results on ordinary differential equations (see, for example, CODDINGTON & LEVINSON (1955)).

The analysis of the eternity of the dynamics is provided by theorem 3.

The regularity hypothesis I could seem very restrictive. However, dropping it would make us run into troubles.

Counter-example 3. Consider a rigid homogeneous bar of length l. The two extremities of the bar are constrained to remain on a fixed circle of diameter l. The two corresponding bilateral constraints are supposed to be perfect. This is a simple occurrence of bilateral constraint which does not satisfy hypothesis I. At initial instant, the bar is at rest. A constant force is applied at the middle point of the bar. This force is directed in the plane of the circle but not along the bar. The reader will convince himself that the corresponding evolution problem II admits no solution.

2.4 Illustrations and comments

The configuration manifold Q of the rigid body system with no constraint is often referred to as the *primitive* configuration manifold, whereas the submanifold S is called the *reduced* configuration manifold. In practice, the reduced configuration manifold can be often constructed directly, without introducing first a primitive configuration manifold. In such a case, the forces mapping is directly introduced with respect to the reduced configuration manifold.

Example 4. Consider a plane system of two homogeneous rigid bars 1 and 2. The bar 1, of length l_1 and mass m_1 is connected to a fixed support by means of a perfect ball-and-socket joint equipped with a spiral spring of stiffness k_1. The bar 2, of length l_2 and mass m_2 is connected to the free extremity of the bar 1 by means of another ball-and-socket joint also equipped with a spiral spring of stiffness k_2. A force acts on the free extremity of the bar 2. This force remains parallel to the direction of the bar 2 and is of constant magnitude $\lambda > 0$ (see Figure 1.1).

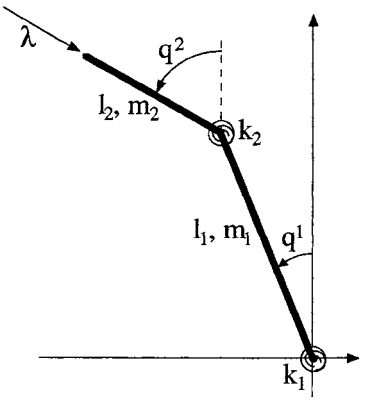

Figure 1.1. Geometry of the double pendulum.

- The configuration space is $\mathbb{R}^2$ equipped with its canonical structure of C^∞ manifold (it is not the 2-torus since the spiral springs impose to be able to count the 'number of turns'). This manifold may be represented by a single chart; in other terms, there exists a global parametrization of the system. In the sequel, we shall only use the chart (q^1, q^2) defined by the angular measures associated with each of the joints.

- The kinetic energy is:

$$\begin{aligned} K &= \frac{1}{2}\int_0^{l_1} \frac{m_1}{l_1} s^2 \left(\dot{q}^1\right)^2 \mathrm{d}s + \frac{1}{2}\int_0^{l_2} \frac{m_2}{l_2}\Big(l_1^2\left(\dot{q}^1\right)^2 \\ &\qquad + s^2\left(\dot{q}^2\right)^2 + 2l_1 s\cos\left(q^1-q^2\right)\dot{q}^1\dot{q}^2\Big)\,\mathrm{d}s, \\ &= \frac{1}{2}\left(\frac{1}{3}m_1 l_1^2\left(\dot{q}^1\right)^2 + m_2 l_1^2\left(\dot{q}^1\right)^2 \right. \\ &\qquad \left. + \frac{1}{3}m_2 l_2^2\left(\dot{q}^2\right)^2 + m_2 l_1 l_2\cos\left(q^1-q^2\right)\dot{q}^1\dot{q}^2\right). \end{aligned}$$

  This kinetic energy defines a Riemannian structure on the configuration space. The expression of the metric tensor in the considered chart is:

$$\begin{aligned} g_{11}\left(q^1,q^2\right) &= \left(\frac{m_1}{3}+m_2\right)l_1^2, \\ g_{12}\left(q^1,q^2\right) &= \frac{1}{2}m_2 l_1 l_2\cos\left(q^1-q^2\right) = g_{21}\left(q^1,q^2\right), \\ g_{22}\left(q^1,q^2\right) &= \frac{1}{3}m_2 l_2^2. \end{aligned}$$

- The forces mapping has for expression in the considered chart:

$$\begin{aligned} f(q,\dot{q};t) &= \left[\lambda l_1\sin\left(q^1-q^2\right) - \left(k_1+k_2\right)q^1 + k_2 q^2\right]e^1(q) \\ &\quad + \left[k_2 q^1 - k_2 q^2\right]e^2(q). \end{aligned}$$

The equations of motion in the chart under consideration is easily formed by use of proposition 1:

$$\begin{aligned} &\left(\tfrac{m_1}{3}+m_2\right)l_1^2\ddot{q}^1 + \tfrac{m_2}{2}l_1 l_2\cos\left(q^1-q^2\right)\ddot{q}^2 + \tfrac{m_2}{2}l_1 l_2\sin\left(q^1-q^2\right)\left(\dot{q}^2\right)^2 \\ &\qquad = \lambda l_1\sin\left(q^1-q^2\right) - \left(k_1+k_2\right)q^1 + k_2 q^2, \\ &\tfrac{m_2}{2}l_1 l_2\cos\left(q^1-q^2\right)\ddot{q}^1 + \tfrac{m_2}{3}l_2^2\ddot{q}^2 - \tfrac{m_2}{2}l_1 l_2\sin\left(q^1-q^2\right)\left(\dot{q}^1\right)^2 \\ &\qquad = k_2\left(q^1-q^2\right). \end{aligned}$$

By proposition 6, one can conclude that a unique maximal motion is associated with any initial condition. Moreover, this maximal motion is analytic and is defined for all time. Indeed, it is easily seen that there exists a positive real constant C, depending only on (l_1, l_2, m_1, m_2) such that:

$$\|f(q)\|_q^* \leq C\Big[\lambda l_1 + 4(k_1+k_2)\left|(q_0^1, q_0^2)\right| + 4(k_1+k_2)d(q,q_0)\Big],$$

where $|\cdot|$ denotes the canonical Euclidean norm on $\mathbb{R}^2$. Therefore, the assumptions of theorem 3 are satisfied.

It should be underlined that the framework of perfect bilateral constraints does not require that there should be no energy dissipation physically associated with a constraint. Indeed, such an energy dissipation can be described, in some cases, in terms of internal forces. For example, suppose that, in the system described above, some viscous damping with coefficients η_1 and η_2 is associated with each ball-and-socket joint. Then, it is incorporated in the forces mapping f which should be changed into

$$\begin{aligned} f(q,\dot{q};t) \;=\; & \left[\lambda l_1 \sin\left(q^1 - q^2\right) - (k_1 + k_2)\, q^1 + k_2 q^2 \right. \\ & \left. - (\eta_1 + \eta_2)\, \dot{q}^1 + \eta_2 \dot{q}^2\right] e^1(q) \\ & + \left[k_2 q^1 - k_2 q^2 + \eta_2 \dot{q}^1 - \eta_2 \dot{q}^2\right] e^2(q). \end{aligned}$$

The above remark does not apply to the case of Coulomb type friction.

Remark 2. As problems II and II′ are equivalent, we see that the dynamics of the constrained system depends only on the geometry of the submanifold S and *not* on the particular choice of the functions φ_i used to define it. In other words, consider a constraint, say constraint 1, defined by n functionally independent functions φ_i and another constraint, say constraint 2, defined by n functionally independent functions φ'_i. Suppose, in addition, that:

$$S = \left\{q \in Q \;;\; \forall\, i,\; \varphi_i(q) = 0\right\} = \left\{q \in Q \;;\; \forall\, i,\; \varphi'_i(q) = 0\right\}.$$

Then, the dynamics of the system subjected to constraint 1 is identical to the dynamics of the system subjected to constraint 2. Moreover, the reaction forces in the motion are the same in both cases.

Since the modelling of rigid bodies system with no constraint or with perfect holonomic bilateral constraint leads to the construction of mathematical structures of the same type, we state the following definition.

Definition 7 *A simple discrete mechanical system is a pair* (Q, f) *where:*

- Q *is a finite-dimensional smooth Riemannian manifold called the configuration manifold.*
- $f : TQ \times \mathbb{R} \to T^*Q$ *is a smooth mapping satisfying:*

$$\forall\, (q,v) \in TQ, \quad \forall t \in \mathbb{R}, \qquad \Pi_Q^*(f(q,v;t)) = q,$$

called the forces mapping.

3. Perfect unilateral constraints

The consideration of elementary examples shows that the dynamics of rigid bodies systems can lead to some prediction of the motion where some bodies of the system *overlap* in the real world. Of course, this should not be allowed. Hence, very often, one has to add the statement of non-penetration conditions to a simple discrete mechanical system. This is a simple occurrence of unilateral constraint. In this section, we shall discuss the consideration of perfect unilateral constraints in simple discrete mechanical systems.

3.1 The geometric description

Consider a simple discrete mechanical system with configuration manifold Q. A *unilateral constraint* is a restriction on the admissible motions of the system which is expressed by means of a finite number n of smooth real-valued functions φ_i defined on the configuration manifold Q:

$$\forall\, i \in \{1, 2, \cdots, n\}, \quad \varphi_i(q) \leq 0. \tag{1.6}$$

We denote by A the set of all admissible configurations:

$$A = \{q \in Q \,;\, \forall\, i \in \{1, 2, \cdots, n\}, \quad \varphi_i(q) \leq 0\}.$$

The set of all active constraints in the admissible configuration $q \in A$ is defined by:

$$J(q) = \big\{i \in \{1, 2, \cdots, n\} \;;\; \varphi_i(q) = 0\big\}.$$

The following hypothesis should be brought aside regularity hypothesis I of Section 2.2.1.

Regularity hypothesis I. The functions φ_i are *functionally independent* in the sense that, for all $q \in A$, the $\mathrm{d}\varphi_i(q)$ $(i \in J(q))$ are linearly independent in T^*Q.

Straightforward consequences of this hypothesis are:

- A is a closed subset of Q,
- $\partial A \subset \bigcup_{i=1}^{n} \varphi_i^{-1}(\{0\})$ (∂A is the boundary of A),
- $\overset{\circ}{A} = J^{-1}(\{\emptyset\})$ ($\overset{\circ}{A}$ is the interior of A).

Consider a motion $q(t)$ in A and assume that a right velocity $\dot{q}^+(t) \in T_{q(t)}Q$ exists at instant t, then we necessarily have:

$$\forall\, i \in J(q(t)), \quad \big\langle \mathrm{d}\varphi_i(q(t)), \dot{q}^+(t) \big\rangle_{q(t)} \leq 0,$$

or, equivalently,

$$\forall i \in J(q(t)), \quad \left(\nabla\varphi_i(q(t)), \dot{q}^+(t)\right)_{q(t)} \leq 0,$$

where $\nabla\varphi_i(q)$ is the gradient of φ_i at q defined by $\nabla\varphi_i(q) = \sharp\,(\mathrm{d}\varphi_i(q))$. Thus, if the system has configuration q and if a right velocity $\dot{q}^+$ exists, then $\dot{q}^+$ necessarily belongs to the closed convex cone $V(q)$ of T_qQ defined by:

$$V(q) = \left\{v \in T_qQ \ ; \ \forall i \in J(q), \quad \langle \mathrm{d}\varphi_i(q), v\rangle_q \leq 0\right\}.$$

$V(q)$ is called the cone of admissible right velocities at the configuration q. In particular,

$$q \in \overset{\circ}{A} \text{ (i.e. } J(q) = \emptyset) \Longrightarrow V(q) = T_qQ.$$

Similarly, if a left velocity $\dot{q}^- \in T_qQ$ exists, then $\dot{q}^- \in -V(q)$

3.2 Formulation of the dynamics

The formulation of the dynamics follows the lines of MOREAU (1983, 1988a).

3.2.1 Equation of motion. As for bilateral constraints, the realization of the constraints induces some reaction force R. The following hypotheses are made.

Constitutive hypothesis II. The unilateral constraints are of type contact without adhesion:

$$\forall v \in V(q), \quad \langle R, v\rangle_q \geq 0.$$

Constitutive hypothesis III. The unilateral constraints are perfect:

$$\forall v \in \left\{v \in T_qQ \ ; \ \forall i \in J(q), \quad \langle \mathrm{d}\varphi_i(q), v\rangle_q = 0\right\}, \qquad \langle R, v\rangle_q = 0.$$

As an easy consequence of constitutive hypotheses II and III, we get:

$$\exists\ (\lambda_i)_{i=1,2,\cdots,n} \in \mathbb{R}^n, \quad R = \sum_{i=1}^{n} \lambda_i\, \mathrm{d}\varphi_i(q), \text{ and } \left| \begin{array}{lcl} i \in J(q) & \Rightarrow & \lambda_i \leq 0, \\ i \notin J(q) & \Rightarrow & \lambda_i = 0. \end{array} \right.$$

Thus, the reaction force $R \in T^*Q$ must be such that:

$$-R \in N^*(q) \overset{\text{def}}{=} \left\{ \sum_{i=1}^{n} \lambda_i\, \mathrm{d}\varphi_i(q) \ ; \ \forall i \in J(q),\ \lambda_i \geq 0,\ \forall i \notin J(q),\ \lambda_i = 0 \right\}. \tag{1.7}$$

$N^*(q)$ is a closed convex cone of T_q^*Q and it is the polar cone of $V(q)$ in the duality $\left(T_qQ, T_q^*Q\right)$. We will also have to consider the polar cone $N(q)$ of $V(q)$ for the Euclidean structure of T_qQ:

$$N(q) = \left\{ \sum_{i=1}^{n} \lambda_i \nabla\varphi_i(q) \; ; \; \forall i \in J(q), \; \lambda_i \geq 0, \quad \forall i \notin J(q), \; \lambda_i = 0 \right\}.$$

Now, consider a motion $q(t)$ starting at $q_0 \in \overset{\circ}{A}$ at time t_0 with velocity v_0. Assumed to be continuous, $q(t)$ remains in $\overset{\circ}{A}$ on a right neighbourhood of t_0. By formula (1.7), the reaction force R vanishes as long as $q(t)$ is in $\overset{\circ}{A}$ and the motion is governed by the ordinary differential equation:

$$\begin{aligned} &(q(t_0), \dot{q}(t_0)) = (q_0, v_0), \\ &\flat \frac{\mathrm{D}\dot{q}}{\mathrm{d}t} = f(q, \dot{q}; t). \end{aligned}$$

Suppose that the solution of this Cauchy problem meets ∂A at some instant greater than t_0. Denote by T the smallest of these instants. The motion admits a left velocity vector v_T^- at time T. Of course, there may happen: $v_T^- \notin V(q(T))$. In this case, no differentiable prolongation of the motion can exist in A for t greater than T. The requirement of differentiability has to be dropped. An instant such T is called an instant of *impact*.

However, we are still going to require the existence of a right velocity vector $\dot{q}^+(t) \in V(q(t))$ at every instant t. The right velocity need not to be a continuous function of time and the equation of motion

$$\flat \frac{\mathrm{D}\dot{q}^+}{\mathrm{d}t} = f(q, \dot{q}^+; t) + R,$$

should be understood in sense of Schwartz's distribution. Actually, we require R to be a *vector valued measure* rather than a general distribution.

We denote by $MMA(I; Q)$ (motions with measure acceleration) the set of all absolutely continuous motions $q(t)$ from a real interval I to Q admitting a right velocity $\dot{q}^+(t)$ at every instant t of I and such that the function $\dot{q}^+(t)$ has locally bounded variation over I. Naturally, bounded variation is classically defined only for functions taking values in a normed vector space. However, for any absolutely continuous curve $q(t)$ on a Riemannian manifold, parallel translation along $q(t)$ classically provides intrinsic identification of the tangent spaces at different points of the curve and so, the definitions can easily be carried over to this case. The precise mathematical setting is postponed to Appendix A. The reader will notice from Appendix A that any motion $q \in MMA(I; Q)$ admits a left and right velocity, $\dot{q}^-$ and $\dot{q}^+$, in the classical sense at any instant. Moreover, with any motion $q \in MMA(I; Q)$ is intrinsically associated the

covariant Stieltjes measure $\mathrm{D}q^+$ of its right velocity q^+. The equation of motion ta es the form:

$$\flat \mathrm{D}\dot{q}^+ = f(q, \dot{q}^+; t)\,\mathrm{d}t + R,$$

where $\mathrm{d}t$ denotes the Lebesgue measure. We have to give a precise meaning to condition (1.7) with R being a vector valued measure.

Convention. We shall write:

$$R \in -N^*(q(t))$$

to mean: there exist n nonpositive real measures λ_i such that:

$$R = \sum_{i=1}^{n} \lambda_i\,\mathrm{d}\varphi_i(q(t)),$$
$$\forall i \in \{1, 2, \cdots, n\}, \quad \mathrm{Supp}\,\lambda_i \subset \{t\,;\ \varphi_i(q(t)) = 0\}. \tag{1.8}$$

Using this convention, the final form of the equation of motion is:

$$R = \flat \mathrm{D}\dot{q}^+ - f(q(t), \dot{q}^+(t); t)\,\mathrm{d}t \in -N^*(q(t)) \tag{1.9}$$

A straightforward consequence of the equation of motion is that an impact (that is, a discontinuity of the right velocity $\dot{q}^+$ by proposition 43) can only occur at an instant t such that $J(q(t) \neq \emptyset$. This fact is a justification for the following definition.

Definition 8 *An impact occuring at time t is said* simple *if $J(q(t)$ contains exactly one element. If $J(q(t))$ contains at least two elements, the impact is said* multiple.

3.2.2 The impact constitutive equation. We begin this section by an example. Consider the one degree-of-freedom mechanical system whose configuration space is $\mathbb{R}$ equipped with its canonical Euclidean structure. The forces mapping f vanishes identically and the unilateral constraint is represented by the single function $\varphi_1(q) = q$ so that the admissible configuration set A is $\mathbb{R}^-$. At initial time $t_0 = 0$, we consider an initial state (q_0, v_0) such that $q_0 < 0$ and $v_0 > 0$. It is readily seen from the equation of motion (1.9) that an impact necessarily occurs at time $t = -q_0/v_0$. At this time, the left velocity is v_0. But, the right velocity can take any negative value and whatever it is, it is compatible with the equation of motion.

The reason for this indetermination lies in the phenomenological nature of the interaction of the system with the obstacle. This missing information has to be added.

Constitutive hypothesis IV. The interaction of the system with the obstacle at time t is completely determined by the present configuration $q(t)$ and the

present left velocity $\dot{q}^-(t)$. In other terms, we postulate the existence of a mapping $\mathcal{F} : TQ \to TQ$ describing the interaction of the system with the obstacle during an impact:

$$\forall t, \quad \dot{q}^+(t) = \mathcal{F}\left(q(t), \dot{q}^-(t)\right). \tag{1.10}$$

To ensure compatibility with the equation of motion (1.9), the mapping $\mathcal{F}$ should satisfy:

$$\forall q \in A, \quad \forall v^- \in -V(q), \qquad \begin{array}{l} \mathcal{F}(q, v^-) \in V(q), \\ \mathcal{F}(q, v^-) - v^- \in -N(q). \end{array} \tag{1.11}$$

Moreover, we add the assumption that the kinetic energy of the system can not increase during an impact:

$$\forall q \in A, \quad \forall v^- \in -V(q), \qquad \left\|\mathcal{F}\left(q, v^-\right)\right\|_q \leq \left\|v^-\right\|_q. \tag{1.12}$$

Let us comment on hypothesis IV. When two solids hit, their bouncing is actually governed by the propagation of deformation waves in each the two solids. But, from the very beginning, we have adopted the simple framework in which each solid is supposed to be rigid, that is, for sake of simplicity, we have chosen to do not take under consideration any phenomena relying on the deformation of the solids. Thus, we cannot expect the theory to be able to predict the outcome of an impact experiment. The aim of constitutive hyposthesis IV is to introduce in the theory the missing information. Of course, in practical situations, we have to identify the unknown mapping $\mathcal{F}$. This can be done either by means of experiments or by use of a refined theory. For example, the theory of elastodynamics could be used to predict the outcome of an impact in every impact configuration. The result would be an identification of the mapping $\mathcal{F}$. In any case, there is a very big amount of work to get a precise identification of $\mathcal{F}$. This is the price we have to pay to describe sophisticated physical phenomena in a very simple framework. Actually, this issue is faced in any mechanical theory (one could think of the theory of elasticity). Naturally, in each mechanical theory, the question arises to know what amount of lacking constitutive information should be introduced. Most of the time, well-posedness of the resulting evolution problem serves as a guideline to state the constitutive hypotheses.

Definition 9 *Equation (1.10), with mapping $\mathcal{F}$ fulfilling both requirements (1.11) and (1.12) is called the* impact constitutive equation. *An impact constitutive equation which ensures the conservation of kinetic energy during an impact:*

$$\forall q \in A, \quad \forall v^- \in -V(q), \qquad \left\|\mathcal{F}\left(q, v^-\right)\right\|_q = \left\|v^-\right\|_q,$$

is called elastic.

There always exist many mappings $\mathcal{F}$ satisfying requirements (1.11) and (1.12).

Example 5. Let $e : TQ \to [0,1]$ be an arbitrary function. The mapping $\mathcal{F}$ defined by:

$$\mathcal{F}\left(q,v^-\right) = \mathrm{Proj}_q\left[v^-; V(q)\right] - e\left(q,v^-\right)\mathrm{Proj}_q\left[v^-; N(q)\right], \qquad (1.13)$$

is easily seen to satisfy requirements (1.11) and (1.12). The associated impact constitutive equation is often called the *canonical* impact constitutive equation. It is elastic if and only if $e \equiv 1$. The function e is classically called the *restitution* coefficient.

The reason why the canonical impact constitutive equation is distinguished is that in situations where only simple impact can occur (for example, if the unilateral constraint is represented by a single function φ_1), then the impact constitutive equation must be the canonical one (it is a simple consequence of requirements (1.11) and (1.12)). However, in case of multiple impacts, the canonical impact constitutive equation has no specific physical relevance. A simple occurence of multiple impact is provided by Newton's cradle. The principle of the experiment is sketched on Figure 1.2.a. Its outcome is sketched on Figure 1.2.b. It should be compared with the prediction of the canonical elastic impact constitutive equation which is sketched on Figure 1.2.c.

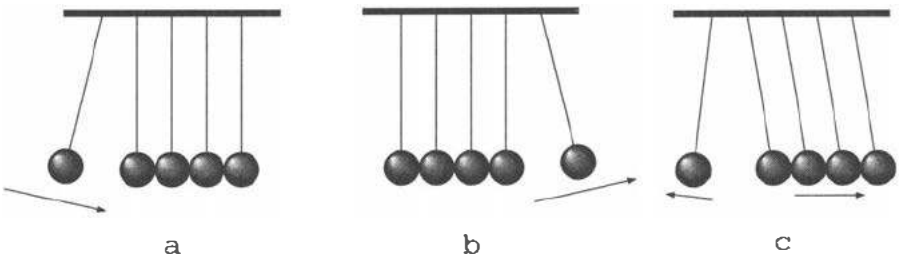

Figure 1.2. Newton's cradle.

The following proposition is a straightforward and useful consequence of requirements (1.11) and (1.12).

Proposition 10 *Let $\mathcal{F}$ be a constitutive mapping satisfying requirements (1.11) and (1.12). Then, we have:*

$$\forall q \in A, \quad \forall v^- \in V(q) \cap (-V(q)), \qquad \mathcal{F}(q,v^-) = v^-.$$

Proof. Define $v^+ = \mathcal{F}(q, v^-)$. By requirement (1.11), we have $v^- - v^+ \in N(q)$. Since $v^- \in V(q) \cap (-V(q))$, we obtain:

$$\left(v^- - v^+, v^-\right)_q = 0,$$

that is,

$$\left(v^+, v^-\right)_q = \left\|v^-\right\|_q^2.$$

The use of Cauchy-Schwarz inequality and requirement (1.12) gives the desired result. □

We conclude this section by a comment on requirement (1.12). At first glance, it could seem to be unnecessary. The following counter-example proves that if it was omitted, then, uniqueness of solution for the resulting evolution problem would surely not hold.

Counter-example 6. Consider the one degree of freedom discrete mechanical system whose configuration space is $\mathbb{R}$ equipped with its canonical structure of Riemannian manifold. The forces mapping is supposed to be constant: $f(q, \dot{q}; t) \equiv 2$. To this simple discrete mechanical system, we add the unilateral constraint described by the single function $\varphi_1(q) = q$. Thus, $A = \mathbb{R}^-$. The impact constitutive equation is given by formula (1.13) where the restitution coefficient is supposed to be the constant $1/2$: $e(q, \dot{q}^-) \equiv 1/2$. This mechanical system is a formal description of the physical occurence of a single particle subjected to gravity and bouncing on the floor. Consider the initial instant $t_0 = 0$ and the initial state $(q_0, v_0) = (-1, 0)$. It is readily seen that the function $q : \mathbb{R}^+ \to \mathbb{R}^-$ defined by:

$$\begin{array}{ll}
\forall t \in [0,1], & q(t) = t^2 - 1, \\
\forall t \in [1,2], & q(t) = t^2 - 3t + 2, \\
\forall t \in \left[3 - \frac{1}{2^{n-1}}, 3 - \frac{1}{2^n}\right], & q(t) = t^2 - \left(6 - \frac{3}{2^n}\right) t + \left(3 - \frac{1}{2^{n-1}}\right)\left(3 - \frac{1}{2^n}\right), \\
\forall t \in [3, +\infty[, & q(t) = 0,
\end{array}$$

$(n \in \mathbb{N})$ belongs to $MMA(\mathbb{R}^+; \mathbb{R}^-)$ and satisfies:

- the initial condition,
- the equation of motion (1.9) (with $f(q, \dot{q}; t) \equiv 2$),
- the impact constitutive equation (1.13) (with $e(q, \dot{q}) \equiv 1/2$).

This motion is pictured on Figure 1.3. Note, by the way, that it exhibits an infinite number of impacts on a compact time subinterval. It could easily be proved that no motion, defined on $[0, \infty[$, with finite number of impact on every compact interval can exist.

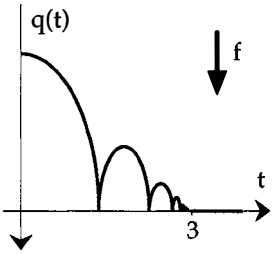

Figure 1.3. Motion of a punctual particle subjected to gravity and bouncing on the floor.

Now, we are going to analyse what happens when the flow of time is reversed. Define q' by:

$$q' \left\{ \begin{array}{rcl} [0,4] & \to & \mathbb{R}^- \\ t & \mapsto & q(4-t) \end{array} \right.$$

Considering the initial state $(q_0, v_0) = (0,0)$ at $t_0 = 0$, it is easily seen that q' satisfies:

- that initial condition,
- the equation of motion (1.9) (with $f(q, \dot{q}; t) \equiv 2$),
- the impact constitutive equation (1.13) (with $e(q, \dot{q}) \equiv 2$).

But, $q'' \equiv 0$ is also seen to satisfy the same initial condition, equation of motion and impact constitutive equation. This example demonstrates that we cannot expect uniqueness of solution when adopting the canonical impact constitutive equation (1.13) with restitution coefficient $e \equiv 2$ (or any real number strictly greater than 1). But the canonical impact constitutive equation with restitution coefficient strictly greater than 1 violates requirement (1.12).

3.2.3 The evolution problem. Now, we formulate the evolution problem associated with the dynamics of rigid bodies systems with perfect bilateral and unilateral constraints. The initial condition is assumed to be compatible with the realization of the constraint: $v_0 \in V(q_0)$.

Problem III. Find $T > t_0$ and $q \in MMA([t_0, T[; Q)$ such that:

- $(q(t_0), \dot{q}^+(t_0)) = (q_0, v_0)$,
- $\forall\, t \in [t_0, T[, \quad q(t) \in A$,
- $R \stackrel{\text{def}}{=} \flat \mathrm{D}\dot{q}^+ - f(q(t), \dot{q}^+(t); t)\, \mathrm{d}t \in -N^*(q(t))$,
- $\forall\, t \in]t_0, T[, \quad \dot{q}^+(t) = \mathcal{F}\left(q(t), \dot{q}^-(t)\right)$.

The equation of motion is understood in sense of convention (1.8) and the impact constitutive equation is supposed to fulfill requirements (1.11) and (1.12).

Yet, no regularity assumption has been made on the mapping f. This will be done in the next section where well-posedness of problem III is studied. However, we can infer from Section 1.1.3 that f will be assumed to be at least of class C^1. We can state an elementary property of any solution (if there are any) of problem III.

Proposition 11 (Energy inequality) *Let (T, q) be an arbitrary solution of problem III. Then, it satisfies:*

$$\forall t_1, t_2 \in [t_0, T[, \quad t_1 \leq t_2, \qquad K\left(q(t_2), \dot{q}^+(t_2)\right) - K\left(q(t_1), \dot{q}^+(t_1)\right) =$$
$$\frac{1}{2}\left\|\dot{q}^+(t_2)\right\|^2_{q(t_2)} - \frac{1}{2}\left\|\dot{q}^+(t_1)\right\|^2_{q(t_1)} \leq \int_{t_1}^{t_2} \left\langle f\left(q(s), \dot{q}^+(s); s\right), \dot{q}^+(s)\right\rangle_{q(s)} \mathrm{d}s$$

Proof. We have the following equality between real measures:

$$\left(\frac{\dot{q}^+(t) + \dot{q}^-(t)}{2}, \mathrm{D}\dot{q}^+\right)_{q(t)} =$$
$$\left\langle \frac{\dot{q}^+(t) + \dot{q}^-(t)}{2}, f\left(q(t), \dot{q}^+(t); t\right)\right\rangle_{q(t)} \mathrm{d}t + \left\langle \frac{\dot{q}^+(t) + \dot{q}^-(t)}{2}, R\right\rangle_{q(t)}.$$

Integrating over $]t_1, t_2]$ and using proposition 41 of Appendix A, we get:

$$\frac{1}{2}\left\|\dot{q}^+(t_2)\right\|^2_{q(t_2)} - \frac{1}{2}\left\|\dot{q}^+(t_1)\right\|^2_{q(t_1)} =$$
$$\int_{]t_1,t_2]} \left\langle \frac{\dot{q}^+(t) + \dot{q}^-(t)}{2}, f\left(q(t), \dot{q}^+(t); t\right)\right\rangle_{q(t)} \mathrm{d}t + \int_{]t_1,t_2]} \left\langle \frac{\dot{q}^+ + \dot{q}^-}{2}, R\right\rangle_q.$$

Consider

$$D = \left\{ t \in]t_1, t_2] \; ; \; \frac{\dot{q}^+(t) + \dot{q}^-(t)}{2} \neq \dot{q}^+(t) \right\}.$$

D is (at most) countable and therefore Lebesgue-negligible. We obtain:

$$\frac{1}{2}\left\|\dot{q}^+(t_2)\right\|^2_{q(t_2)} - \frac{1}{2}\left\|\dot{q}^+(t_1)\right\|^2_{q(t_1)} =$$
$$\int_{t_1}^{t_2} \left\langle \dot{q}^+(t), f\left(q(t), \dot{q}^+(t); t\right)\right\rangle_{q(t)} \mathrm{d}t + \int_{]t_1,t_2]} \left\langle \frac{\dot{q}^+ + \dot{q}^-}{2}, R\right\rangle_q.$$

Therefore, to prove proposition 11, there remains only to prove:

$$\int_{]t_1,t_2]} \left\langle \frac{\dot{q}^+ + \dot{q}^-}{2}, R\right\rangle_q \leq 0. \tag{1.14}$$

But, on one hand,

$$\int_{]t_1,t_2]\setminus D} \left\langle \frac{\dot{q}^+ + \dot{q}^-}{2}, R \right\rangle_q = \int_{]t_1,t_2]\setminus D} \langle \dot{q}^+, R \rangle_q = \int_{]t_1,t_2]\setminus D} \langle \dot{q}^-, R \rangle_q ,$$

where the second integral is nonnegative by convention (1.8) whereas the third integral is nonpositive. As a consequence:

$$\int_{]t_1,t_2]\setminus D} \left\langle \frac{\dot{q}^+ + \dot{q}^-}{2}, R \right\rangle_q = 0. \tag{1.15}$$

On the other hand,

$$\begin{aligned} \int_D \langle \frac{\dot{q}^+(t) + \dot{q}^-(t)}{2}, R \rangle_{q(t)} &= \int_D \left(\frac{\dot{q}^+(t) + \dot{q}^-(t)}{2}, \mathrm{D}\dot{q}^+ \right)_{q(t)}, \\ &= \frac{1}{2} \sum_{t \in D} \left(\|\dot{q}^+(t)\|^2_{q(t)} - \|\dot{q}^-(t)\|^2_{q(t)} \right), \\ &\leq 0, \end{aligned} \tag{1.16}$$

by virtue of formula (1.12). Bringing together formulae (1.15) and (1.16), we get inequality (1.14). □

3.3 Well-posedness of the dynamics

To study the well-posedness of problem III, we need to state regularity assumptions on the data. Looking at those of Section 2.2.3, we could expect to be able to prove well-posedness of problem III under the assumption that the functions φ_i and the mapping f are of class C^2 and C^1 respectively. The following counter-example originally due to BRESSAN (1960) and SCHATZMAN (1978) shows that uniqueness does not hold in general even if the data are supposed to be of class C^∞.

Counter-example 7. Consider a simple discrete mechanical system whose configuration space is $\mathbb{R}$ equipped with its canonical structure of Riemannian manifold. This is the configuration space of a particle with unit mass constrained to move along a line. A fixed obstacle at the origin is taken into consideration. It gives rise to a unilateral constraint described by the single function:

$$\varphi_1(q) = q$$

Therefore, the admissible configuration set is $A = \mathbb{R}^-$. The impact constitutive equation is supposed to be elastic. Here, the geometry is so poor that this statement determines completely the impact constitutive equation. It is necessarily the canonical one with restitution coefficient $e \equiv 1$. The forces mapping f is

supposed not to depend on the state but only on time. It will be denoted by $f(t)$. The initial instant is $t_0 = 0$ and the initial state is $(q_0, v_0) = (0, 0)$. The corresponding problem III admits here the simple formulation:
find $T > 0$ and $q \in MMA([0, T[; \mathbb{R})$ such that:

- $(q(0), \dot{q}^+(0)) = (0, 0)$,
- $\forall t \in [0, T[, \quad q(t) \leq 0$,
- $R \stackrel{\text{def}}{=} \mathrm{d}\dot{q}^+ - f(t)\,\mathrm{d}t$ is a nonpositive real measure such that: $\mathrm{Supp}\, R \subset \{t \in [0, T[\ ;\ q(t) = 0\}$,
- $\forall t \in]0, T[, \quad \begin{cases} q(t) \neq 0 & \Rightarrow \quad \dot{q}^+(t) = \dot{q}^-(t) \\ q(t) = 0 & \Rightarrow \quad \dot{q}^+(t) = -\dot{q}^-(t) \end{cases}$

Here $\mathrm{d}\dot{q}^+$ is merely the classical Stieltjes measure associated with the function with locally bounded variation $\dot{q}^+$. We investigate uniqueness under the assumption that f is of class C^∞ and nonnegative:

$$\forall t \in \mathbb{R}^+, \quad f(t) \geq 0.$$

Then, it is readily seen that the null function $\tilde{q} \equiv 0$ on $\mathbb{R}^+$ is a solution of that problem, whatever is the nonnegative C^∞ function f. Now, we are going to construct an explicit example of such a function f in such a way that the associated evolution problem III admits another solution, distinct from the identically vanishing one.

First, define a Massin function ρ by:

$$\rho \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & \left| \begin{array}{ll} 0 & \text{if} \quad x \in]-\infty, 0] \cup [1, +\infty[\\ C e^{\frac{1}{x(x-1)}} & \text{if} \quad x \in]0, 1[\end{array} \right. \end{array} \right.$$

where C is a real constant which is chosen to get:

$$\int_{-\infty}^{+\infty} \rho(x)\,\mathrm{d}x = 1.$$

Define:

$$T = \sum_{n=0}^{\infty} \frac{(n+5)^2}{(n+1)(n+2)(n+3)(n+4)},$$

and, for every $n \in \mathbb{N}$,

$$\begin{aligned} a_n &= \sum_{i=n}^{\infty} \frac{(i+5)^2}{(i+1)(i+2)(i+3)(i+4)}, \\ \delta_n &= \frac{n+5}{(n+1)(n+2)(n+4)} \quad \left(\text{i.e. } \delta_n = \frac{n+3}{n+5}(a_n - a_{n+1})\right), \\ f_n &= \frac{1}{n!}, \\ v_n &= -\frac{1}{(n+3)!}. \end{aligned}$$

Now, the functions $f(t)$ and $v(t)$, from $[0,T[$ to $\mathbb{R}$, are defined by:

$$f(0) = 0$$
$$f(t) = \begin{cases} 0, & \text{if } \ t \in [a_{n+1}, a_{n+1} + \delta_n[\\ \dfrac{f_n}{2}\rho\left(\dfrac{t-a_{n+1}-\delta_n}{a_n-a_{n+1}-\delta_n}\right), & \text{if } \ t \in [a_{n+1}+\delta_n, a_n[\end{cases}$$

and:

$$v(0) = 0,$$
$$v(t) = \begin{cases} v_{n+1}, & \text{if } t \in [a_{n+1}, a_{n+1} + \delta_n[\\ v_{n+1} + \dfrac{f_n}{2}\displaystyle\int_{a_{n+1}+\delta_n}^{t} \rho\left(\dfrac{s-a_{n+1}-\delta_n}{a_n-a_{n+1}-\delta_n}\right)\mathrm{d}s, & \text{if } t \in [a_{n+1}+\delta_n, a_n[\end{cases}$$

Finally the function $q : [0,T[\to \mathbb{R}$ is defined by:

$$q(t) = \int_0^t v(s)\,\mathrm{d}s.$$

The graph of the functions $f(t)$ and $q(t)$ is sketched on Figure 1.4. The reader will easily check that:

- $f(t)$ is a C^∞ nonnegative function on $[0,T[$,
- (T,q) is a solution of the considered evolution problem,
- the only instants at which $q(t) = 0$ are 0 and the a_n.

Therefore, q and $\tilde{q}$ provide two solutions of the evolution problem. These two solutions do not coincide on any open subinterval of $[0,T[$. Therefore, uniqueness of solution for problem III cannot be asserted, even in the case

where the data are supposed to be of class C^∞. PERCIVALE (1985, 1991) was the first to notice that, in the above example, if $f(t)$ is supposed to be *analytic*, then uniqueness of solution does hold. Recently, a complete discussion of the one-degree-of freedom problem was obtained by SCHATZMAN (1998). The general case is treated in BALLARD (2000) and is now recalled. Let us just mention that prior existence results had been obtained, but they were limited to the case where the unilateral constraint is represented by a single function (see MONTEIRO MARQUES (1993) and SCHATZMAN (1978)).

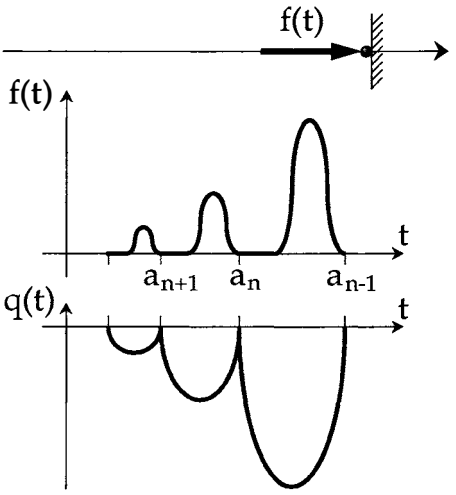

Figure 1.4. Bressan-Schatzman counterexample.

Regularity hypothesis V. The Riemannian configuration manifold, the functions φ_i and the mapping $f : TQ \times \mathbb{R} \to T^*Q$ are *analytic.*

The proof of the following proposition can be found in BALLARD (2000). An earlier proof can also be found in LÖTSTEDT (1982).

Proposition 12 *Let $q_0 \in A$ and $v_0 \in V(q_0)$. Then, there exist $T_a > t_0$, an analytic curve $q_a : [t_0, T_a[\to Q$ and n analytic functions $\lambda_{ai} : [t_0, T_a[\to \mathbb{R}$ such that:*

- $(q_a(0), \dot{q}_a^+(0)) = (q_0, v_0)$,
- $\forall t \in [t_0, T_a[, \quad \flat \dfrac{D}{dt}\dot{q}_a(t) = f(q_a(t), \dot{q}_a(t); t) + \sum_{i=1}^{n} \lambda_{ai}(t)\, d\varphi_i(q_a(t))$,
- $\forall t \in [t_0, T_a[$, $\forall i = 1, 2, \cdots, n$, $\quad \lambda_{ai}(t) \leq 0, \quad \varphi_i(q_a(t)) \leq 0, \quad \lambda_{ai}(t)\, \varphi_i(q_a(t)) = 0.$

Moreover, the solution of this evolution problem is unique in the sense that any other analytic solution $(T, q, \lambda_1, \cdots, \lambda_n)$ *is either a restriction or analytic extension of* $(T_a, q_a, \lambda_{a1}, \cdots, \lambda_{an})$.

Corollary 13 *There exists an analytic solution* (T_a, q_a) *for problem III.*

Proof. Consider the motion q_a furnished by proposition 12. It obviously satisfies the initial condition, the unilateral constraint and the equation of motion. The only thing which remains to prove is that it satisfies the impact constitutive equation. Since q_a is analytic and satisfies the unilateral constraint, we have:

$$\forall t \in]t_0, T_a[, \quad \dot{q}_a^-(t) = \dot{q}_a^+(t) \in V(q_a(t)) \cap (-V(q_a(t)))\,,$$

and therefore,

$$\forall t \in]t_0, T_a[, \quad \dot{q}_a^+(t) = \dot{q}_a^-(t) = \mathcal{F}\left(q_a(t), \dot{q}_a^-(t)\right),$$

by proposition 10. □

Naturally, the analytic solution furnished by corollary 13 will cease to exist at the first instant of impact. This is the reason why we have considered the wider class *MMA* which contains motions which are not differentiable in the classical sense. Considering motions in *MMA* will allow to extend the solution beyond the first instant of impact. But, it must be made sure that admitting the wider class of solutions *MMA* will not introduce parasitic solutions. This is the aim of the following theorem.

Theorem 14 *Let* (T_a, q_a) *be the solution for problem III furnished by corollary 13, and* (T, q) *be an arbitrary solution for problem III. Then, there exists a real number* T_0 *(*$t_0 < T_0 \leq \min\{T_a, T\}$*) such that:*

$$q_{|[t_0, T_0[} \equiv q_{a|[t_0, T_0[}\,.$$

In other terms, there is local uniqueness for problem III.

The proof of theorem 14 makes extensive use of the following corollary of Gronwall-Bellman lemma (lemma 4).

Lemma 15 *Let m be a nonnegative integer, and $\psi : [0,T] \to \mathbb{R}$ an integrable function. If $\phi : [0,T] \to \mathbb{R}$ is any absolutely continuous function such that $\phi(t) = o(t^{m+1})$ when t tends towards 0 and such that there exists a nonnegative real constant C such that:*

$$\text{for almost all } t \in]0,T[, \quad t\frac{\mathrm{d}}{\mathrm{d}t}\phi(t) \leq (1+m+Ct)\,\phi(t) + t^{m+2}\psi(t),$$

then,

$$\forall\, t \in [0,T], \quad \phi(t) \leq t^{m+1}e^{Ct}\int_0^t \psi(s)e^{-Cs}\mathrm{d}s.$$

Proof. This is almost obvious. Dividing each member of the inequality by t^{m+2}, we obtain:

$$\text{for almost all } t \in]0,T[, \quad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\phi(t)}{t^{m+1}}\right) \leq C\frac{\phi(t)}{t^{m+1}} + \psi(t).$$

After integration, Gronwall-Bellman lemma yields:

$$\forall\, t \in]0,T], \quad \frac{\phi(t)}{t^{m+1}} \leq \int_0^t \psi(s)\,\mathrm{d}s + \int_0^t Ce^{C(t-s)}\int_0^s \psi(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}s.$$

Then, an integration by part gives the desired conclusion. □

Proof of theorem 14.

Step 1. *Parametrization of the problem and notations.*

We denote by $d_0 \leq d$ the number of elements of $J(q_0)$. Consider a local chart $\psi : U \subset Q \to \mathbb{R}^d$ on Q centered at q_0 such that the d_0 first components of $\psi(q)$ are $(\varphi_i(q))_{i\in J(q_0)}$. Such a chart exists since $(\mathrm{d}\varphi_i(q_0))_{i\in J(q_0)}$ is linearly independent in $T^*_{q_0}Q$ by regularity hypothesis I. Next, choose $\alpha > 0$, sufficiently small to have, for all $t \in [t_0, t_0+\alpha]$,

- $q_{\mathrm{a}}(t) \in U, \quad q(t) \in U,$
- $\forall\, i \in J(q_0), \quad \dfrac{\mathrm{d}}{\mathrm{d}t}\varphi_i(q_{\mathrm{a}}(t)) = \langle \mathrm{d}\varphi_i(q_{\mathrm{a}}(t)), \dot{q}_{\mathrm{a}}(t)\rangle_{q_{\mathrm{a}}(t)} \leq 0, \quad (1.17)$
- $\forall\, i \in \{1,2,\cdots,n\} \setminus J(q_0), \quad \varphi_i(q_{\mathrm{a}}(t)) < 0, \quad \varphi_i(q(t)) < 0. (1.18)$

Such a choice for α is possible because the functions $\varphi_i(q_{\mathrm{a}}(t))$ are analytic and the functions $\varphi_i(q(t))$ are continuous. We denote by f_i the components of f in the natural basis (e^i) associated with the chart under consideration. Since q_{a} and q are local solutions for problem III, we have, for all $i \in \{1,2,\cdots,n\}$,

$$g_{ij}(q_{\mathrm{a}})\left(\ddot{q}^j_{\mathrm{a}} + \Gamma^j_{kl}(q_{\mathrm{a}})\dot{q}^k_{\mathrm{a}}\dot{q}^l_{\mathrm{a}}\right) = f_i(q_{\mathrm{a}}, \dot{q}_{\mathrm{a}}; t) + \lambda_{\mathrm{a}i}, \quad (1.19)$$

$$g_{ij}(q)\left(\mathrm{d}\dot{q}^{+j} + \Gamma^j_{kl}(q)\dot{q}^{+k}\dot{q}^{+l}\mathrm{d}t\right) = f_i(q, \dot{q}^+; t)\,\mathrm{d}t + \mu_i, \quad (1.20)$$

where the λ_{ai} and μ_i are respectively d nonpositive analytic functions on $[t_0, t_0+\alpha]$ and d nonpositive measures on $[t_0, t_0+\alpha]$. Note incidentally that the λ_{ai} and μ_i vanish identically for $i > d_0$, by (1.18). We denote by $|.|$ the standard euclidean norm on $\mathbb{R}^d$. Confusing (abusively) q and $\psi(q)$, we shall write:

$$|q|^2 = \sum_{i=1}^{d} \left(q^i\right)^2, \quad \text{and} \quad \left|\dot{q}^+\right|^2 = \sum_{i=1}^{d} \left(\dot{q}^{+i}\right)^2.$$

Step 2. *There exist some positive real constants C_1 and C_2 such that the following estimate:*

$$\int_{t_0}^{t} \left(|q-q_a|^2(s) + \left|\dot{q}^+ - \dot{q}_a\right|^2(s)\right) \mathrm{d}s \leq$$

$$- C_1 \int_{t_0}^{t} e^{C_2(t-s)} \int_{t_0}^{s} \sum_{i=1}^{d_0} \lambda_{ai}(\sigma)\, \dot{q}^{+i}(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s. \quad (1.21)$$

holds for all $t \in [t_0, t_0+\alpha]$.

By proposition 41 of Appendix A, we have:

$$\mathrm{d}\left(\frac{1}{2}\left(\dot{q}^{+i} - \dot{q}_a^i\right) g_{ij}(q) \left(\dot{q}^{+j} - \dot{q}_a^j\right)\right) =$$

$$\left(\frac{\dot{q}^{-i} + \dot{q}^{+i}}{2} - \dot{q}_a^i\right) g_{ij}(q) \left(\mathrm{d}\dot{q}^{+j} - \ddot{q}_a^j\, \mathrm{d}t + \Gamma_{kl}^j(q)\dot{q}^{+k}\left(\dot{q}^{+l} - \dot{q}_a^l\right) \mathrm{d}t\right),$$

and, therefore, using equation of motion (1.20),

$$\mathrm{d}\left(\frac{1}{2}\left(\dot{q}^{+i} - \dot{q}_a^i\right) g_{ij}(q) \left(\dot{q}^{+j} - \dot{q}_a^j\right)\right) =$$

$$\left(\dot{q}^{+i} - \dot{q}_a^i\right) f_i(q, \dot{q}^+; t)\mathrm{d}t - \left(\dot{q}^{+i} - \dot{q}_a^i\right) g_{ij}(q) \left(\ddot{q}_a^j + \Gamma_{kl}^j(q)\dot{q}^{+k}\dot{q}_a^l\right) \mathrm{d}t$$

$$+ \sum_{k=1}^{d_0} \left(\frac{\dot{q}^{-k} + \dot{q}^{+k}}{2} - \dot{q}_a^k\right) \mu_k.$$

But, each $\dot{q}_a^k \mu_k$ is a nonnegative measure by (1.17), and,

$$\sum_{k=1}^{d_0} \frac{\dot{q}^{-k} + \dot{q}^{+k}}{2} \mu_k = \left\langle \frac{\dot{q}^- + \dot{q}^+}{2}, R \right\rangle_q$$

is a nonpositive real measure by proposition 11. Therefore,

$$\mathrm{d}\left(\frac{1}{2}\left(\dot{q}^{+i} - \dot{q}_a^i\right) g_{ij}(q) \left(\dot{q}^{+j} - \dot{q}_a^j\right)\right) \leq$$

$$\left[\left(\dot{q}^{+i} - \dot{q}_a^i\right) f_i(q, \dot{q}^+; t) - \left(\dot{q}^{+i} - \dot{q}_a^i\right) g_{ij}(q) \left(\ddot{q}_a^j + \Gamma_{kl}^j(q)\dot{q}^{+k}\dot{q}_a^l\right)\right] \mathrm{d}t,$$

in the sense of ordering of real measures. Integrating over $]t_0, t]$ ($t \in [t_0, t_0+\alpha]$), we get:

$$\frac{1}{2}\left(\dot{q}^{+i} - \dot{q}_{\mathrm{a}}^{i}\right) g_{ij}(q)\left(\dot{q}^{+j} - \dot{q}_{\mathrm{a}}^{j}\right) \leq$$
$$\int_{t_0}^{t}\left[\left(\dot{q}^{+i} - \dot{q}_{\mathrm{a}}^{i}\right) f_i(q, \dot{q}^{+}; s) - \left(\dot{q}^{+i} - \dot{q}_{\mathrm{a}}^{i}\right) g_{ij}(q)\left(\ddot{q}_{\mathrm{a}}^{j} + \Gamma_{kl}^{j}(q)\dot{q}^{+k}\dot{q}_{\mathrm{a}}^{l}\right)\right] \mathrm{d}s.$$

The term within the integral sign is an analytic function of the three variables q, $\dot{q}^{+}$ and s. Therefore, it is also an analytic function of the three variables $q - q_{\mathrm{a}}$, $\dot{q}^{+} - \dot{q}_{\mathrm{a}}$ and s. It is written under the form:

$$\left(\dot{q}^{+i} - \dot{q}_{\mathrm{a}}^{i}\right) F_i(q - q_{\mathrm{a}}, \dot{q}^{+} - \dot{q}_{\mathrm{a}}; s).$$

But, each function F_i can be decomposed under the form:

$$F_i(q - q_{\mathrm{a}}, \dot{q}^{+} - \dot{q}_{\mathrm{a}}; s) = F_i(0, 0; s) + G_i(q - q_{\mathrm{a}}, \dot{q}^{+} - \dot{q}_{\mathrm{a}}; s),$$

where the G_i are analytic and $G_i(0, 0; s) \equiv 0$. Hence, there exists a positive constant M such that, for all $t \in [t_0, t_0 + \alpha]$,

$$\left|G_i(q(s) - q_{\mathrm{a}}(s), \dot{q}^{+}(s) - \dot{q}_{\mathrm{a}}(s); s)\right| \leq$$
$$M\sqrt{|q(s) - q_{\mathrm{a}}(s)|^2 + |\dot{q}^{+}(s) - \dot{q}_{\mathrm{a}}(s)|^2}$$

Hence, we have proved:

$$\frac{1}{2}\left(\dot{q}^{+i} - \dot{q}_{\mathrm{a}}^{i}\right) g_{ij}(q)\left(\dot{q}^{+j} - \dot{q}_{\mathrm{a}}^{j}\right) \leq$$
$$\int_{t_0}^{t}\left\{\left(\dot{q}^{+i} - \dot{q}_{\mathrm{a}}^{i}\right)\left[f_i(q_{\mathrm{a}}, \dot{q}_{\mathrm{a}}; s) - g_{ij}(q_{\mathrm{a}})\left(\ddot{q}_{\mathrm{a}}^{j} + \Gamma_{kl}^{j}(q_{\mathrm{a}})\dot{q}_{\mathrm{a}}^{k}\dot{q}_{\mathrm{a}}^{l}\right)\right]\right.$$
$$\left. + Md\left|\dot{q}^{+} - \dot{q}_{\mathrm{a}}\right|\sqrt{|q - q_{\mathrm{a}}|^2 + |\dot{q}^{+} - \dot{q}_{\mathrm{a}}|^2}\right\} \mathrm{d}s.$$

Moreover, by a compactness argument, there exists a positive constant m such that for all $t \in [t_0, t_0 + \alpha]$,

$$\frac{1}{2}\left(\dot{q}^{+i} - \dot{q}_{\mathrm{a}}^{i}\right) g_{ij}(q)\left(\dot{q}^{+j} - \dot{q}_{\mathrm{a}}^{j}\right) \geq m\left|\dot{q}^{+} - \dot{q}_{\mathrm{a}}\right|^2.$$

We obtain :

$$\left|\dot{q}^{+} - \dot{q}_{\mathrm{a}}\right|^2(t) \leq \frac{Md}{m}\int_{t_0}^{t}\left(|q - q_{\mathrm{a}}|^2(s) + \left|\dot{q}^{+} - \dot{q}_{\mathrm{a}}\right|^2(s)\right) \mathrm{d}s$$
$$- \frac{1}{m}\int_{t_0}^{t}\sum_{i=1}^{d_0}\lambda_{\mathrm{a}i}(s)\left(\dot{q}^{+i} - \dot{q}_{\mathrm{a}}^{i}\right) \mathrm{d}s,$$

where equation of motion (1.19) has been used. Note that, actually:

$$\forall i \in \{1, 2, \cdots, d_0\}, \quad \lambda_{ai} q_a^i \equiv 0,$$

and, so, by the analyticity of functions q_a^i and λ_{ai}:

$$\forall i \in \{1, 2, \cdots, d_0\}, \quad \lambda_{ai} \dot{q}_a^i \equiv 0.$$

By use of Cauchy-Schwarz inequality, we get:

$$|q - q_a|^2(t) + \left|\dot{q}^+ - \dot{q}_a\right|^2(t) \leq$$
$$\left(\frac{Md}{m} + \alpha\right) \int_{t_0}^t \left(|q - q_a|^2(s) + \left|\dot{q}^+ - \dot{q}_a\right|^2(s)\right) \mathrm{d}s - \frac{1}{m} \int_{t_0}^t \sum_{i=1}^{d_0} \lambda_{ai} \dot{q}^{+i} \mathrm{d}s,$$

Defining:

$$C_1 = \frac{1}{m}, \qquad C_2 = \frac{Md}{m} + \alpha,$$

multiplying each term of the above inequality by $e^{-C_2 t}$ and integrating, we obtain estimate (1.21).

Step 3. *Estimate (1.21) implies that the function* $t \mapsto \sum_{i=1}^{d_0} \lambda_a^i(t)\, \dot{q}^{+i}(t)$ *vanishes identically on a right neighbourhood of* t_0

Indeed, by estimate (1.21):

$$\forall t \in [t_0, t_0 + \alpha], \quad \int_{t_0}^t e^{-C_2 s} \int_{t_0}^s \sum_{i=1}^{d_0} \lambda_{ai}(\sigma)\, \dot{q}^{+i}(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s \leq 0,$$

which is, after integration by parts:

$$\int_{t_0}^t e^{-C_2 s} \sum_{i=1}^{d} \lambda_{ai}(s)\, q^i(s)\, \mathrm{d}s \leq \int_{t_0}^t e^{-C_2 s} \int_{t_0}^s \sum_{i=1}^{d_0} q^i(\sigma)\, \dot{\lambda}_{ai}(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s. \tag{1.22}$$

But, since,

$$\forall i \in \{1, 2, \cdots, d_0\}, \quad \forall s \in [t_0, t_0 + \alpha], \quad \lambda_{ai}(s) \leq 0 \text{ and } q^i(s) \leq 0,$$

the two members of inequality (1.22) are nonnegative and, therefore, the inequality is preserved when taking the absolute value of each member. We get:

$$\begin{aligned} \int_{t_0}^t e^{-C_2 s} \sum_{i=1}^{d_0} \lambda_{ai}(s)\, q^i(s)\, \mathrm{d}s &\leq \int_{t_0}^t e^{-C_2 s} \int_{t_0}^s \sum_{i=1}^{d_0} \left|q^i(\sigma)\right| \left|\dot{\lambda}_{ai}(\sigma)\right| \mathrm{d}\sigma\, \mathrm{d}s, \\ &\leq \int_{t_0}^t \int_{t_0}^s e^{-C_2 \sigma} \sum_{i=1}^{d_0} \left|q^i(\sigma)\right| \left|\dot{\lambda}_{ai}(\sigma)\right| \mathrm{d}\sigma\, \mathrm{d}s. \end{aligned}$$

Define:

$$\begin{aligned} Q^i(s) &= -e^{-C_2(s+t_0)} q^i(s+t_0), \\ L^i(s) &= -\lambda_{ai}(s+t_0). \end{aligned}$$

With these notations, we obtain:

$$\forall\, t \in [0,\alpha], \quad \int_0^t \sum_{i=1}^{d_0} L^i(s)\, Q^i(s)\, \mathrm{d}s \leq \int_0^t \int_0^s \sum_{i=1}^{d_0} \left|\dot{L}^i(s)\right| Q^i(s)\, \mathrm{d}\sigma\, \mathrm{d}s, \tag{1.23}$$

where the L^i are nonnegative real-analytic functions and the Q^i are nonnegative continuous functions which all vanish at $t = 0$ and are right-differentiable at $t = 0$. We are going to prove that inequality (1.23) implies that:

$$\exists\, \beta \in]0,\alpha], \quad \forall\, t \in [0,\beta], \quad \forall\, i \in \{1,2,\cdots,d_0\}, \quad L^i(t)\, Q^i(t) = 0.$$

The functions L^i being analytic nonnegative, there exist nonnegative integers $n_1 < n_2 < \cdots < n_m$, a partition $I_1, I_2, \cdots, I_m$ of $\{1,2,\cdots,d_0\}$, and analytic nonnegative functions G^i such that:

$$\forall\, k \in \{1,2,\cdots,m\}, \quad \forall\, i \in I_k, \qquad L^i(s) = s^{n_k} G^i(s),$$

with either $G^i(0) > 0$ or $G^i \equiv 0$. Inequality (1.23) may be rewritten as:

$$\begin{aligned} &\int_0^t \sum_{k=1}^{m} \sum_{i \in I_k} \sigma^{n_k} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma \leq \\ &\qquad \int_0^t \int_0^s \sum_{k=1}^{m} \sum_{i \in I_k} n_k \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s \\ &\qquad\qquad + \int_0^t \int_0^s \sum_{k=1}^{m} \sum_{i \in I_k} \sigma^{n_k} \left|\dot{G}^i(\sigma)\right| Q^i(\sigma)\, d\sigma\, \mathrm{d}s. \end{aligned}$$

But, by the analyticity of the functions G^i:

$$\exists\, \beta > 0, \quad \exists\, N > 0, \quad \forall\, i \in J(q_0), \quad \forall\, \sigma \in [0,\beta], \qquad \left|\dot{G}^i(\sigma)\right| \leq N G^i(\sigma).$$

Therefore, for all $t \in [0, \beta]$,

$$\int_0^t \sum_{k=1}^{m} \sum_{i \in I_k} \sigma^{n_k} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma \leq$$
$$\int_0^t \int_0^s \sum_{k=1}^{m} \sum_{i \in I_k} n_k \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s$$
$$+ Nt \int_0^t \int_0^s \sum_{k=1}^{m} \sum_{i \in I_k} \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s.$$

Integrating by parts the left member of the inequality, we obtain:

$$t \int_0^t \sum_{k=1}^{m} \sum_{i \in I_k} \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma \leq$$
$$\int_0^t \int_0^s \sum_{k=1}^{m} \sum_{i \in I_k} (n_k + 1) \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s$$
$$+ Nt \int_0^t \int_0^s \sum_{k=1}^{m} \sum_{i \in I_k} \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s. \quad (1.24)$$

Since each function $G^i(\sigma) Q^i(\sigma)/\sigma$ is bounded over $[0, \beta]$, there exists a nonnegative real constant H such that, for all $k \in \{1, 2, \cdots, m\}$ and for all $t \in [0, \beta]$,

$$\int_0^t \int_0^s \sum_{i \in I_k} \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s \leq H t^{n_k + 2}.$$

Since it can be assumed that $\beta < 1$, inequality (1.24) gives, for all $t \in [0, \beta]$,

$$t \int_0^t \sum_{i \in I_1} \sigma^{n_1 - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma \leq$$
$$(1 + n_1 + Nt) \int_0^t \int_0^s \sum_{i \in I_1} \sigma^{n_1 - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s + H_1 t^{n_2 + 2},$$

where H_1 is a non negative real constant. Applying lemma 15, we get:

$$\int_0^t \int_0^s \sum_{i \in I_1} \sigma^{n_1 - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s = O(t^{n_2 + 2}).$$

Coming back to inequality (1.24), we get, for all $t \in [0, \beta]$,

$$t \int_0^t \sum_{k=1}^{2} \sum_{i \in I_k} \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma \leq$$

$$(1 + n_2 + Nt) \int_0^t \int_0^s \sum_{k=1}^{2} \sum_{i \in I_k} \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s + H_2 t^{n_3 + 2}.$$

Applying once more lemma 15, we obtain:

$$\int_0^t \int_0^s \sum_{k=1}^{2} \sum_{i \in I_k} \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s = O(t^{n_3 + 2}).$$

Proceeding inductively, we obtain:

$$\int_0^t \int_0^s \sum_{k=1}^{m-1} \sum_{i \in I_k} \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s = O(t^{n_m + 2}).$$

But, by inequality (1.24), for all $t \in [0, \beta]$,

$$t \int_0^t \sum_{k=1}^{m} \sum_{i \in I_k} \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma \leq$$

$$(1 + n_m + Nt) \int_0^t \int_0^s \sum_{k=1}^{m} \sum_{i \in I_k} \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s.$$

Using lemma 15 for the last time, we get:

$$\forall\, t \in [0, \beta], \quad \int_0^t \int_0^s \sum_{k=1}^{m} \sum_{i \in I_k} \sigma^{n_k - 1} G^i(\sigma)\, Q^i(\sigma)\, \mathrm{d}\sigma\, \mathrm{d}s = 0,$$

which implies:

$$\forall\, i \in \{1, 2, \cdots, d_0\}, \quad \forall\, t \in [0, \beta], \qquad G^i(t)\, Q^i(t) = 0,$$

which is nothing but:

$$\forall\, i \in \{1, 2, \cdots, d_0\}, \quad \forall\, t \in [t_0, t_0 + \beta], \qquad \lambda_{ai}(t)\, \dot{q^i}(t) = 0.$$

But, the analyticity of the functions λ_{ai} implies:

$$\forall\, i \in \{1, 2, \cdots, d_0\}, \quad \forall\, t \in [t_0, t_0 + \beta], \qquad \lambda_{ai}(t)\, \dot{q}^{+i}(t) = 0,$$

and the assertion of step 3 is proved.

Step 4. *Conclusion of the proof of local uniqueness.*

Bringing together the results of steps 2 and 3, we get:

$$\forall t \in [t_0, t_0 + \beta], \quad \int_{t_0}^{t} \left(|q - q_a|^2 (s) + \left|\dot{q}^+ - \dot{q}_a\right|^2 (s) \right) \mathrm{d}s \leq 0,$$

which yields the desired conclusion:

$$\forall t \in [t_0, t_0 + \beta], \quad q(t) = q_a(t).$$

□

Corollary 16 *There exists a unique maximal solution for problem III.*

It was noticed above that the analytical solution for problem III furnished by corollary 13 stops to exist at the first instant of impact. To overcome this fact, we have proved that local uniqueness still holds in the wider class of motion *MMA* which allows impacts. But, this does not guarantee that the maximal solution for problem III is not going to stop to exist at finite time for unphysical reasons. In other terms, we still do not know if the class *MMA* is wide enough. Actually, it is wide enough as shown by the following theorem which should be brought aside theorem 3.

Theorem 17 *The configuration manifold Q is assumed to be a complete Riemannian manifold and the mapping f is supposed to admit the following estimate:*

$$\forall (q, v) \in TQ, \quad \textit{for almost all } t \in [t_0, +\infty[, \qquad \|f(q, v; t)\|_q^* \leq l(t) \left(1 + d(q, q_0) + \|v\|_q\right),$$

where $d(\cdot, \cdot)$ is the Riemannian distance and $l(t)$, a (necessarily nonnegative) function of $L^1_{loc}(\mathbb{R}; \mathbb{R})$.

Then, the dynamics is eternal, that is, the maximal solution for problem III is defined on $[t_0, +\infty[$.

Proof. We proceed as for the proof of theorem 3. We assume that the maximal solution q is defined on $[t_0, T[$, with T finite and try to obtain a contradiction.

Step 1. *The function $t \mapsto \|\dot{q}^+(t)\|_{q(t)}$ is bounded over $[t_0, T[$:*

$$\exists V > 0, \quad \forall t \in [t_0, T[, \qquad \left\|\dot{q}^+(t)\right\|_{q(t)} \leq V. \tag{1.25}$$

By proposition 11, we have:

$$\frac{1}{2} \left\|\dot{q}^+(t)\right\|_{q(t)}^2 \leq \frac{1}{2} \|v_0\|_{q_0}^2 + \int_{t_0}^{t} \left\|f(q(s), \dot{q}^+(s); s)\right\|_{q(s)}^* \left\|\dot{q}^+(s)\right\|_{q(s)} \mathrm{d}s.$$

Applying lemma 5, we get:

$$\left\|\dot{q}^+(t)\right\|_{q(t)} \leq \|v_0\|_{q_0} + \int_{t_0}^{t} \left\|f(q(s), \dot{q}^+(s); s)\right\|_{q(s)}^* \mathrm{d}s,$$

which yields, by virtue of the hypotheses of the theorem:

$$\left\|\dot{q}^+(t)\right\|_{q(t)} \leq \|v_0\|_{q_0} + \int_{t_0}^{t} l(s) \left(1 + d(q(s), q_0) + \left\|\dot{q}^+(s)\right\|_{q(s)}\right) \mathrm{d}s.$$

But,

$$\forall\, t \in [t_0, T[, \quad d(q(t), q_0) \leq \int_{t_0}^{t} \left\|\dot{q}^+(s)\right\|_{q(s)} \mathrm{d}s,$$

therefore,

$$d(q(t), q_0) + \left\|\dot{q}^+(t)\right\|_{q(t)} \leq$$
$$\|v_0\|_{q_0} + \int_0^t l(s)\,\mathrm{d}s + \int_{t_0}^{t} (1 + l(s)) \left(d(q(s), q_0) + \left\|\dot{q}^+(s)\right\|_{q(s)}\right) \mathrm{d}s.$$

By Gronwall-Bellman lemma (lemma 4), we have, for all $t \in [t_0, T[$:

$$d(q(t), q_0) + \left\|\dot{q}^+(t)\right\|_{q(t)} \leq \left(\|v_0\|_{q_0} + \int_0^t l(s)\,\mathrm{d}s\right) e^{\int_{t_0}^{t} (1+l(s))\,\mathrm{d}s},$$

which yields (1.25).

Step 2. *The right velocity $\dot{q}^+$ has bounded variation over $[t_0, T[$:*

$$\mathrm{Var}\left(\dot{q}^+, [t_0, T[\right) < \infty. \tag{1.26}$$

By step 1, we have:

$$\forall\, s_1, s_2 \in [t_0, T[, \quad d\left(q(s_1), q(s_2)\right) \leq V \left|s_2 - s_1\right|.$$

Since Q is assumed to be complete, we deduce that:

$$q_T = \lim_{t \to T^-} q(t)$$

exists in TQ. We denote by d_T the number of elements of $J(q_T)$. Let (U, ψ) be a local chart on Q at q_T such that the d_T first components of $\psi(q)$ in $\mathbb{R}^d$ are $(\varphi_i(q))_{i \in J(q_T)}$. Consider a compact neighbourhood K of q_T in Q such that:

- $K \subset U$,
- $\forall\, q \in K, \quad J(q) \subset J(q_T)$.

Define:

$$t_0' = \min\{t \in [t_0, T[\ ;\ \forall\, s \in [t, T[,\quad q(s) \in K\}\,.$$

Since $[t_0, t_0']$ is compact, one has:

$$\mathrm{Var}\left(\dot{q}^+; [t_0, t_0']\right) < \infty,$$

therefore, it remains only to prove:

$$\mathrm{Var}\left(\dot{q}^+;]t_0', T[\right) < \infty.$$

Denote by $\lambda^{\max}$ (resp. $\lambda^{\min}$) the maximum (resp. the minimum) of the greatest (resp. least) eigenvalue of the matrix $(g_{ij}(q))_{i,j=1,2,\cdots,d}$ when q wanders in K. It is readily seen that:

$$\forall\, i \in \{1, 2, \cdots, d\}\,,\quad \forall\, t \in [t_0', T[,\qquad \begin{aligned} &\left|g_{ij}(q(t))\dot{q}^{+j}(t)\right| \le \sqrt{\lambda^{\max}}V, \\ &\left|\dot{q}^{+i}(t)\right| \le \frac{V}{\sqrt{\lambda^{\min}}}. \end{aligned} \tag{1.27}$$

We denote by $B_q(0, V)$ the closed ball of T_qQ with radius V and centered at the origin. Considering the following compact subset K' of TQ:

$$K' = \bigcup_{q \in K} B_q(0, V),$$

we define the following nonnegative real constants:

$$F = \max_{\substack{(q,v;t) \in K' \times [t_0', T], \\ i \in \{1,2,\cdots,d\}}} \left|f_i(q, v; t)\right|,$$

and:

$$G = \max_{\substack{i,j,k \in \{1,2,\cdots,d\}, \\ q \in K}} \left|\frac{\partial g_{ij}(q)}{\partial q^k}\right|.$$

Writing the equation of motion (1.9) in the local chart (U, ψ), we obtain:

$$\forall\, i \in \{1, 2, \cdots, d\}\,,$$
$$g_{ij}(q)\left(\mathrm{d}\dot{q}^{+j} + \Gamma^j_{kl}(q)\dot{q}^{+k}\dot{q}^{+l}\,\mathrm{d}t\right) = f_i(q, \dot{q}^+; t)\,\mathrm{d}t + \lambda_i,$$

where the λ_i are d nonpositive real measures on $]t_0'; T[$. Expressing the Christoffel symbols in terms of the metric, we have:

$$g_{ij}(q)\mathrm{d}\dot{q}^{+j} + \frac{\partial g_{ij}(q)}{\partial q^k}\dot{q}^{+j}\dot{q}^{+k}\mathrm{d}t - \frac{1}{2}\frac{\partial g_{kl}(q)}{\partial q^i}\dot{q}^{+k}\dot{q}^{+l}\mathrm{d}t = f_i(q, \dot{q}^+; t)\mathrm{d}t + \lambda_i, \tag{1.28}$$

or, equivalently,

$$\mathrm{d}\left(g_{ij}(q)\dot{q}^{+j}\right) = \frac{1}{2}\frac{\partial g_{kl}(q)}{\partial q^i}\dot{q}^{+k}\dot{q}^{+l}\,\mathrm{d}t + f_i(q,\dot{q}^+;t)\,\mathrm{d}t + \lambda_i.$$

We deduce:

$$\begin{aligned}\int_{]s_1,s_2]}(-\lambda_i) &= g_{ij}(q(s_1))\dot{q}^{+j}(s_1) - g_{ij}(q(s_2))\dot{q}^{+j}(s_2) \\ &\quad + \int_{s_1}^{s_2}\left(f_i(q,\dot{q}^+;t) + \frac{1}{2}\frac{\partial g_{kl}(q)}{\partial q^i}\dot{q}^{+k}\dot{q}^{+l}\right)\mathrm{d}t \\ &\leq 2\sqrt{\lambda^{\max}}V + \left(F + \frac{d^2GV^2}{2\lambda^{\min}}\right)(s_2 - s_1),\end{aligned}$$

for all $i \in \{1,2,\cdots,d\}$ and all $s_1, s_2 \in [t_0', T[$ with $s_1 < s_2$. There results that the λ_i are d *bounded* measures on $]t_0', T[$. Thanks to equation (1.28), it is readily seen that the measures $\mathrm{d}\dot{q}^{+i}$ are also bounded measures on $]t_0', T[$. Therefore, the d functions $\dot{q}^{+i} :]t_0', T[\to \mathbb{R}$ have bounded variation over the interval $]t_0', T[$. Then, corollary 36 of Appendix A yields the desired result.

Step 3. *Conclusion of the proof of theorem 17.*

By Steps 1 and 2 and by proposition 38 of Appendix A,

$$(q_T, v_T^-) = \lim_{t \to T^-}(q(t), \dot{q}^+(t))$$

exists in TQ. Define:

$$v_T = \mathcal{F}\left(q_T, v_T^-\right)$$

Take it as a new initial condition at $t = T$. Then, theorem 13 furnishes $T' > T$ and an extension of q on $[T, T'[$ such that $q \in MMA\,([t_0, T'[; Q)$ is a solution of problem III. But, this contradicts the definition of T. □

3.4 Illustrative examples and comments

It is readily seen that the function q displayed in counter-example 6 is the unique maximal solution of problem III corresponding to the situation under consideration. This solution exhibits an accumulation of impacts on the *left* side of instant $t = 3$. However, as predicted by corollary 13, for each instant $t \in \mathbb{R}^+$, there exists a right neighbourhood $[t, t+\eta[$ of t, such that the restriction of q to $[t, t+\eta[$ is analytic. A straightforward and general consequence of this is the following.

Proposition 18 *Let q be the maximal solution of problem III furnished by corollary 16. Although infinitely many impacts can accumulate at the* left *of a given instant, such an accumulation of impacts can* never *occur at the* right *of any*

instant. Moreover, in the particular case where the impact constitutive equation is elastic, the instants of impact are isolated *and therefore in finite number in any compact interval of time.*

Proof. Since for each instant $t \in [t_0, T[$, there exists a right neighbourhood $[t, t+\eta[$ of t, such that the restriction of q to $[t, t+\eta[$ is analytic, we get the first part of the proposition. For the second part, let τ be an arbitrary instant in $]t_0, T[$ and consider the problem III associated with the initial condition $(q(\tau), -\dot{q}^-(\tau))$, the elastic constitutive impact equation and the force mapping $g(q, v; t)$ defined by:

$$g(q, v; t) = f(q, -v; \tau - t)$$

which is analytic. By theorem 14, there exists an analytic function $q_a : [0, T_a[\to Q$ which is a solution of this problem III. Any other solution of problem III coincides with q_a on a right neighbourhood of $t = 0$. Actually, as seen in the proof of local uniqueness (theorem 14), a little bit more is proved: any function $q' \in MMA([0, T[; Q)$ satisfying the initial condition, the unilateral constraint, the equation of motion (1.9) and the energy inequality (proposition 11) has to coincide with q_a on a right neighbourhood of $t = 0$. But, it is readily seen that the function defined by:

$$q'(t) = q(\tau - t), \quad t \in [0, \tau - t_0[$$

fulfill these requirements. Thus, q' can not have right accumulation of impacts at $t = \tau$ and, therefore, q can not have left accumulation of impacts at $t = \tau$ and the instants of impact are isolated. Of course, if q is the maximal solution defined on $[t_0, T[$, impacts can still accumulate at the left of T, as seen on simple examples. □

The fact that infinitely many impacts can accumulate at the left of a given instant but not at the right is a specific feature of the analytical setting that is lost in the C^∞ setting as seen in counterexample 7. Actually, this counterexample shows that pathologies of nonuniqueness in the C^∞ setting are intimately connected to the possibility of right accumulations of impacts. The fact that the analytical setting prevents from such right accumulations is the thorough reason why we could prove uniqueness in this case.

We conclude this section by a come back to the double pendulum of example 4. The aim of the following example is to illustrate the generality of the above theory.

Example 8. Consider the double pendulum described in example 4 and add a rigid obstacle on the vertical coordinate axis as represented on Figure 1.5. This obstacle may be represented by two analytic functions whose expressions in

the global chart of Q described in example 4 are:

$$\begin{aligned}\varphi_1(q^1, q^2) &= -l_1 \sin q^1 \leq 0,\\ \varphi_2(q^1, q^2) &= -l_1 \sin q^1 - l_2 \sin q^2 \leq 0.\end{aligned}$$

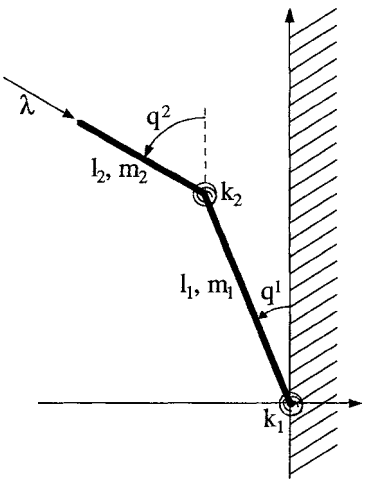

Figure 1.5. Double pendulum with obstacle.

It is readily seen that, except in the particular case where $l_1 = l_2$, these constraints are functionally independent, that is, they satisfy regularity hypothesis I. An arbitrary initial state (q_0, v_0) such that $v_0 \in V(q_0)$ is given at time $t_0 = 0$. To fix ideas, we adopt the canonical constitutive equation with arbitrary restitution coefficient $e(q, \dot{q}^-)$. Then, writing the evolution problem in the chart under consideration is straightforward. By corollary 16, we get a unique maximal solution for this evolution problem. By theorem 17, we can state that this maximal solution is defined all over $\mathbb{R}^+$, that is, the dynamics is eternal.

4. Perfect non-holonomic bilateral constraints

In this section, we come back to simple discrete mechanical systems. Perfect holonomic bilateral constraints were defined to be constraints on the configurations of type:

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

Considering an arbitrary motion satisfying the constraint and differentiating with respect to time, gives:

$$\langle \mathrm{d}\varphi_i(q), \dot{q} \rangle_q = 0.$$

Thus, the constraint may be viewed as acting on the velocity. There are practical situations where the constraint is given in this way. A typical occurence is the "rolling without slipping". Thus, we are led to consider general constraints of type:

$$\langle \alpha_i(q), \dot{q} \rangle_q = 0,$$

where the $\alpha_i(q)$ are cotangent vector fields (we say also 1-form) on the configuration manifolds.

The reason that makes here desirable the study of non-holonomic constraints is that the "rolling without slipping" can be seen as a frictional bilateral constraint with a friction of infinite magnitude. Therefore, this section prepares the full discussion of frictional constraints in the sequel.

4.1 The geometric description

A *non-holonomic bilateral constraint* is a restriction on the admissible motions of the system which is expressed by means of a finite number n of smooth 1-form α_i defined on the configuration manifold:

$$\forall\, i \in \{1, 2, \cdots, n\}, \quad \langle \alpha_i(q), \dot{q} \rangle_q = 0. \tag{1.29}$$

As in the case of holonomic constraints, the constraints are required to be independant in the following sense:

Regularity hypothesis I. For all q in Q, the $\alpha_i(q)$ ($i \in \{1, 2, \cdots, n\}$) are linearly independent in T^*Q.

A straightforward consequence of this hypothesis is that the set E of all admissible velocities:

$$E = \left\{ (q, v) \in TQ \; ; \; \forall\, i \in \{1, 2, \cdots, n\}, \quad \langle \alpha_i(q), v \rangle_q = 0 \right\}, \tag{1.30}$$

is a tangent subbundle of Q (that is, a vector bundle over Q which is also a submanifold of TQ) of dimension $2d - n$.

Of course, the terminology is a little bit confusing (but it is classical) since a non-holonomic constraint may turn out to be holonomic ('holonomic' is greek for 'integrable'). A trivial example is provided in the case $n = 1$ when the 1-form α_1 is *exact* (that is, there exists φ_1 such that $\mathrm{d}\varphi_1 = \alpha_1$). In this case, the non-holonomic constraint is equivalent to the holonomic one: $\varphi_1(q) =$ constant. The constant is determined by the initial configuration q_0. The non-holonomic constraint defined by α_1 may turn out to be holonomic even in the case where α_1 is not exact. Indeed, even if α_1 is not exact, there may exist some real valued function $h(q)$ such that $h(q)\alpha_1(q)$ is exact. We shall say that the non-holonomic constraint defined by the α_i is holonomic if there exist (locally) n real-valued functions φ_i such that (1.29) is equivalent to:

$$\forall\, i \in \{1, 2, \cdots, n\}, \quad \langle \mathrm{d}\varphi_i(q), \dot{q} \rangle_q = 0.$$

Deciding, in the general case, whether a non-holonomic constraint is holonomic or not, is a difficult issue. One answer is provided by Frobenius' theorem (see, for example, ABRAHAM & MARSDEN (1985), p. 93).

Theorem 19 (Frobenius) *The non-holonomic constraint defined by the* α_i *(*$i \in \{1, 2, \cdots, n\}$*) is holonomic if and only if for any two vector fields* X *and* Y *defined on open sets of* Q *and which take values in* E*, the Lie bracket* $[X, Y]$ *takes values in* E *as well.*

Hence, the study of non-holonomic bilateral constraints is more general than the study of holonomic ones, since the former contains formally the latter. However, the handling of holonomic constraints is simpler since it allows immediately the elimination of the redundant parameters in any parametrization. So, each time a non-holonomic constraint turns out to be holonomic, it should be integrated.

4.2 Formulation of the dynamics

Here also, the realization of the constraints necessarily involves some reaction forces R which should be specified through a constitutive assumption.

Constitutive hypothesis II. The non-holonomic bilateral constraint (1.29) is supposed to be *perfect*, that is, the virtual power of the reaction forces R vanishes in any virtual velocity compatible with the bilateral constraint:

$$\forall\, (q, v) \in E, \quad \langle R, v \rangle_q = 0.$$

Hypotheses I and II imply that there exists n real-valued functions λ_i, unique, such that:

$$R(t) = \sum_{i=1}^{n} \lambda_i(t)\, \alpha_i(q).$$

Now, we formulate the evolution problem associated with the dynamics of rigid bodies systems with perfect bilateral constraints, either non-holonomic or holonomic (the holonomic constraint is included in the definition of the configuration manifold Q). The initial condition is assumed to be compatible with the realization of the constraint: $(q_0, v_0) \in E$.

Problem IV. Find $T > t_0$, $q \in C^2([t_0, T[; Q)$ and n functions $\lambda_i \in C^0([t_0, T[; \mathbb{R})$ such that:

- $(q(t_0), \dot{q}(t_0)) = (q_0, v_0)$,
- $\forall\, t \in [t_0, T[, \quad (q(t), \dot{q}(t)) \in E$,
- $\forall\, t \in [t_0, T[, \quad \flat \dfrac{\mathrm{D}}{\mathrm{d}t} \dot{q}(t) = f(q(t), \dot{q}(t), t) + \displaystyle\sum_{i=1}^{n} \lambda_i(t)\, \alpha_i(q(t))$.

4.3 Well-posedness of the dynamics

By similarity with that of Section 2.3, we state the following regularity hypothesis.

Regularity hypothesis III. The configuration manifold Q is of class C^2, the mapping $f : TQ \times \mathbb{R} \to T^*Q$ and the 1-forms α_i are of class C^1.

The fundamental reason why problem IV is well-posed, is, that it reduces to a (first order) ordinary differential equation on TE. To describe how this is realized, we need to introduce some new notations and definitions.

In Section 1.2, it has been stated briefly that the equation of motion:

$$\frac{\mathrm{D}}{\mathrm{d}t}\dot{q}(t) = \sharp \circ f(q(t), \dot{q}(t), t), \tag{1.31}$$

is a second order differential equation on the configuration manifold Q. We are going to express more precisely what is meant by that. Consider a local chart $\psi : U \to \mathbb{R}^d$. With ψ, we associate a natural local chart $\Psi : \Pi_Q^{-1}(U) \to \mathbb{R}^{2d}$ on TQ by:

$$\begin{aligned} \Psi(q, v) &= \left(q^1, \ldots, q^d, v^1, \ldots, v^d\right), \\ &= \left(\psi^1(q), \ldots, \psi^d(q), \langle \mathrm{d}\psi^1(q), v\rangle_q, \ldots, \langle \mathrm{d}\psi^d(q), v\rangle_q\right). \end{aligned}$$

Actually, $\Psi = T\psi$ is nothing but the classical tangent map of ψ (see, for example, ABRAHAM & MARSDEN (1985), p. 45). We find it convenient to write the basis of tangent spaces to TQ at points of $\Pi_Q^{-1}(U)$ by:

$$\left(\frac{\partial}{\partial q^1}, \ldots, \frac{\partial}{\partial q^d}, \frac{\partial}{\partial v^1}, \ldots, \frac{\partial}{\partial v^d}\right)$$

(this notation is standard and expresses the fact that tangents can be viewed as derivation operators on real valued functions and reciprocally). It is easy to write equation (1.31) in the chart ψ as a first order differential equation:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}q^i &= v^i, \\ \frac{\mathrm{d}}{\mathrm{d}t}v^i &= -\Gamma^i_{jk}(q)v^j v^k + g^{ij}(q) f_j(q, v; t). \end{aligned}$$

Hence, the solution of equation (1.31) is nothing but an integral curve (see ABRAHAM & MARSDEN (1985), Section 2.1) of the time-dependent vector field $\mathcal{G}(\cdot; t)$ defined on $\Pi_Q^{-1}(U)$ by:

$$\mathcal{G}(t) = v^i \frac{\partial}{\partial q^i} - \Gamma^i_{jk}(q)v^j v^k \frac{\partial}{\partial v^i} + g^{ij}(q) f_j(q, v; t) \frac{\partial}{\partial v^i}.$$

Since the geodesic equations are independent of the choice of coordinates on Q, we conclude that $\mathcal{G}(\cdot;t)$ defines a global time-dependent vector field on TQ. Now, if $\Omega :]\alpha, \beta[\to TQ$ is any integral curve of $\mathcal{G}(\cdot;t)$ and ω is the curve on Q defined by $\omega = \Pi_Q \circ \Omega$, then it is readily seen that:

$$\frac{\mathrm{d}}{\mathrm{d}t}\omega = \Omega.$$

This last property is easily seen to be equivalent to the following property of $\mathcal{G}$:

$$\forall\, (q,v) \in TQ, \quad T\Pi_Q\,(\mathcal{G}(q,v;t)) = (q,v),$$

and motivates the following definition.

Definition 20 *Let Q be a manifold and E any tangent subbundle of Q. A time-dependent vector field $\mathcal{X}(\cdot;t)$ on E is said to determine a second-order differential equation on Q if:*

$$T\Pi_Q(\mathcal{X}(\cdot;t)) = id_E$$

Now, E will be the tangent subbundle of the configuration manifold defined by formula (1.30). We denote by E_q the fiber over $q \in Q$. We define a map P_E by:

$$P_E \left\{ \begin{array}{ccc} TQ & \to & E \\ (q,v) & \mapsto & \left(q, \mathrm{Proj}_q[v; E_q]\right) \end{array} \right.$$

Recall that $\mathrm{Proj}_q[v; E_q]$ was defined in Section 2.2 to be the orthogonal projection of v on the subspace E_q of T_qQ. The tangent map TP_E of P_E maps the second tangent bundle TTQ of Q onto TE. Thus, $TP_E\,(\mathcal{G}(\cdot;t))$ is a C^1 vector field on E (we have used regularity hypothesis III). It is readily seen that the vector field $TP_E\,(\mathcal{G}(\cdot;t))$ determines a second-order differential equation on Q.

Theorem 21 *Any solution of problem IV defines an integral curve of the time-dependent vector field $TP_E\,(\mathcal{G}(\cdot;t))$ on E and reciprocally.*

Proof. Let $q(t)$ be an arbitrary solution of problem IV. We shall denote by $\Omega(t) = (q(t), \dot{q}(t))$ the corresponding curve in E. We have:

$$\frac{\mathrm{D}}{\mathrm{d}t}\dot{q}(t) = \sharp \circ f(q(t), \dot{q}(t), t) + r(t),$$

where $r : [t_0, T[\to TQ$ is such that, for all t, $r(t) = r^i(t)\,\partial/\partial q^i$ lies in the orthogonal complement of $E_{q(t)}$ in $T_{q(t)}Q$. As a result,

$$\frac{\mathrm{d}}{\mathrm{d}t}\Omega(t) = \mathcal{G}(\Omega(t);t) + \mathcal{R}(t),$$

where $\mathcal{R}(t)$ is the curve in TE which is expressed by $\mathcal{R}(t) = r^i(t)\partial/\partial v^i$ in any local chart. By $P_E(r(t)) = 0$, we get immediately $TP_E(\mathcal{R}(t)) = 0$. Moreover, since $\Omega(t)$ is in E for all t, we have:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Omega(t) &= TP_E\frac{\mathrm{d}}{\mathrm{d}t}\Omega(t)\\ &= TP_E\left(\mathcal{G}(\Omega(t);t)\right) + TP_E(\mathcal{R}(t))\\ &= TP_E\left(\mathcal{G}(\Omega(t);t)\right),\end{aligned}$$

and, therefore, the first part of the proposition. Reciprocally, let Ω be an integral curve of $TP_E\left(\mathcal{G}(\cdot;t)\right)$. We define $q(t)$ by:

$$q(t) = \Pi_Q(\Omega(t))$$

Since $TP_E\left(\mathcal{G}(\cdot;t)\right)$ determines a second-order differential equation on Q, we have:

$$(q(t), \dot{q}(t)) = \Omega(t) \in E.$$

Moreover, we easily have:

$$\forall t, \quad P_E\left(\frac{\mathrm{D}}{\mathrm{d}t}\dot{q}(t) - \sharp \circ f(q(t), \dot{q}(t), t)\right) = 0,$$

which yields the desired result. □

Corollary 22 *Problem IV admits a unique maximal solution q_m. Moreover, if Q is of class C^p ($p \geq 2$), and f and the α_i are of class C^{p-1} then q_m is of class C^p. If Q, f and the α_i are analytic functions then so are q_m and the functions λ_i.*

Similarly to theorem 3, we have:

Theorem 23 *The configuration manifold Q is assumed to be a complete Riemannian manifold and the mapping f is supposed to admit the following estimate:*

$$\forall (q,v) \in TQ, \quad \textit{for almost all } t \in [t_0, +\infty[,$$
$$\|f(q,v;t)\|_q^* \leq l(t)\left(1 + d(q,q_0) + \|v\|_q\right),$$

where $d(\cdot,\cdot)$ is the Riemannian distance and $l(t)$, a (necessarily nonnegative) function of $L^1_{loc}(\mathbb{R};\mathbb{R})$.

Then, the dynamics is eternal, that is, q_m is defined on $[t_0, +\infty[$.

Theorem 23 is proved exactly along the same lines as theorem 3.

4.4 Illustrative example and comments

Example 9. In the usual three-dimensional space, consider a rigid homogeneous ball of radius R and mass M. The center of the ball is constrained to remain at distance R of a given fixed affine plane (perfect holonomic bilateral constraint). The ball is initially at rest and a prescribed punctual force applies at the center of the ball. Also, the ball is constrained to roll without slipping on the plane (perfect non-holonomic bilateral constraint). The holonomic bilateral constraint is taken into account by using the reduced configuration manifold $Q = \mathcal{P} \times \mathbb{SO}3$ where $\mathcal{P}$ is the affine plane containing the center of the ball. There is no global parametrization of that system. As a local chart at the initial configuration, we can use some Cartesian orthonormal coordinates (x, y) in $\mathcal{P}$ and some Euler angles (ψ, θ, ϕ) (the ball is supposed to lie 'above the plane in the z-direction' and the initial configuration has Euler angle $\psi = 0$, $\theta = \pi/2$ and $\phi = 0$) in $\mathbb{SO}3$. The kinetic energy in the considered chart is given by:

$$K(q, \dot{q}) = \frac{M}{2}\left(\dot{x}^2 + \dot{y}^2\right) + \frac{MR^2}{5}\left(\dot{\psi}^2 + \dot{\theta}^2 + \dot{\phi}^2 + 2\cos\theta\,\dot{\psi}\dot{\phi}\right),$$

which provides immediately the components $g_{ij}(q)$ of the kinetic metric on Q. The force mapping $f(t)$ is given by:

$$f(t) = F_x(t)\,\mathrm{d}x + F_y(t)\,\mathrm{d}y,$$

where $F_x(t)$, $F_y(t)$ are the components of the real world force along x, y. The non-holonomic constraint is obtained in the given chart, by writing that the real world velocity of the contact point must vanish. It is readily seen that we need two 1-forms α_1 and α_2 to express this. They are given in the chart under consideration by:

$$\begin{aligned} \alpha_1 &= \mathrm{d}x - R\sin\psi\,\mathrm{d}\theta + R\cos\psi\,\sin\theta\,\mathrm{d}\phi, \\ \alpha_2 &= \mathrm{d}y + R\cos\psi\,\mathrm{d}\theta + R\sin\psi\,\sin\theta\,\mathrm{d}\phi, \end{aligned}$$

which are clearly independent. Using a covering of the manifold Q by such charts, these definitions are easily globalized. Using the results of the present section, it is easy to form the evolution problem associated with the dynamics of this system. Straightforward application of corollary 22 allows to conclude to the existence of a unique maximal motion, provided $F_x(t)$ and $F_y(t)$ are of class C^1. By corollary 23, we have that this maximal motion is defined for all time.

Given two arbitrary configuration q_i and q_f in Q, it can be proved that there is a smooth motion $q(t)$, starting at q_i, ending at q_f and satisfying the non-holonomic constraint at every instant:

$$\forall t, \quad \langle \alpha_1(q(t)), \dot{q}(t) \rangle_{q(t)} = \quad \langle \alpha_2(q(t)), \dot{q}(t) \rangle_{q(t)} = 0.$$

This fact demonstrates that the non-holonomic constraint defined by α_1 and α_2 *is not* holonomic. An alternative way to see it would have been to apply Frobenius theorem (theorem 19).

To conclude this example, let us write the evolution problem in the parametrization described above. We have to find smooth functions $x(t)$, $y(t)$, $\psi(t)$, $\theta(t)$, $\phi(t)$, $\lambda_1(t)$ and $\lambda_2(t)$, satisfying the initial condition and such that:

$$\begin{aligned}
M\ddot{x} &= F_x + \lambda_1, \\
M\ddot{y} &= F_y + \lambda_2, \\
\frac{2MR^2}{5}\left(\ddot{\psi} + \cos\theta\,\ddot{\phi} - \sin\theta\,\dot{\theta}\dot{\phi}\right) &= 0, \\
\frac{2MR^2}{5}\left(\ddot{\theta} + \sin\theta\,\dot{\psi}\dot{\phi}\right) &= -R\sin\psi\,\lambda_1 + R\cos\psi\,\lambda_2, \\
\frac{2MR^2}{5}\left(\ddot{\phi} + \cos\theta\,\ddot{\psi} - \sin\theta\,\dot{\psi}\dot{\theta}\right) &= R\cos\psi\sin\theta\,\lambda_1 + R\sin\psi\sin\theta\,\lambda_2, \\
\dot{x} - R\sin\psi\,\dot{\theta} + R\cos\psi\sin\theta\,\dot{\phi} &= 0, \\
\dot{y} + R\cos\psi\,\dot{\theta} + R\sin\psi\sin\theta\,\dot{\phi} &= 0,
\end{aligned}$$

To solve this system, we can eliminate the unknown functions $\lambda_1(t)$ and $\lambda_2(t)$ in order to get a first order differential equation with unknown $(x, y, \psi, \theta, \phi, \dot{\psi}, \dot{\theta}, \dot{\phi})$. It turns out that this is nothing but particularizing the proof of theorem 21 to the given system with the particular chart under consideration. The intrinsic point-of-view has provided a valuable guide to perform this in a systematic way. Moreover, it has allowed to lighten the notations very much.

Remark 3. A comment similar to remark 2 can be made here. The dynamics of the constrained system depends only on the geometry of the tangent subbundle E and not on the particular choice of the 1-forms α_i used to define it.

5. Non-firm bilateral constraints

In Section 4, we have discussed general perfect bilateral constraints on simple discrete mechanical systems. They are described by means of a finite number n of linearly independent smooth 1-forms α_i defined on the configuration manifold. The reaction forces were seen to have general expression:

$$R(t) = \sum_{i=1}^{n} \lambda_i(t)\,\alpha_i(q),$$

where the λ_i are *a priori* unknown smooth real valued functions of time. Once the evolution problem associated with the dynamics is solved, they are uniquely

determined. Actually, to write the evolution problem associated with the dynamics, we have implicitly assumed that the constraint is firm in the following sense.

Definition 24 *A general perfect bilateral constraint is said to be firm if any value of the associated reaction force can be assumed by the system.*

In some cases, it may turn to be physically relevant to deal with *non-firm* bilateral constraints. This is the object of this section.

5.1 Formulation of the dynamics

We are given an arbitrary simple discrete mechanical system according to definition 7 and a general perfect bilateral constraint defined by n linearly independent smooth 1-forms α_i defined on the configuration manifold Q. The general expression for the reaction force associated with that constraint is given by:

$$R = \sum_{i=1}^{n} \lambda_i \, \alpha_i(q), \tag{1.32}$$

where $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is an arbitrary element of $\mathbb{R}^n$ in the case where the constraint is assumed to be firm. To discuss the case of non-firm constraints, it is natural to introduce a closed convex subset C_0 of $\mathbb{R}^n$, containing the origin, and to require the following restriction for the reaction force:

$$(\lambda_1, \lambda_2, \ldots, \lambda_n) \in C_0.$$

Actually, to get more generality and in view of discussing dry friction, it will be convenient to allow that the convex of admissible reaction forces can depend on time and also on the state. We state in the following constitutive hypothesis, the general form of the dependency that we allow. It will be enough for our purpose.

Constitutive hypothesis I. The bilateral constraint defined by the n linearly independent 1-forms α_i is non-firm in the sense that the associated reaction force $R = \sum_{i=1}^{n} \lambda_i \, \alpha_i(q)$ can not assume values out of the subset $C(q, \dot{q}; t)$ of T_q^*Q defined by:

$$C(q, \dot{q}; t) = \left\{ \sum_{i=1}^{n} \lambda_i \, \alpha_i(q) \; ; \right.$$
$$\left. (\lambda_1, \lambda_2, \ldots, \lambda_n) \in M(q) \cdot \left[C_0 + \sum_{i=1}^{m} \mu_i(q, \dot{q}; t) C_i \right] \right\},$$

where:

- C_0 is a given closed convex subset of $\mathbb{R}^n$, possibly unbounded and containing the origin,
- the C_i $(i = 1, 2 \ldots, m)$ are given *bounded* closed convex subsets of $\mathbb{R}^n$, containing the origin,
- the $\mu_i : TQ \times \mathbb{R} \to \mathbb{R}^+$ are given functions whose regularity will be stated later on,
- $M(q)$ is a given invertible square real matrix of order n, which depends smoothly on q.

It is readily seen that $C(q, \dot{q}; t)$ is a closed convex subset of T_q^*Q which contains the origin.

Naturally, this formalism contains the case of firm constraints as a particular case: take $m = 0$ and $C_0 = \mathbb{R}^n$.

Of course, this restriction on the admissible reaction forces will not be compatible any more, in general, with the kinematical realization of the constraint:

$$\forall\, i \in \{1, 2, \ldots, n\}, \quad \langle \alpha_i(q), \dot{q} \rangle_q = 0, \tag{1.33}$$

which, therefore, should be relaxed. But relaxing the constraint is equivalent to admit some dissipation of energy associated with the reaction force. The following constitutive hypothesis gives precise information on the way the constraint is relaxed.

Constitutive hypothesis II. The non-firm bilateral constraint obeys to the so-called *Principle of Maximal Dissipation*:

$$\forall\, \tilde{R} \in C(q, \dot{q}; t), \quad -\langle R, \dot{q} \rangle_q \geq -\left\langle \tilde{R}, \dot{q} \right\rangle_q .$$

In the particular case where $m = 0$ and $C_0 = \mathbb{R}^n$ (firm bilateral constraint), constitutive hypothesis II implies nothing but the realization (1.33) of the constraint. In the general case, constitutive hypothesis II specifies the way the constraint is relaxed. Hence, the general firm perfect bilateral constraint, as discussed in Section 4, appears as a particular case of non-firm bilateral constraints as discussed in this section. Constitutive hypothesis II can be given a synthetic formulation by use of elementary convex analysis and its notations (see Appendix B):

$$-\dot{q} \in \partial I_{C(q,\dot{q};t)}(R),$$

where ∂I_C is the subdifferential (in the sense of the duality (T_qQ, T_q^*Q)) of the indicator function of the closed convex subset C of T_q^*Q. Introducing the support function $S_{C(q,\dot{q};t)}$ of $C(q, \dot{q}; t)$ (that is, the conjugate or dual function

of $I_{C(q,\dot{q};t)}$ in the duality (T_qQ, T_q^*Q)), we have the equivalent formulation (see Proposition 48 of Appendix B).

$$R \in \partial S_{C(q,\dot{q};t)}(-\dot{q}).$$

Now, given any initial condition $(q_0, v_0) \in TQ$ compatible with the non-firm constraint:

$$-v_0 \in \text{Dom}\, S_{C(q_0,v_0;t_0)}, \tag{1.34}$$

the evolution problem associated with the dynamics of simple discrete mechanical systems subjected to non-firm bilateral constraints is formulated as follows.

Problem V. Find $T > t_0$ and $q \in W^{2,\infty}([t_0, T[; Q)$ such that:

- $(q(t_0), \dot{q}(t_0)) = (q_0, v_0)$,
- $\flat \dfrac{D}{dt}\dot{q}(t) - f(q(t), \dot{q}(t); t) \in \partial S_{C(q(t),\dot{q}(t);t)}(-\dot{q}(t))$.

The reason why we look for solutions in the Sobolev class $W^{2,\infty}$ which is larger than the usual class C^2 will be made clear later on (Section 5.2 and example 10).

5.2 Well-posedness of the dynamics

Regularity hypothesis III. The configuration manifold Q is of class C^2, the mapping $f : TQ \times \mathbb{R} \to T^*Q$, the 1-forms α_i and the mapping $M : q \mapsto M(q)$ are of class C^1. Also, the functions $\mu_i : TQ \times \mathbb{R} \to \mathbb{R}^+$ are locally lipschitzian (in the sense that the representative in a local chart at an arbitrary (q_0, v_0) is locally lipschitzian with respect to $(q, \dot{q}; t) \in \mathbb{R}^{2d+1}$).

Then, we can prove well-posedness for problem V.

Theorem 25 *There exists a solution (T, q) for problem V.*

Proof. First, we are going to write the evolution equation (actually, inclusion) in a local chart. Let (U, ψ) be a local chart on Q at q_0. Also, taking U smaller if necessary, we can complete the $\alpha_i(q)$ so as to get a basis of T_q^*Q at each q. Now, define a new basis $(w^{i*}(q))$ of T_q^*Q by:

$$w^{i*}(q) = \sum_{j=1}^{d} M_{ji}(q)\alpha_j(q),$$

where the matrix $M(q)$, which has been defined as a real matrix of order n, is transformed into a matrix of order d by adding zeroes everywhere except on

the diagonal where we add some ones. Define a basis $(w_i(q))$ of T_qQ to be the dual basis of $(w^{i*}(q))$. If $\dot{q}$ is any element of T_qQ, then we have:

$$\dot{q} = \eta^i w_i(q), \quad \text{with } \eta^i = A_{ij}(q)\dot{q}^j,$$

where $A(q)$ is an invertible real matrix of order d, depending smoothly on q. Its inverse matrix will be denoted by $B(q)$. Hence, the $\mathbb{R}^{2d}$-valued mapping $(q^1, \ldots, q^d, \eta^1, \ldots, \eta^d)$ defines a vector bundle local chart on TQ at (q_0, v_0). We shall write:

$$f(q, \dot{q}; t) = f_i(q, \eta, t) w^{i*}(q),$$

where the f_i are C^1 functions defined on an open set of $\mathbb{R}^{2d+1}$. The C_i which have been defined as closed subsets of $\mathbb{R}^n$ are now seen as closed subsets of $\mathbb{R}^d$. We denote by S_i their support functions which are, thus, defined on $\mathbb{R}^d$. We define some convex functions φ_i $(i = 0, 1, \ldots, m)$ on $\mathbb{R}^{2d}$ by:

$$\varphi_i(q, \eta) = S_i(\eta).$$

We shall keep the same notation for the μ_i and their representatives in the chart (q, η). With these notations, the evolution inclusion takes the following form in the chart under consideration, thanks to propositions 44 and 47 of Appendix B:

$$-G(q) \cdot \begin{pmatrix} \dot{q} \\ -\dot{\eta} \end{pmatrix} - F(q, \eta, t) \in \partial\varphi_0(q, -\eta) + \sum_{i=1}^{m} \mu_i(q, \eta, t) \partial\varphi_i(q, -\eta). \quad (1.35)$$

In evolution inclusion 1.35, we used the following notations.

$$G(q) = \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & {}^tB(q) \cdot g(q) \cdot B(q) \end{pmatrix},$$

where $g(q)$ is the real matrix of order d defined by the $g_{ij}(q)$. It is clear that the real matrix of order $2d$ $G(q)$ is symmetric positive definite for all q. Moreover, it is a C^1 function of the variable q. Also, we have denoted by $F(q, \eta, t)$ the element of $\mathbb{R}^{2d}$ defined by:

$$\begin{aligned} F_i(q, \eta, t) &= -\sum_{j=1}^{d} B_{ij}(q)\eta^j, \\ F_{d+i}(q, \eta, t) &= f_i(q, \eta, t) + B_{ji}(q) g_{jk}(q) \left[\frac{\partial B_{kl}(q)}{\partial q^m} B_{mn}(q) \eta^l \eta^n \right. \\ &\qquad \left. + \Gamma^k_{lm}(q) B_{ln}(q) B_{mo}(q) \eta^n \eta^o \right], \end{aligned}$$

for $i = 1, 2, \ldots, d$. It is clear that the function F is of class C^1. To express the initial condition, we introduce η_0 which is easily expressed in terms of q_0 and v_0.

Next, we are given a positive real number R such that the closed ball $B(q_0, R) \subset \mathbb{R}^d$ is contained in $\psi(U)$. We denote by B the closed ball of $\mathbb{R}^{2d}$ centered at $(q_0, -\eta_0)$ with radius R. Given a function $(\tilde{q}, -\tilde{\eta}) \in W^{1,\infty}(t_0, T; \mathbb{R}^{2d})$, taking values in B, consider the following evolution problem:
Find $(q, -\eta) \in W^{1,\infty}(t_0, T; \mathbb{R}^{2d})$ such that:

- $(q(t_0), -\eta(t_0)) = (q_0, -\eta_0)$,
- for a.e. $t \in [t_0, T[$,

$$-G(\tilde{q}) \cdot \begin{pmatrix} \dot{q} \\ -\dot{\eta} \end{pmatrix} - F(\tilde{q}, \tilde{\eta}, t) \in \partial\varphi_0(q, -\eta) + \sum_{i=1}^{m} \mu_i(\tilde{q}, \tilde{\eta}, t) \partial\varphi_i(q, -\eta).$$

This evolution problem falls exactly into the type of those which are studied in proposition 52 of Appendix B. Hence, it admits a unique solution. Using estimate (1.B.1) (proposition 52 of Appendix B), we can easily construct a $T > t_0$ depending only on R, G, F, φ_0, the μ_i and the φ_i, which ensures that the solution of the above evolution problem takes values in B. In the sequel of the proof, we adopt the notation $u = (q, -\eta)$. We define by induction a sequence of such functions u. First, u_0 is the constant function $(q_0, -\eta_0)$. The function u_1 is defined to be the solution of the above well-posed evolution problem with the choice $(\tilde{q}, -\tilde{\eta}) = u_0$. Going on inductively, we have built a sequence u_N. By use of estimate (1.B.2) of proposition 52 of Appendix B, we prove easily by induction:

$$\forall\, t \in [t_0, T], \quad |u_{N+1}(t) - u_N(t)| \leq \frac{(Ct)^N}{N!} \max_{s \in [t_0, T]} |u_1(s) - u_0(s)|,$$

where $|\cdot|$ is the standard norm on $\mathbb{R}^{2d}$ and C denotes a real constant independent on N. Therefore, the sequence u_N converges towards a limit u in the Banach space $C^0([t_0, T]; \mathbb{R}^{2d})$. Moreover, use of estimate (1.B.1) of proposition 52 of Appendix B together with the definition of T shows that the sequence $\|\dot{u}_N\|_{L^\infty}$ is bounded. Thus, we have $u \in W^{1,\infty}(t_0, T; \mathbb{R}^{2d})$. Also, reproducing the reasoning of the proof of proposition 52 of Appendix B, we can conclude that u solves evolution inclusion (1.35) and so, we have constructed a solution for problem V. □

Theorem 26 *There is local uniqueness for problem V, that is, if* (T_1, q_1) *and* (T_2, q_2) *are two solutions of problem V, then, there exists* $T_0 \leq \min\{T_1, T_2\}$ *such that:*

$$q_{1|[t_0, T_0[} \equiv q_{2|[t_0, T_0[}$$

Proof. We stick to the notations of the proof of theorem 25. The real number $T > t_0$ being defined as in the above proof, define $T_0 = \min\{T, T_1, T_2\}$. The

two solutions q_1 and q_2 define two solutions $u_1 = (q_1, -\eta_1)$ and $u_2 = (q_2, -\eta_2)$ in $W^{1,\infty}(t_0, T_0; \mathbb{R}^{2d})$ of evolution inclusion (1.35). Use of estimate (1.B.2) of proposition 52 of Appendix B yields:

$$\forall t \in [t_0, T_0], \quad |u_2(t) - u_1(t)| \leq C \int_{t_0}^{t} |u_2(s) - u_1(s)| \, \mathrm{d}s,$$

where C is a positive real constant. Now, use of Gronwall lemma (lemma 4) yields the claim. □

Corollary 27 *There exists a unique maximal solution for problem V.*

Proposition 28 (Energy inequality) *Let (T, q) be an arbitrary solution of problem V. Then, we have:*

$$\forall t_1, t_2 \in [t_0, T[, \quad t_1 \leq t_2, \qquad K\left(q(t_2), \dot{q}(t_2)\right) - K\left(q(t_1), \dot{q}(t_1)\right) =$$
$$\frac{1}{2} \left\|\dot{q}(t_2)\right\|^2_{q(t_2)} - \frac{1}{2} \left\|\dot{q}(t_1)\right\|^2_{q(t_1)} \leq \int_{t_1}^{t_2} \left\langle f\left(q(s), \dot{q}(s); s\right), \dot{q}(s) \right\rangle_{q(s)} \mathrm{d}s$$

Proof. For all $w \in \partial S_{C(q(t),\dot{q}(t);t)}(-\dot{q}(t))$,

$$\langle w, \dot{q}(t) \rangle_{q(t)} \leq S_{C(q(t),\dot{q}(t);t)}(0) - S_{C(q(t),\dot{q}(t);t)}(-\dot{q}(t)) \leq 0,$$

since $S_{C(q(t),\dot{q}(t);t)}$ can only take nonnegative values. □

Corollary 29 *The configuration manifold Q is assumed to be a complete Riemannian manifold and the mapping f is supposed to admit the following estimate:*

$$\forall (q, v) \in TQ, \quad \textit{for almost all } t \in [t_0, +\infty[,$$
$$\left\|f(q, v; t)\right\|^*_q \leq l(t) \left(1 + d(q, q_0) + \|v\|_q\right),$$

where $d(\cdot, \cdot)$ is the Riemannian distance and $l(t)$, a (necessarily nonnegative) function of $L^1_{loc}(\mathbb{R}; \mathbb{R})$.

Then, the dynamics is eternal, that is, the maximal solution for problem V is defined on $[t_0, +\infty[$.

5.3 Illustrative examples and comments

Non-firm bilateral constraints have been introduced principally in view of discussing dry friction. This is postponed to next section. However, we shall provide here a simple example where a non-firm constraint appears naturally.

Example 10. Consider the back wheel of a bicycle and its gear. We shall provide a simple model of their assembly in which appears naturally a non-firm constraint according to the above formalism.

Two homogeneous disks, with mass M_1, M_2 and radius R_1, R_2 are constrained to rotate around the same axis, passing through the centers of the disks and perpendicular to their common plane. The configuration manifold is the 2-torus and we shall use the global parametrization defined by the two angular measures (θ_1, θ_2). The kinetic energy is given by:

$$K = \frac{1}{4} M_1 R_1^2 \dot{\theta}_1^2 + \frac{1}{4} M_2 R_2^2 \dot{\theta}_2^2$$

The forces is supposed to be a constant torque Γ_1 applied on the disk 1. This defines a simple discrete mechanical system according to definition 7. Next, we want to describe the fact that the relative velocity of disk 1 with respect to disk 2 has constant sign. To do this, introduce the *non-firm* bilateral constraint defined by the 1-form $\alpha_1 = \mathrm{d}\theta_1 - \mathrm{d}\theta_2$. Sticking to the notations of the beginning of the present section, choose $m = 0$, $C_0 = \mathbb{R}^-$ and $M(q) = \mathrm{Id}$. The corresponding evolution problem V can be written in the parametrization under consideration in the following manner.

Find $\theta_1, \theta_2 \in W^{2,\infty}(0, T; \mathbb{R})$ and $\lambda \in L^\infty(0, T; \mathbb{R})$ such that, for almost every $t \in [0, T]$:

- $\frac{1}{2} M_1 R_1^2 \ddot{\theta}_1(t) = \Gamma_1 + \lambda(t)$,
- $\frac{1}{2} M_2 R_2^2 \ddot{\theta}_2(t) = -\lambda(t)$,
- $\lambda(t) \leq 0$,
- $\dot{\theta}_1(t) - \dot{\theta}_2(t) \leq 0$,
- $\lambda(t) \left(\dot{\theta}_1(t) - \dot{\theta}_2(t) \right) = 0$,
- + initial conditions.

By corollaries 27 and 29, we know that this evolution problem admits a unique solution whatever is $T > 0$. Next, choose:

$$\dot{\theta}_1(0) = -1 \text{ and } \dot{\theta}_2(0) = 0,$$

as initial conditions. Suppose, in addition, that $\Gamma_1 > 0$. Then, it is readily seen that the solution of the dynamics is given by:

$$\dot{\theta}_1(t) = -1 + \frac{2\Gamma_1}{M_1 R_1^2} t, \quad \dot{\theta}_2(t) = 0, \qquad \text{if } 0 \leq t \leq \frac{M_1 R_1^2}{2\Gamma_1},$$

$$\dot{\theta}_1(t) = \dot{\theta}_2(t) = \frac{2\Gamma_1}{M_1 R_1^2 + M_2 R_2^2} \left(t - \frac{M_1 R_1^2}{2\Gamma_1} \right), \qquad \text{if } t \geq \frac{M_1 R_1^2}{2\Gamma_1}.$$

The acceleration does not depend continuously on time. This example illustrates the fact that we can not require, in general, that the solution of problem V belongs to $C^2([0,T];Q)$.

Also, this example explains the reason why we have allowed one of the C_i in the general theory, to be unbounded.

More generally, it is seen that the formalism of non-firm constraints can handle those cases where the constraints appears as inequalities applying on the velocity.

6. Bilateral constraints with dry friction

Usually, the dynamics of rigid bodies systems involving dry friction is formulated in terms of the *real world* reactions. However, this standard approach leads to two major difficulties.

- In case where the contact between two solids occur at more than two points, the real world reactions are generally not defined. The only reaction force which makes sense is the generalized reaction. With respect to this, the reader is referred to example 12.
- Such a formulation leads to situations where the dynamics is ill-posed. There may happen non-uniqueness of solutions and even non-existence (see LÖTSTEDT (1981)). As stated in the introduction, my opinion is that well-posedness should be a requirement for any theory in classical dynamics.

In this section, the formalism of non-firm constraints is applied to derive a general formulation of the dynamics of rigid bodies systems involving dry friction associated with a bilateral constraint. This formulation relies on the following principles. First, we consider a holonomic bilateral constraint. The associated reaction forces define the so-called 'normal reaction'. Next, we are given in addition a general non-holonomic bilateral constraint. The associated reaction forces define the 'tangential reaction'. This supplementary constraint is supposed to be non-firm and the tangential reaction is required to belong to a closed convex set depending on the magnitude of the normal reaction. Then, as for the general case of non-firm constraints, the flow rule is supposed to be governed by the Principle of Maximal Dissipation.

As illustrated by example 11, whenever we study systems of punctual particles, this new formulation encompass the usual formulation (for example, Coulomb friction law), since, in that case, the real world reactions are components of the generalized reaction.

Formulation and well-posedness of the dynamics are first derived. Next, we discuss in length, through many examples, how the theory is to be applied in practical situations.

6.1 Formulation of the dynamics

Consider a simple discrete mechanical system according to definition 7. First, we superimpose a perfect holonomic bilateral constraint described by a single function φ_1 as in Section 2. Hence, the motion is required to take place in the submanifold:

$$S = \{q \in Q \; ; \; \varphi_1(q) = 0\} \, .$$

The equation of motion was seen to be:

$$\flat \frac{\mathrm{D}_Q}{\mathrm{d}t} \dot{q}(t) = f(q(t), \dot{q}(t); t) + \lambda_1(t) \, \mathrm{d}\varphi_1(q(t)),$$

where λ_1 is *a priori* unknown, but it is completely determined once the evolution problem has been solved. Physically, it could be said that $\|\lambda(t) \, \mathrm{d}\varphi_1(q(t))\|^*_{q(t)}$ is a measure of 'how much the system is constrained' at instant t to remain in S. It was also noted in Section 2 that the equation of motion can be written as:

$$\flat \frac{\mathrm{D}_S}{\mathrm{d}t} \dot{q}(t) = \mathrm{Proj}^*_{q(t)} \left[f(q(t), \dot{q}(t); t); T^*_{q(t)} S \right] .$$

Then, $q(t)$ being the motion of the system, we have:

$$\flat \frac{\mathrm{D}_Q}{\mathrm{d}t} \dot{q}(t) = \flat \frac{\mathrm{D}_S}{\mathrm{d}t} \dot{q}(t) + \mathrm{Proj}^*_{q(t)} \left[f(q(t), \dot{q}(t); t); \mathbb{R} \, \mathrm{d}\varphi_1(q(t)) \right] + \lambda_1(t) \, \mathrm{d}\varphi_1(q(t)).$$

To describe the physical phenomenon of dry friction, we shall superimpose a non-firm, non-holonomic bilateral constraint whose threshold depends on the magnitude $\|\lambda_1(t) \, \mathrm{d}\varphi_1(q(t))\|^*_{q(t)}$ of the normal reaction. More precisely, consider a non-holonomic constraint described by n 1-forms $\alpha_{1j} \in T^*S$. This constraint will be supposed to be non-firm. According to the formalism of Section 5, we are given:

- a bounded closed convex subset C_1 of $\mathbb{R}^n$, containing the origin,
- an invertible square real matrix $M(q)$ of order n, which depends smoothly on q,
- a smooth function $\kappa_1 : TQ \times \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^+$.

Define:

$$C(q, \dot{q}; t; r) = \left\{ \sum_{i=1}^{n} \lambda_i \, \alpha_i(q) \; ; \; (\lambda_1, \lambda_2, \ldots, \lambda_n) \in M(q) \cdot [\kappa_1(q, \dot{q}; t; r) C_1] \right\}$$

Then, following the formalism of Section 5, the equation of motion of the system subjected to the frictional bilateral constraint φ_1 is written as:

- $q(t) \in S$,
- $\flat \dfrac{\mathrm{D}_S}{\mathrm{d}t}\dot{q}(t) - \mathrm{Proj}^*_{q(t)}\left[f(q(t),\dot{q}(t);t);T^*_{q(t)}S\right]$
$$\in \partial_S S_{C(q(t),\dot{q}(t);t;r(t))}(-\dot{q}(t)),$$
- $\flat \dfrac{\mathrm{D}_Q}{\mathrm{d}t}\dot{q}(t) = \flat \dfrac{\mathrm{D}_S}{\mathrm{d}t}\dot{q}(t) + \mathrm{Proj}^*_{q(t)}\left[f(q(t),\dot{q}(t);t);\mathbb{R}\,\mathrm{d}\varphi_1(q(t))\right]$
$$+ \lambda_1(t)\,\mathrm{d}\varphi_1(q(t)),$$
- $r(t) = \|\lambda_1(t)\,\mathrm{d}\varphi_1(q(t))\|^*_{q(t)}$,

where the S in ∂_S recalls that the subdifferential is to be understood in the sense of the duality (T_qS, T^*_qS).

Now, we are going to obtain a generalization to the case of a frictional bilateral constraint described by l smooth and functionally independent functions φ_i. The submanifold S containing the constrained motions is now defined by:

$$S = \{q \in Q\ ;\ \forall\, i = 1,2,\ldots,l, \quad \varphi_i(q) = 0\}\,.$$

The other data are as follows.

- α_i are n linearly independent 1-forms in T^*S,
- C_0 is a given closed convex subset of $\mathbb{R}^n$, possibly unbounded and containing the origin,
- the C_i $(i = 1,2\ldots,m)$ are given *bounded* closed convex subsets of $\mathbb{R}^n$, containing the origin,
- $M(q)$ is a given invertible square real matrix of order n, which depends smoothly on $q \in S$,
- the $\kappa_i : TQ \times \mathbb{R} \times (\mathbb{R})^l \to \mathbb{R}^+$ are given functions whose regularity will be stated later on.

How these data are to be constructed in practical situations will be seen through the examples of Section 6.3. Next, we define:

$$C(q,\dot{q};t;r) = \left\{ \sum_{i=1}^{n} \lambda_i\,\alpha_i(q)\ ; \right.$$
$$\left. (\lambda_1,\lambda_2,\ldots,\lambda_n) \in M(q)\cdot\left[C_0 + \sum_{i=1}^{m} \kappa_i(q,\dot{q};t;r)C_i\right]\right\}.$$

The closed convex subset $C(q, \dot{q}; t; r)$ of $T_q^* S$ could be called 'the set of all admissible tangential reactions'.

Now, given any initial condition $(q_0, v_0) \in TQ$ compatible with the frictional constraint:

$$-v_0 \in \mathrm{Dom}\, S_{C(q_0, v_0; t_0)} \subset T_{q_0} S,$$

the evolution problem associated with the dynamics of simple discrete mechanical systems subjected to frictional bilateral constraints is formulated as follows.

Problem VI. Find $T > t_0$, $q \in W^{2,\infty}([t_0, T[; Q)$ and $\lambda_i \in C^0([t_0, T[; \mathbb{R})$ $(i = 1, 2, \ldots, l)$ such that:

- $(q(t_0), \dot{q}(t_0)) = (q_0, v_0)$,
- $q(t) \in S, \quad \forall\, t \in [t_0, T[$,
- $\flat \dfrac{\mathrm{D}_S}{\mathrm{d}t} \dot{q}(t) - \mathrm{Proj}^*_{q(t)} \left[f(q(t), \dot{q}(t); t); T^*_{q(t)} S \right]$
$$\in \partial_S S_{C(q(t), \dot{q}(t); t; r_n(t))}(-\dot{q}(t)),$$
- $\flat \dfrac{\mathrm{D}_Q}{\mathrm{d}t} \dot{q}(t) = \flat \dfrac{\mathrm{D}_S}{\mathrm{d}t} \dot{q}(t) + \mathrm{Proj}^*_{q(t)} \left[f(q(t), \dot{q}(t); t); \bigoplus_{i=1}^{l} \mathbb{R}\, \mathrm{d}\varphi_i(q(t)) \right]$
$$+ \sum_{i=1}^{l} \lambda_i(t)\, \mathrm{d}\varphi_i(q(t)), \quad \text{for a.e. } t \in [t_0, T[,$$
- $r_n(t) = (\lambda_1(t), \lambda_2(t), \ldots, \lambda_l(t))$.

6.2 Well-posedness of the dynamics

Regularity hypothesis I. The configuration manifold Q is of class C^2, the mapping $f : TQ \times \mathbb{R} \to T^*Q$, the 1-forms α_i and the mapping $M : q \mapsto M(q)$ are of class C^1. Also, the functions $\kappa_i : TQ \times \mathbb{R} \times (\mathbb{R}^+)^l \to \mathbb{R}^+$ are locally lipschitzian.

Theorem 30 *There exists a unique maximal solution for problem VI.*

Proof. We are going to prove that problem VI reduces to a problem V on S. Then, corollary 27 will yield the claim.

First, consider an arbitrary chart (U_S, ψ_S) at q_0 on S. We shall denote $\psi_S(q) \in \mathbb{R}^{d-l}$ by $(q^{l+1}, q^{l+2}, \ldots q^d)$. Next, we construct a local chart at q_0 on:

$$S_{l-1} = \{q \in Q\ ;\ \forall\, i = 1, 2, \ldots, l-1, \quad \varphi_i(q) = 0\},$$

by use of the flow of the vector field

$$\frac{\nabla\varphi_l(q)}{\|\nabla\varphi_l(q)\|_q^2}$$

on S_{l-1}. The supplementary coordinate is denoted by q^l and it is readily checked that:

$$q^l = \varphi_l(q).$$

Repeating that construction successively on $S_{l-2}, \ldots, S_1, S_0 = Q$, we obtain a chart (U, ψ) at q_0 on Q such that:

- the l first coordinates of $\psi(q)$ are the $\varphi_i(q)$ $(i = 1, 2, \ldots, l)$.
- for all $q \in S$, the $\partial/\partial q^i$ $(i = 1, 2, \ldots, l)$ are orthogonal to the $\partial/\partial q^j$ $(j = l+1, l+2, \ldots, d)$.

As a consequence, the representative of the metric tensor in that chart satisfies:

$$\forall q \in S, \quad \forall i \in \{1, 2, \ldots, l\}, \quad \forall j \in \{l+1, l+2, \ldots, d\}, \qquad g_{ij}(q) = g^{ij}(q) = 0.$$

Writing the evolution problem in the chart under consideration gives:

$$\lambda_i = g_{ij}(q)\Gamma^j_{kl}(q)\dot{q}^k\dot{q}^l - f_i(q, \dot{q}; t),$$

for $i = 1, 2, \ldots, l$. Therefore, $r_n(t)$ is determined by $(q, \dot{q}; t) \in TS \times \mathbb{R}$. Moreover, the induced mapping $r_n : TS \times \mathbb{R} \to (\mathbb{R}^+)^l$ is clearly locally Lipschitzian. Defining:

$$\mu_i(q, \dot{q}; t) = \kappa_i\left(q, \dot{q}; t; r_n(q, \dot{q}; t)\right),$$

we see that the evolution problem reduces to a problem V on S. Thus, the existence and uniqueness of a maximal solution for problem VI is provided by corollary 27. □

Proposition 31 *The configuration manifold Q is assumed to be a complete Riemannian manifold and the mapping f is supposed to admit the following estimate:*

$$\forall (q, v) \in TQ, \quad \textit{for almost all } t \in [t_0, +\infty[, \qquad \|f(q, v; t)\|_q^* \leq l(t)\left(1 + d(q, q_0) + \|v\|_q\right),$$

where $d(\cdot, \cdot)$ is the Riemannian distance and $l(\cdot)$ a nonnegative function in $L^1_{loc}(\mathbb{R}; \mathbb{R})$.

Then, the dynamics is eternal, that is, the maximal solution for problem VI is defined on $[t_0, +\infty[$.

Proof. If Q is complete, then so is S. Moreover, we have:

$$\forall\,(q,v;t) \in TS \times \mathbb{R}, \quad \left\|\mathrm{Proj}_q^*\left[f(q,v;t);T_q^*S\right]\right\|_q^* \leq \|f(q,v;t)\|_q^*\,.$$

Therefore, use of corollary 29 in the proof of theorem 30 yields the claim. □

6.3 Illustrative examples and comments

Our general formulation of dry friction relies on the Principle of Maximal Dissipation through the formalism of non-firm constraints. In some cases, it is the same that the usual formulation of Coulomb friction, as seen on next example.

Example 11. Consider a punctual particle of mass 1 moving in the usual Euclidean $\mathbb{R}^3$. This particle is free of external forces but is constrained to move in a two-dimensional submanifold of $\mathbb{R}^3$. In order to simplify the equations, we shall assume that this submanifold can be represented by the Cartesian equation:

$$z = s(x,y) \quad \text{(that is, } \varphi_1(x,y,z) = z - s(x,y)\text{)}.$$

The associated 'normal reaction' has general expression:

$$R_N = \frac{\lambda}{\sqrt{1+(\partial s/\partial x)^2+(\partial s/\partial y)^2}}\left(-\frac{\partial s}{\partial x}\,\mathrm{d}x - \frac{\partial s}{\partial y}\,\mathrm{d}y + \mathrm{d}z\right),$$

In order to express that this bilateral constraint is frictional, we shall superimpose a non-firm constraint of immobility on the constrained submanifold. So, we have to introduce two 1-form fields to express that the tangential velocity vanishes. It could be natural to use $\mathrm{d}x$ and $\mathrm{d}y$, but to stick to our formalism of non-firm constraint, we project them on the orthogonal complement of

$$\mathrm{d}\varphi_1 = -\frac{\partial s}{\partial x}\,\mathrm{d}x - \frac{\partial s}{\partial y}\,\mathrm{d}y + \mathrm{d}z$$

to get:

$$\begin{aligned}
\alpha_1 &= \frac{1}{\sqrt{1+(\partial s/\partial x)^2}}\left(\mathrm{d}x + \frac{\partial s}{\partial x}\,\mathrm{d}z\right),\\
\alpha_2 &= \frac{1}{\sqrt{1+(\partial s/\partial x)^2}}\left(\mathrm{d}y + \frac{\partial s}{\partial y}\,\mathrm{d}z\right).
\end{aligned}$$

These 1-form fields have been normalized for sake of simplicity (this normalization could also have been done by introducing appropriate matrix $M(q)$

in the definition of $C(q, R_N)$ below). Postulating the following convex set of admissible 'tangential reactions' associated with the non-firm constraint of immobility:

$$C(x, y, z, \lambda) = \left\{ \lambda_1 \alpha_1 + \lambda_2 \alpha_2 \; ; \; \sqrt{\lambda_1^2 + \lambda_2^2} \leq \mu |\lambda| \right\},$$

where μ is a positive real constant (the so-called Coulomb friction coefficient), it is now an easy matter to write the equation of motion in the parametrization (x, y, z). Writing the corresponding evolution problem VI, the reader will check that one recovers the usual formulation of Coulomb friction. It is of interest to notice, that whenever the function s is not linear, some frictional dissipation can be activated during the motion even if the particle is free of external forces.

Many authors prefer to write directly the usual Coulomb friction law in any case rather than coming back to the Principle of Maximal Dissipation as we did. This method necessarily requires that the evolution problem should be written in terms of the 'real world reactions' instead of the 'generalized reaction' as in our formulation. At first glance, this makes it easier to identify the constitutive equations in practical situations. But, it should be stressed that the concept of 'real world reaction' is in general meaningless in the framework of rigid bodies system. Indeed, the whole theory relies on the rigid geometric description which determines the structure of the space of all virtual velocities. By duality, through the Virtual Power Principle, we obtain the representation of forces as linear forms on the space of virtual velocities. This is the most general representation of forces which is consistent with the geometric description of the system.

We are going to try to illustrate these general considerations by examining some more complicated examples.

Example 12. Consider a rigid four-feet table lying upon a plane floor. The extremity of each of the feet is supposed to be constrained to remain on the floor. This is a holonomic constraint which is described by *three* independent smooth functions. Some external forces are applied on the table. We aim at writing the evolution problem associated with the dynamics, with some dry friction between the table and the floor taken into account. In this example, the 'real world' reactions in each of the feet of the table are *undetermined* and we have actually no other choice than using the generalized reactions to express the dry friction.

We use the coordinates (x, y, z) of the 'center' of the table and Euler angles (ψ, θ, ϕ) to parametrize the system in such a way that the initial configuration is given by:

$$x_0 = y_0 = z_0 = 0, \quad \psi_0 = \phi_0 = 0, \quad \theta_0 = \frac{\pi}{2},$$

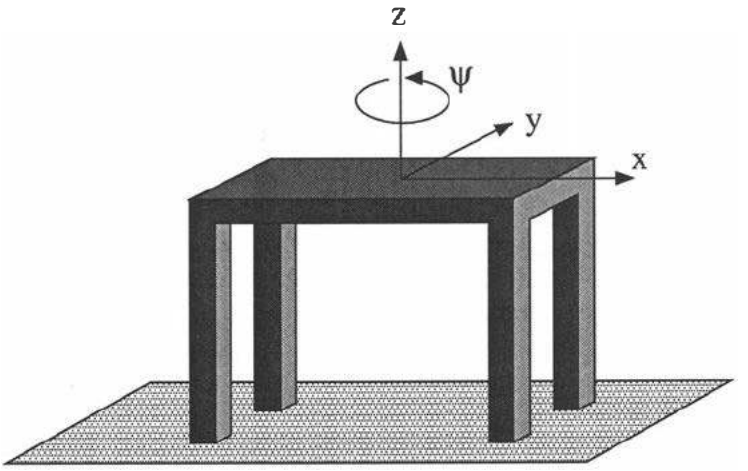

Figure 1.6. Four-feet table on a frictional floor.

and the holonomic constraint is described by:

$$z = 0, \quad \theta = \frac{\pi}{2}, \quad \phi = 0,$$

(see Figure 1.6). The associated reaction force (normal reaction) admits the general form:

$$R_N = R_z \, \mathrm{d}z + R_\theta \, \mathrm{d}\theta + R_\phi \mathrm{d}\phi.$$

To describe the dry friction with the floor, we superimpose the non-holonomic constraint defined by the three 1-forms $\mathrm{d}x$, $\mathrm{d}y$ and $\mathrm{d}\psi$. The associated reaction force (tangential reaction) is written as:

$$R_T = R_x \, \mathrm{d}x + R_y \, \mathrm{d}y + R_\psi \, \mathrm{d}\psi.$$

This non-holonomic constraint will be assumed to be non-firm with convex of admissible (tangential) reaction defined by:

$$C(R_N) = \Big\{ R_x \, \mathrm{d}x + R_y \, \mathrm{d}y + R_\psi \, \mathrm{d}\psi \; ; \\ (R_x, R_Y, R_\psi) \in \kappa_1(|R_z|, |R_\theta|, |R_\phi|) C_1 \Big\},$$

where C_1 is a given bounded closed convex subset of $\mathbb{R}^3$, containing the origin and κ_1 an arbitrary smooth function taking positive values. It is then an easy matter to write the equation of motion of the system in the parametrization under consideration.

What we want to stress is that this is the most general description of the physical phenomenon of dry friction which is consistent with our choice for the geometric description of the table.

A general feature of this formulation is that, for each configuration q_0 and each given constant 'generalized' external forces f_0, immobility will be the further motion of the system if and only if f_0 belongs to a given convex set. But, we have seen, in Section 1, that the 'generalized' forces f_0 are given, in general, in terms of some 'real world' forces distribution ϕ (notations of Section 1). It is easy to design experiments (for example, on the system of the above example), in which it could be observed that there exist two 'real world' forces distribution ϕ_1 and ϕ_2 consistent with the same generalized forces f_0 and such that ϕ_1 induces immobility of the system whereas ϕ_2 does not. In such a case, what is questionable is not our general formulation of dry friction, it is the geometric assumption of rigidity which is too rough to describe the physical phenomenon (dry friction) under consideration. In such a case, the only way to obtain a more realistic model is to refine the geometric description by adding some degrees-of-freedom. In the above example, this could be done by allowing that each foot of the table is connected to the table through a joint equipped with springs in such a way that some components of the generalized reaction can be interpreted in terms of real world reaction components.

Of course, in any case, there remains to identify the convex set C_1 and the function κ_1 which may turn out to be not so easy. But, once more this is the price we have to pay for the simplicity of the geometric description that we have adopted. Making that choice of simplicity requires to inject a lot of information which is not necessarily at hand in the constitutive equation. We could say that the geometric assumption that has been made is not in accordance with the physical phenomena we wish to describe.

In our formulation, the constitutive equation is completely determined by the data of the convex of admissible 'tangential generalized reaction' through the Principle of Maximal Dissipation. Of course, nothing forbids to use considerations based on 'real world reactions' to derive the convex set of admissible 'tangential generalized reaction' which should be postulated in one practical situation or the other. However, this method is far from working all the time as seen in next example.

Example 13. Consider a rigid homogeneous bar with length L and mass M which is constrained to move in a fixed plane. One of the extremities of the bar is constrained to remain on a fixed bar. We suppose that dry friction is associated to that bilateral constraint. We shall use the parametrization $q = (x, y, \theta)$ as represented on Figure 1.7. The kinetic energy in this parametrization is given

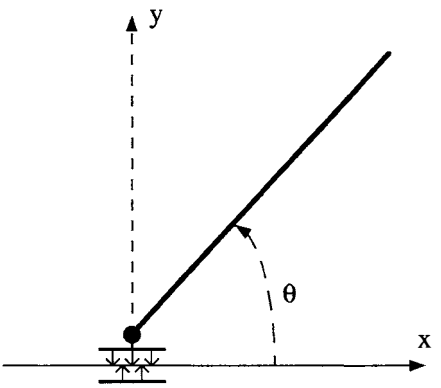

Figure 1.7. Rigid bar with frictional constraint at one extremity.

by:

$$K(q,\dot{q}) = \frac{M}{2}\left(\dot{x}^2 + \dot{y}^2 + \frac{L^2}{3}\dot{\theta}^2 - L\sin\theta\,\dot{x}\dot{\theta} + L\cos\theta\,\dot{y}\dot{\theta}\right),$$

and the external forces have general expression:

$$f(q,\dot{q};t) = f_x(q,\dot{q};t)\,\mathrm{d}x + f_y(q,\dot{q};t)\,\mathrm{d}y + f_\theta(q,\dot{q};t)\,\mathrm{d}\theta.$$

The holonomic bilateral constraint is represented by the single function:

$$\varphi_1(x,y,\theta) = y,$$

which defines the 'normal' reaction as:

$$R_N\,\mathrm{d}y$$

To describe dry friction, we superimpose a non-firm non-holonomic constraint which requires that $\dot{x}$ should vanish. In the formalism of non-firm constraint, it involves the 1-form α_1 obtained by projection of $\mathrm{d}x$ on the orthogonal complement of $\mathrm{d}y$:

$$\alpha_1 = \mathrm{d}x + \frac{3\cos\theta\sin\theta}{4 - 3\sin^2\theta}\,\mathrm{d}y.$$

The convex of admissible tangential reaction has general form:

$$C(q,\dot{q};t;R_N) = \{R_T\alpha_1\ ;\ R_T \in \kappa_1(q,\dot{q};t;|R_N|)C_1\},$$

where C_1 is a bounded closed convex subset of $\mathbb{R}$, containing the origin and κ_1 a function. To identify the constitutive data C_1 and κ_1, a natural démarche

is to specify that the 'real world' reaction $R_x\,\mathrm{d}x + R_y\,\mathrm{d}y$ should belong to the Coulomb cone:

$$C = \{R_x\,\mathrm{d}x + R_y\,\mathrm{d}y\ ;\ |R_x| \leq \mu|R_y|\} \tag{1.36}$$

(where μ denotes the usual Coulomb friction coefficient) and to translate it in terms of the 'normal and tangential generalized reaction'. We obtain:

$$C(q,\dot{q};t;R_N) = \left\{R_T\alpha_1\ ;\ |R_T| \leq \left|\frac{3\cos\theta\sin\theta}{4-3\sin^2\theta}R_T + R_N\right|\right\}$$

which *is not convex in general* and prevents from applying the Principle of Maximal Dissipation. If we postulate Coulomb flow rule in such a case instead of the Principle of Maximal Dissipation, then we obtain an ill-posed evolution problem with possible multiple solutions or also no solutions at all, as it is well-known (see LÖTSTEDT (1981) and the references of that paper). In such a case, the only way to write the equation of motion which remains consistent with the initial geometric description is to stick to the above formalism of non-firm constraints. Of course, some structural effects are incorporated in the definition of the convex of admissible 'tangential' reaction and it is hard to see which convex set should be postulated in that situation. Moreover, it is also possible that we obtain unrealistic predictions, in which case the geometric description should be refined.

Now, we are going to discuss a last example which illustrates the interest of allowing the possibly unbounded convex set C_0 in the formalism of this section. It also demonstrates that some structural effects can play a role in the definition of the convex of admissible 'tangential' reactions.

Example 14. Consider the same system as in example 13, but suppose that, in addition, the free extremity of the bar is ideally constrained to remain on a fixed bar as on Figure 1.8. We keep the primitive parametrization $q = (x, y, \theta)$ as defined in example 13. The external forces have general form:

$$f(q,\dot{q};t) = f_x(q,\dot{q};t)\,\mathrm{d}x + f_y(q,\dot{q};t)\,\mathrm{d}y + f_\theta(q,\dot{q};t)\,\mathrm{d}\theta.$$

The bilateral constraint associated with the 'bottom' fixed bar is still represented by:

$$\varphi_1(q) = y,$$

whereas the bilateral constraint associated with the 'top' bar is represented by:

$$\varphi_2(q) = y + L\sin\theta - d,$$

where d denotes the distance between the two fixed bars. The (normal) generalized reaction associated with that constraint has general expression:

$$\lambda_1\,\mathrm{d}\varphi_1(q) + \lambda_2\,\mathrm{d}\varphi_2(q).$$

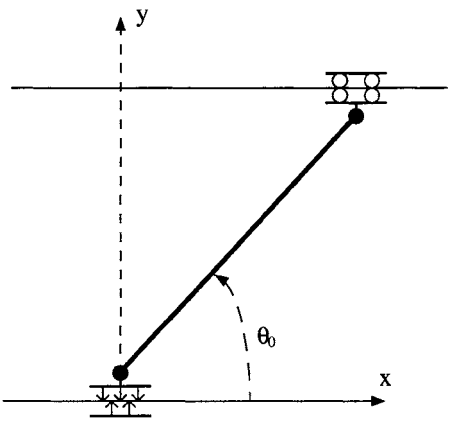

Figure 1.8. Rigid bar with frictional constraint at one extremity and perfect constraint at the other.

Let us note that we could equally have represented the bilateral constraint associated with the 'top' bar by the function:

$$\varphi_2'(q) = \theta - \theta_0,$$

in which case the normal generalized reaction would be expressed as:

$$\mu_1 \, \mathrm{d}\varphi_1(q) + \mu_2 \, \mathrm{d}\varphi_2'(q).$$

Since neither $(\mathrm{d}\varphi_1(q), \mathrm{d}\varphi_2(q))_q^* = 0$ nor $(\mathrm{d}\varphi_1(q), \mathrm{d}\varphi_2'(q))_q^* = 0$, we don't have $\lambda_1 = \mu_1$. In other words, we cannot *intrinsically define the normal reaction associated with one fixed bar or the other.* Actually, this is meaningless to say that one of the constraint is ideal and the other is frictional. The reason is that they are coupled by the kinetic metric. In the framework of rigid bodies dynamics, the only thing which can be expressed is that there is a dissipation mechanism associated with the bilateral constraint associated with the two functions φ_1 and φ_2. We shall see that this does not contradict our general formulation through the non-firm constraint formalism. Actually, this is an indication that some structural effects are to be incorporated in the expression of the constitutive law associated with the dissipation mechanism. Therefore, the Coulomb friction law, which is local by nature, cannot be enough to build the constitutive law associated with such a dissipation mechanism.

We are going to write the equation of motion by applying the formalism of frictional bilateral constraint to the bilateral constraint defined by the two functions φ_1 and φ_2. To express the non-firm non-holonomic constraint associated with the dissipation mechanism, we first project the 1-form $\mathrm{d}x$ onto the

orthogonal complement of the subspace containing the 'normal generalized' reaction:

$$\alpha_1 = \mathrm{d}x - \frac{1}{2} L \sin\theta_0 \, \mathrm{d}\theta,$$

We have now to postulate the constitutive law associated with the dissipation mechanism by defining a convex of admissible 'tangential generalized' reaction. One way to proceed is to make experiments to identify the set of values of (f_x, f_y, f_θ) are compatible with equilibrium. Once more, there may happen that different experimental ways of loading the system, which correspond to the same value of (f_x, f_y, f_θ), give different outcomes of the experiment (equilibrium or not). In such a case, it is the simplicity of the geometric description which sould be questioned and refined. Another way to identify the convex of admissible 'tangential generalized' reaction is to express that the 'real world' reaction at the bottom extremity of the bar belong to some Coulomb cone. This method should never be considered as systematic, since, fundamentally, it mixes two different geometric descriptions of the system (rigid and deformable). However, in some particular cases, this method can be a good guideline to identify the constitutive law corresponding to the dissipation mechanism. In the example under consideration, we denote by:

$$R_T \, \alpha_1 + R_{Ny} \, \mathrm{d}y + R_{N\theta} \, \mathrm{d}\theta$$

the'generalized' reaction, by:

$$R_{x1} \, \mathrm{d}x + R_{y1} \, \mathrm{d}y$$

the 'real world' reaction associated with the top bar and by:

$$R_{y2} \, \mathrm{d}y$$

the 'real world' reaction associated with the bottom bar. Assuming $\theta_0 \neq \pi/2$, we obtain easily:

$$\begin{aligned} R_{x1} &= R_T, \\ R_{y1} &= \frac{1}{2} \tan\theta_0 \, R_T + R_{Ny} - \frac{1}{L \cos\theta_0} R_{N\theta}, \\ R_{y2} &= -\frac{1}{2} \tan\theta_0 \, R_T + \frac{1}{L \cos\theta_0} R_{N\theta}. \end{aligned}$$

Expressing that the 'real world' reaction $R_{x1} \, \mathrm{d}x + R_{y1} \, \mathrm{d}y$ belongs to the Coulomb cone (1.36), we have:

$$|R_T| \leq \mu \left| \frac{1}{2} \tan\theta_0 \, R_T + R_{Ny} - \frac{1}{L \cos\theta_0} R_{N\theta} \right|. \tag{1.37}$$

In the case $\mu \tan_{00} < 2$, we easily identify:

$$C(R_{Ny}, R_{N\theta}) = \Big\{ R_T \alpha_1 \ ; \ R_T, R_{Ny}, R_{N\theta} \text{ satisties inequality (1.37)} \Big\},$$

which is easily put under the form:

$$C(R_{Ny}, R_{N\theta}) = \Big\{ R_T \alpha_1 \ ; \\ R_T \in \kappa^-(R_{Ny}, R_{N\theta})[-1, 0] + \kappa^+(R_{Ny}, R_{N\theta})[0, 1] \Big\},$$

where κ^- and κ^+ are positive Lipschitzian function. The equation of motion in the parametrization under consideration is:

$$\begin{aligned} M\ddot{x} &= f_x(x, \dot{x}; t) + R_T, \\ 0 &= f_y(x, \dot{x}; t) + R_{Ny}, \\ -\frac{ML}{2} \sin\theta_0 \ddot{x} &= f_\theta(x, \dot{x}; t) + R_{N\theta} - \frac{ML}{2} \sin\theta_0 R_T, \\ -\dot{x} &\in \partial I_{\kappa^-(R_{Ny}, R_{N\theta})[-1,0] + \kappa^+(R_{Ny}, R_{N\theta})[0,1]}(R_T). \end{aligned}$$

It is well-posed by virtue of theorem 30.

In the case $\mu \tan\theta_0 \geq 2$, inequality (1.37) does not allow any more to identify the convex of admissible tangential reaction. Some structural effects are to be incorporated in the definition of $C(q, \dot{q}; t)$. In this situation, there may happen what is often called 'dynamical locking': some arbitrary large values of the tangential reaction can be compatible with equilibrium. To model such a situation, it may turn out convenient to use some unbounded convex subset C_0 of $\mathbb{R}$ in the definition of $C(q, \dot{q}; t)$. Let us underline that this situation of possible dynamical locking has to be postulated in the constitutive law. It can not be theoretically investigated in the framework of the simple geometric description of the system that has been adopted. The only way to lead this investigation would be to refine the geometric description.

One word to conclude and summarize this section. The point of view on dry friction that we have developed is the following: there is a dissipation mechanism associated with a bilateral constraint which depends on the reaction force associated with that constraint. The flow rule associated with this mechanism obeys to the Principle of Maximal Dissipation. This point of view allows a systematic and intrinsic formulation of the dynamics which is proved to be well-posed. In case where the system contains only punctual particles (or is a deformable body), we recover the usual local Coulomb friction law. In the other cases, we obtain the most general formulation of the dynamics which is consistent with the geometric description of the system. Trying to use the local Coulomb friction law in every case is not consistent with the geometric description of the system and produces numerous paradoxes, as is well-known.

7. On frictional unilateral constraints and related open problems

In this paper, we have extended the classical theory of the dynamics of simple discrete mechanical systems in two directions.

- In Section 3, we have discussed formulation and well-posedness of the dynamics of simple discrete mechanical systems submitted, in addition, to perfect unilateral constraints.

- In Section 4, we have discussed the same issues for simple discrete mechanical systems undergoing, in addition, non-holonomic bilateral constraints. Since a non-holonomic bilateral constraint can be viewed as a frictional bilateral constraint with infinitely large friction, the idea of non-holonomic constraint has been generalized to non-firm and frictional bilateral constraint. General and systematic formulation of the dynamics of such systems has been derived and well-posedness has been established.

Naturally, having in mind a general theory of the evolution of complex mechanisms, the question arises to take into account both unilateral constraints and frictional constraints. That is, we would like to be able to mix the two above theories. Since frictional bilateral constraint appears to be a generalization of non-holonomic bilateral constraint, we are going to handle the problem of associating unilateral constraints with non-holonomic constraints for sake of simplicity.

There are essentially two ways of associating non-holonomic bilateral constraints with unilateral constraints:

- the unconditional association means that the non-holonomic bilateral constraint is always active, no matter whether the unilateral constraint is active or not,

- the conditional association means that the non-holonomic constraint is active only when the unilateral constraint is active.

A typical occurrence of unconditional association is the rolling without slipping of a billiard ball on a billiard table with possible collisions with the edges of the table. An example of conditional association is the rolling without sliping of a ball on a wavy profile with possible takeoff.

The general theory of unconditional association can easily be derived by combining the contents of Sections 3 and 4. Systematic formulation and well-posedness would be obtained as well as the general conditions that should be satisfied by the impact constitutive equation. Even, it is possible to extend to the theory on unconditional association of unilateral constraints with frictional

bilateral constraints. The only difficulty, which is easily overcome, is to derive sufficient regularity assumptions on the data of problem VI to ensure that its solution is analytic on a right neighbourhood of every instant. An example of application of such a theory could be the dynamics of the double-pendulum with obstacle as in example 8 where we could take into account some friction in each of the ball-and-socket joints in addition.

The general theory of conditional association turns out to be more complicated. Some substantial adaptation of the proof of theorem 14 seems to be necessary.

As a matter of conclusion, let us underline the following remark which comes back in the paper as a *leitmotiv*. A complete theory of the dynamics of rigid bodies systems taking into account complicated phenomena such as impacts or friction is highly desirable in view of a lot of applications (granular dynamics, virtual reality, etc...). However, it turns out that it is the fact that bodies are actually deformable and not rigid that governs those physical phenomena. In principle, it does not prevent to derive a complete theory of the dynamics of rigid bodies including these phenomena, but, we have to keep in mind that the structural effects (those which physically rely on the fact that the bodies are deformable) are incorporated in the constitutive equations. As a result, there will be probably many situations where the theory will be of no use because no realistic constitutive equation will be at hand.

Appendix: The class of motion $MMA(I, Q)$

In this section, I is any interval of the real line, $(E, \|\cdot\|)$ a finite-dimensional normed vector space and Q a d-dimensional Hausdorff manifold. We do not aim at being systematic nor general, but only at stating definitions and elementary results needed in this paper. Since they are very easy, proofs are generally omitted.

Bounded variation of E-valued functions

In this section, we briefly recall standard results whose proofs may be found, for example, in RUDIN (1966) and MOREAU (1988b).

A function $f : I \to E$ is said to have *bounded variation* if

$$\mathrm{Var}(f, I) \stackrel{\text{def}}{=} \sup \sum_{i=1}^{n} \|f(t_i) - f(t_{i-1})\| < \infty,$$

where the supremum is taken over all strictly increasing finite sequences $t_0 < t_1 < \cdots < t_n$ in I. The property of having bounded variation does not depend on the particular choice of a norm on E whereas the real number $\mathrm{Var}(f, I)$ depends on that choice. A function $f : I \to E$ is said to have locally bounded variation if it has bounded variation on any compact subinterval of I. If f has locally bounded variation, then it admits left and right limits at every $t \in \overset{\circ}{I}$ (notation $f^-(t)$, $f^+(t)$). The function f is continuous at every $t \in \overset{\circ}{I}$ (that is, $f^-(t) = f^+(t)$) except, maybe, for some t belonging to a (at most) countable subset of I. The function f is differentiable in the classical sense at every $t \in \overset{\circ}{I}$ except, maybe, for some t belonging to a Lebesgue-negligible subset of I. If $\psi : E \to E$ has class C^1, then $\psi \circ f$ has locally bounded variation.

A E-valued measure on I is any real-valued linear functional μ on $C_c^0(I; E^*)$ satisfying the continuity property:

$$\forall a < b \in I, \quad \exists M_{a,b} \geq 0, \quad \forall \varphi \text{ with Supp}\,\varphi \subset [a,b], \quad |\mu(\varphi)| \leq M_{a,b} \max_{t \in I} \|\varphi(t)\|^* .$$

($(E^*, \|\cdot\|^*)$ denotes the dual space of E). The real number $\mu(\varphi)$ will also be denoted by:

$$\int_I \langle \varphi, \mu \rangle .$$

The support Supp μ of the measure μ is the complement of the union of all open subsets O of I such that:

$$\forall \varphi \in C_c^0(O; E^*), \quad \mu(\varphi) = 0.$$

For μ a E-valued measure on I, and φ a nonnegative function of $C_c^0(I, \mathbb{R})$, define:

$$|\mu|(\varphi) = \sup_{\substack{f \in C_c^0(I, E^*), \\ \|f(t)\| \leq \varphi(t)}} \mu(\varphi), \tag{1.A.1}$$

where the supremum is finite, thanks to the continuity properties included in the definition of measures. For arbitrary φ in $C_c^0(I, \mathbb{R})$, define $|\mu|(\varphi)$ by:

$$|\mu|(\varphi) = |\mu|\left(\langle\varphi\rangle^+\right) - |\mu|\left(\langle\varphi\rangle^-\right),$$

where $\langle x \rangle^{\pm} = \max\{\pm x, 0\}$ are the classical positive and negative parts. Then, $|\mu|$ is a real valued measure called the *modulus* measure of μ. We have Supp $|\mu|$ = Supp μ. Let χ_I the characteristic function of I. Define $\mu(\chi_I)$ by means of formula (1.A.1). If $\mu(\chi_I)$ is finite, then the measure μ is said *bounded*.

Let $f : I \to E$ be a function with locally bounded variation. Then, there exists a unique E-valued measure $\mathrm{d}f$ on I such that:

$$\forall \varphi \in C_c^1(I, E^*), \quad \int_I \langle \varphi, \mathrm{d}f \rangle = -\int_I \left\langle \frac{\mathrm{d}}{\mathrm{d}t}\varphi, f \right\rangle .$$

The measure $\mathrm{d}f$ is called the Stieltjes measure of f. Hence, the distributional derivative of a function with locally bounded variation is a measure. Reciprocally, if μ is a E-valued measure on I, then the E-valued function defined on I by:

$$t \mapsto \int_{[t_0, t]} \mu,$$

($t_0 \in I$) has locally bounded variation. A function with locally bounded variation has bounded variation if and only if its Stieltjes measure is bounded. If f has locally bounded variation, then, we have:

$$\begin{aligned}
\int_{]a,b[} \mathrm{d}f &= f^-(b) - f^+(a) \\
\int_{]a,b]} \mathrm{d}f &= f^+(b) - f^+(a) \\
\int_{[a,b[} \mathrm{d}f &= f^-(b) - f^-(a) \\
\int_{[a,b]} \mathrm{d}f &= f^+(b) - f^-(a)
\end{aligned}$$

If there exists $h \in L^1_{\text{loc}}(I, E)$ such that:

$$\mathrm{d}f = h\,\mathrm{d}t,$$

with $\mathrm{d}t$ denoting the Lebesgue measure on I, then the function f with locally bounded variation is said to be *locally absolutely continuous.* A locally absolutely continuous function is continuous. If $h \in L^1(I, E)$, then f is said absolutely continuous. If $\psi : E \to E$ has class C^1 and f is locally absolutely continuous, then $\psi \circ f$ is locally absolutely continuous.

>From now on, E is assumed to be a Euclidean vector space. If $f, g : I \to E$ have locally bounded variation, then $(f, g) : t \mapsto (f(t), g(t))$ has also locally bounded variation and we have:

$$\mathrm{d}(f, g) = \left(\frac{f^+ + f^-}{2}, \mathrm{d}g\right) + \left(\mathrm{d}f, \frac{g^+ + g^-}{2}\right).$$

continuous curve on Q **Bounded variation of vector fields over a locally absolutely continuous curve on Q**

Definition 32 *A curve $q : I \to Q$ is said locally absolutely continuous if, for all t in I, there exists a neighbourhood J of t in I and local chart (U, ψ) at $q(t)$ such that:*

- $q(J) \subset U$,
- $\psi \circ q : J \to \mathbb{R}^d$ *is locally absolutely continuous.*

In that case, for any local chart (U, ψ) and any subinterval J of I such that $q(J) \subset U$, the mapping $\psi \circ q : J \to \mathbb{R}^d$ is locally absolutely continuous.

Definition 33 *A vector field v over a curve $q : I \to Q$ is a mapping $v : I \to TQ$ satisfying the condition:*

$$\forall t \in I, \quad \Pi_Q\,(v(t)) = q(t).$$

From now on, $q : I \to Q$ denotes a locally absolutely continuous curve.

Definition 34 *A vector field v over q is said to have locally bounded variation if its components in any chart are real valued functions with locally bounded variation.*

It is possible to define the concept of vector fields with bounded variation on an absolutely continuous curve by means of a locally finite covering by charts domain and partition of unity, but we shall not pursue in that direction. Indeed, though the definition of the concept of bounded variation of vector fields is possible on general manifolds, the definition of the variation itself requires a Riemannian structure on Q. So, we are going to particularize the definitions to that case and from now on, we assume that Q is equipped with a Riemannian structure. By Lebesgue's theorem, $q(t)$ admits a tangent vector $\dot{q}(t) \in T_{q(t)}Q$ for $\mathrm{d}t$-almost all t in I. The Riemannian structure on Q and Caratheodory's theorem allow us to define classically a parallel translation operator along q, $\tau_{t,s} : T_{q(s)}Q \to T_{q(t)}Q$ (see, for example, CHAVEL (1993), p. 7). The operator $\tau_{t,s}$ is defined for all $(s, t) \in I^2$.

Proposition 35 *Let t_0 be an arbitrary element of I and v a vector field over q. We denote by θ_{t_0} the mapping:*

$$\theta_{t_0} \left\{ \begin{array}{ccc} I & \to & T_{q(t_0)}Q \\ s & \mapsto & \tau_{t_0,s}(v(s)) \end{array} \right. \tag{1.A.2}$$

which takes values in the d-dimensional normed vector space $T_{q(t_0)}Q$. Then, v has locally bounded variation if and only if θ_{t_0} has locally bounded variation.

Proof. It is a consequence of the identity:

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta_{t_0}(t) = \left[\frac{\mathrm{d}}{\mathrm{d}t}v^i(t) + \Gamma^i_{jk}(q(t))\dot{q}^j(t)v^k(t)\right]\tau_{t_0,t}(e_i(q(t))),$$

which holds in the sense of distributions in any local chart. □

Corollary 36 *Let J be a bounded subinterval of I and (U,ψ) a local chart such that $q(J)\subset U$. Let v^i be the components in that chart of the vector field v over q with locally bounded variation. Then, θ_{t_0} has bounded variation over J if and only if each function v^i has bounded variation over J.*

Definition 37 *Let v be a vector field over q with locally bounded variation, J any subinterval of I and θ_{t_0} as in formula (1.A.2). The variation of v over J is, by definition:*

$$\mathrm{Var}(v(s);J) = \mathrm{Var}(\tau_{t_0,s}(v(s));J).$$

It belongs to $\mathbb{R}\cup\{+\infty\}$.

That $\mathrm{Var}(v(s);J)$ does not depend on a particular choice of t_0, relies on the identity:

$$\|\tau_{t_1,s_1}(v(s_1)) - \tau_{t_1,s_2}(v(s_2))\|_{q(t_1)} = \|\tau_{t_2,s_1}(v(s_1)) - \tau_{t_2,s_2}(v(s_2))\|_{q(t_2)},$$

that holds for all $s_1, s_2, t_1, t_2 \in I$.

Proposition 38 *Let v be a vector field over q with locally bounded variation and J any subinterval of I. Then for all t_0 in $\mathring{J}$, the two one-sided limits $\lim_{t\to t_0^-} v(t)$ and $\lim_{t\to t_0^+} v(t)$ exist in TQ. They satisfy:*

$$\Pi_Q\left(\lim_{t\to t_0^-} v(t)\right) = \Pi_Q\left(\lim_{t\to t_0^+} v(t)\right) = q(t_0),$$

and are denoted respectively by $v^-(t_0)$ and $v^+(t_0)$. They can be distinct only at points belonging to a countable subset of J. If $\mathrm{Var}(v;J)$ is finite, such one-sided limits exist also at the two end-points of J, even if the end-points do not belong to J.

We denote by $C^0_c(I,q;TQ)$ the space of continuous vector fields over q with compact support (similar definition for $C^0_c(I,q;T^*Q)$). By definition, a vector valued measure over q is any linear functional μ on $C^0_c(I,q;T^*Q)$ enjoying the following continuity property:

$$\forall\, a<b\in I,\quad \exists\, M_{a,b}\geq 0,\quad \forall\varphi \text{ with } \mathrm{Supp}\,\varphi\subset[a,b],\qquad |\mu(\varphi)|\leq M_{a,b}\max_{t\in I}\|\varphi(t)\|^*_{q(t)}.$$

The real number $\mu(\varphi)$ will also be denoted by $\int_I\langle\varphi(t),\mu\rangle_{q(t)}$. For μ being a vector valued measure over q, the definitions of $\mathrm{Supp}\,\mu$ and $\flat\mu$ are straightforward.

Definition 39 *Let v be a vector field over q with locally bounded variation and θ_{t_0} as in formula (1.A.2). Then, the linear form on $C^0_c(I,q;T^*Q))$ defined by:*

$$\varphi^*\mapsto\int_I(\tau_{t_0,s}(\sharp\circ\varphi^*(s)),\mathrm{d}\theta_{t_0})_{q(t_0)},$$

turns out to be independent on the particular choice of t_0 and define a vector valued measure over q, denoted by Dv and called the covariant Stieltjes measure of v. In any local chart, we have:

$$Dv = \left[dv^i + \Gamma^i_{jk}(q(t))v^j(t)\dot{q}^k(t)\,dt \right] e_i(q(t)).$$

Proposition 40 *Let v be a vector field over q with locally bounded variation. Then, v^- and v^+ are also vector fields over q with locally bounded variation and the following identities hold:*

$$(v^+)^+ = v^+, \quad (v^+)^- = v^-, \quad (v^-)^+ = v^+, \quad (v^-)^- = v^-,$$
$$Dv^- = Dv^+ = Dv.$$

Proposition 41 *Let v and w be two vector fields over q with locally bounded variation. Then, the function $t \mapsto (v(t), w(t))_{q(t)}$ is a real valued function with locally bounded variation over I and we have:*

$$d\left(v(t), w(t)\right)_{q(t)} = \left(\frac{v^-(t) + v^+(t)}{2}, Dw\right)_{q(t)} + \left(Dv, \frac{w^-(t) + w^+(t)}{2}\right)_{q(t)}.$$

Definition of the class *MMA*

In this section, Q is a d-dimensional Riemannian manifold.

Definition 42 *We denote by $MMA(I;Q)$ (motions with measure acceleration) the set of all locally absolutely continuous motions $q : I \to Q$ such that the right velocity $\dot{q}^+(t)$ exists (in the classical sense) for all t in I and defines a vector field over q with locally bounded variation.*

The following proposition ensures the consistency of our notations.

Proposition 43 *Let q be in $MMA(I;Q)$. Then, $\dot{q}^+ : I \to TQ$ is right continuous:*

$$\forall t \in \overset{\circ}{I}, \quad (\dot{q}^+(t))^+ = \dot{q}^+(t).$$

Moreover, $q(t)$ admits a left velocity vector at each instant and:

$$\forall t \in \overset{\circ}{I}, \quad \dot{q}^-(t) = (\dot{q}^+(t))^-.$$

Proof. Use the Mean Value Inequality in a local chart. □

Appendix: Some convex analysis

In this appendix, we do not aim at providing a systematic list of the main theorems in convex analysis. We just want to recall those that are needed in the paper and also to prove some technical results which, if they had been proved in the paper when needed, could have masked the logical train of ideas.

Basic convex analysis

We denote by $(E, \|\cdot\|)$ a finite-dimensional normed vector space. The dual will be denoted by $(E^*, \|\cdot\|^*)$. The bidual of E is systematically identified with E. We briefly recall some

standard definitions and results whose proofs may be found, for example, in ROCKAFELLAR (1970).

A function $\phi : E \to \mathbb{R} \cup \{+\infty\}$ is said *proper* if its domain:

$$\mathrm{Dom}\,\phi \stackrel{\text{def}}{=} \{v \in E\,;\ \phi(v) \neq +\infty\}\,,$$

is non-void. A function $\phi : E \to \mathbb{R} \cup \{+\infty\}$ is said *lower-semi-continuous* if

$$\phi^{-1}(]-\infty, \lambda]) = \{v \in E\,;\ \phi(v) \leq \lambda\}$$

is closed in E for all $\lambda \in \mathbb{R}$. If ϕ is convex, its domain is convex. If ϕ is lower-semi-continuous, its domain does not need to be closed. If C is any convex subset of E, then C is the domain of its indicator function I_C:

$$I_C(v) = \left| \begin{array}{ll} 0 & \text{if } v \in C \\ +\infty & \text{if } v \notin C \end{array} \right.$$

which is convex. It is lower-semi-continuous if and only if C is closed. A function $\phi : E \to \mathbb{R} \cup \{+\infty\}$ is said *positively homogeneous* if:

$$\forall \lambda \in \mathbb{R}^+,\ \forall v \in E \quad \phi(\lambda v) = \lambda \phi(v).$$

If $\phi : E \to \mathbb{R} \cup \{+\infty\}$ is a proper convex function, its conjugate (or dual) function $\phi^* : E^* \to \mathbb{R} \cup \{+\infty\}$ is defined by:

$$\phi^*(v^*) = \sup_{v \in E} \{\langle v^*, v\rangle - \phi(v)\}\,.$$

It is a proper convex function which is, in addition, lower-semi-continuous. Identifying E with its bidual, the conjugacy is a one-to-one, involutive correspondence in the class of all proper lower-semi-continuous convex functions. Moreover, it maps proper, lower-semi-continuous, convex, positively homogeneous functions onto the class of indicator functions of non-void closed convex sets, and reciprocally. Let $\phi : E \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function. For C being a closed convex subset of E, we define the *support function* S_C of C as being the conjugate function of the indicator function I_C of C. The following result is obvious.

Proposition 44 *Let C_1 and C_2 be two closed convex subsets of E. We have:*

$$S_{C_1 + C_2} = S_{C_1} + S_{C_2}$$

Proposition 45 *Let C be a closed convex subset of E. Then, C is bounded if and only if its support function S_C does not take the value $+\infty$. In that case, S_C is L-Lipschitzian with:*

$$L = \max_{\|v\|^* = 1} S_C(v) = \max_{v \in E^* \setminus \{0\}} \frac{S_C(v)}{\|v\|^*} = \max_{v \in C} \|v\|$$

Definition 46 *The subdifferential $\partial\phi(v)$ of ϕ at point v is:*

$$\partial\phi(v) = \{v^* \in E^*\,;\ \forall w \in E,\ \phi(w) \geq \phi(v) + \langle v^*, w - v\rangle\}\,.$$

It is a closed convex subset of E^*.

Let ϕ and ψ be two proper, lower-semi-continuous, convex functions. We obviously have:

$$\partial\phi + \partial\psi \subset \partial(\phi + \psi),$$

but the equality does not hold in general. However, a sufficient condition to get the equality is that one of the two functions has domain E.

Proposition 47 *Let ϕ and ψ be two proper, lower-semi-continuous, convex functions. If* $\operatorname{Dom}\psi = E$, *then*

$$\partial\phi + \partial\psi = \partial(\phi + \psi).$$

Actually, in the finite-dimensional case that we consider, a convex function with domain E is necessarily lower-semi-continuous and even more continuous.

Proposition 48 *Let $\phi : E \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex, lower-semi-continuous function. We have:*

$$v^* \in \partial\phi(v) \iff v \in \partial\phi^*(v^*) \iff \phi(v) + \phi^*(v^*) = \langle v^*, v\rangle.$$

Most of the above definitions and results can be extended to the infinite-dimensional case. We shall need few results of this sort. In the following proposition, a proper convex lower-semi-continuous function is built on the space $L^2(0, T; E)$. The proof can be found in BREZIS (1973), p. 47).

Proposition 49 *Let ϕ be a proper, lower-semi-continuous, convex function on E and μ be a non-negative integrable function on $[0, T]$. For $u \in L^2(0, T; E)$, we define:*

$$\Phi(u) = \begin{cases} \displaystyle\int_0^T \mu(t)\phi(u(t))\,\mathrm{d}t & \text{if } \mu\phi(u) \in L^1(0, T; \mathbb{R}), \\ +\infty & \text{otherwise.} \end{cases}$$

Then Φ is a proper, lower-semi-continuous, convex function on $L^2(0, T; E)$. Moreover, we have:

$$\partial\Phi(u) = \left\{v \in L^2(0, T; E^*)\ ;\ v(t) \in \partial\mu(t)\phi(u(t)),\quad \text{for a.e. } t \in [0, T]\right\}.$$

Also, it is clear that, if, in addition, ϕ is positively homogeneous with domain E and $\mu \in L^2$, then $\operatorname{Dom}\Phi = L^2(0, T; E)$. Proposition 47 holds true in the case where E is a Hilbert space, possibly infinite-dimensional (see BREZIS (1973), p. 41).

Proposition 50 *Let ϕ and ψ be two proper, lower-semi-continuous, convex functions on a Hilbert space H. If* $\operatorname{Dom}\psi = H$, *then*

$$\partial\phi + \partial\psi = \partial(\phi + \psi).$$

Note that in the case where H is infinite-dimensional, a convex function with domain H needs not be lower-semi-continuous.

Evolution problems associated with subdifferentials

The systematic reference for this section is BREZIS (1973). In the sequel, for H being a Hilbert space, we shall systematically identify the dual of H with H.

First, we recall a well-known result whose proof is to be found, for example, in BREZIS (1973), p. 54.

Theorem 51 *Let H be a Hilbert space, $\varphi : H \to \mathbb{R} \cup \{+\infty\}$ be a convex, proper, lower-semi-continuous function, and u_0 be any element of Dom φ. Then, there exists a unique $u \in W^{1,\infty}(0, T; H)$ such that:*

- $u(0) = u_0$,
- $-\dot{u}(t) \in \partial\varphi(u(t))$, *for a.e.* $t \in [0, T]$.

Moreover,

- *the solution* u *admits a right-derivative* $\dot{u}^+(t)$, *at all* $t \in [0, T[$ *and:*

$$\forall\, t \in [0, T[, \quad \dot{u}^+(t) + \mathrm{Proj}\,[0; \partial\varphi(u(t))] = 0,$$

- *the function:*

$$t \mapsto \|\mathrm{Proj}\,[0; \partial\varphi(u(t))]\|_H = \min_{v \in \partial\varphi(u(t))} \|v\|_H$$

is non-increasing.

Now, we are going to derive a modified version of theorem 51 which is adapted to our needs. In the sequel, we denote by $(u, v) = {}^t u \cdot v$ the canonical scalar product of $\mathbb{R}^n$ and by $|\cdot|$ the associated norm. The n^2-dimensional space of real square matrices of order n is denoted by $\mathcal{M}_n(\mathbb{R})$.

We are given some data as follows.

- $\varphi_0 : \mathbb{R}^n \to \mathbb{R}^+ \cup \{+\infty\}$ is convex, proper, lower-semi-continuous and positively homogeneous.
- $\varphi_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, 2, \ldots, m)$ are convex and positively homogeneous.
- $\mu_i \in W^{1,\infty}(0, T; \mathbb{R}^+)$ $(i = 1, 2, \ldots, m)$.
- $f \in W^{1,\infty}(0, T; \mathbb{R}^n)$.
- $G \in W^{1,\infty}(0, T; \mathcal{M}_n(\mathbb{R}))$ is such that $G(t)$ is symmetric, positive definite, for all $t \in [0, T]$.
- $u_0 \in \mathrm{Dom}\,\varphi_0$.

By proposition 45, we have that the functions φ_i $(i = 1, 2, \ldots, m)$ are all L-Lipschitzian for some L. We denote by $\lambda_G^{\min} > 0$ (respectively $\lambda_G^{\max}$) the minimum (respectively the maximum) of all the eigenvalues of $G(t)$ for t wandering in $[0, T]$.

Proposition 52 *There exists a unique* $u \in W^{1,\infty}(0, T; \mathbb{R}^n)$ *such that:*

- $u(0) = u_0$,
- $-G(t) \cdot \dot{u}(t) - f(t) \in \partial\varphi_0(u(t)) + \sum_{i=1}^m \mu_i(t)\partial\varphi_i(u(t))$, *for a.e.* $t \in [0, T]$.

Moreover,

$$\|\dot{u}\|_{L^\infty} \leq \frac{\lambda_G^{\max}}{\lambda_G^{\min}} e^{\left\|\dot{G}\right\|_{L^\infty} T/\lambda_G^{\min}} \left\{ \min_{v \in \partial\varphi_0(u_0)} |v| + \|f\|_{L^\infty} + L\sum_{i=1}^m \|\mu_i\|_{L^\infty} + T\left\|\dot{f}\right\|_{L^\infty} + TL\sum_{i=1}^m \|\dot{\mu}_i\|_{L^\infty} \right\} \quad (\stackrel{\text{def}}{=} C_1). \tag{1.B.1}$$

Finally, if u *is the solution associated with the data* (G, f, μ_i) *and* $\tilde{u}$ *the one associated with the data* $(\tilde{G}, \tilde{f}, \tilde{\mu}_i)$, *then, for all* $t \in [0, T]$, *the following estimate holds:*

$$|\tilde{u}(t) - u(t)| \leq e^{\left\|\dot{\tilde{G}}\right\|_{L^\infty} T/\lambda_{\tilde{G}}^{\min}} \left\{ \frac{C_1}{\lambda_{\tilde{G}}^{\min}} \int_0^t \left|\tilde{G}(s) - G(s)\right|_{\mathcal{M}_n(\mathbb{R})} \mathrm{d}s + \frac{L+1}{\lambda_{\tilde{G}}^{\min}} \int_0^t \left[\left|\tilde{f}(s) - f(s)\right| + \sum_{i=1}^m |\tilde{\mu}_i(s) - \mu_i(s)| \right] \mathrm{d}s \right\}. \tag{1.B.2}$$

Proof. The proof of proposition 52 is derived from theorem 51 by means of very classical arguments.

To prove uniqueness of solution, consider two solutions u and $\tilde{u}$ and define $\delta(t) = \tilde{u}(t) - u(t)$. As a consequence of the monotonicity of subdifferentials, we have easily:

$$\delta(t) \cdot G(t) \cdot \delta(t) \leq \frac{\left\|\dot{G}\right\|_{L^\infty}}{\lambda_G^{\min}} \int_0^t \delta(s) \cdot G(s) \cdot \delta(s) \, \mathrm{d}s,$$

for all $t \in [0, T]$. Applying the Gronwall lemma (lemma 4), we obtain that $\delta(t)$ vanishes identically. Therefore, the functions u and $\tilde{u}$ coincide identically.

To prove existence, we define, for all $N \in \mathbb{N}$ an approximant u_N of the solution in the following way. First, we require $u_N(0) = 0$. Next, we define u_N on $[(k-1)T/2^N, kT/2^N]$ successively for $k = 1, 2, \ldots 2^N$ by:

$$\text{for a.e. } t \in \left[\tfrac{(k-1)T}{2^N}, \tfrac{(k-1)T}{2^N}\right],$$

$$-G\left(\tfrac{(k-1)T}{2^N}\right) \cdot \dot{u}_N(t) - f\left(\tfrac{(k-1)T}{2^N}\right) \in \partial\varphi_0(u_N(t)) + \sum_{i=1}^m \mu_i\left(\tfrac{(k-1)T}{2^N}\right) \partial\varphi_i(u_N(t)).$$

To see that u_N is well-defined, it is enough to apply proposition 47 and theorem 51 with H being $\mathbb{R}^n$ equipped with the scalar product induced by the matrix $G((k-1)T/2^N)$ and the function φ being defined by:

$$\varphi(v) = {}^t f\left(\tfrac{(k-1)T}{2^N}\right) \cdot v + \varphi_0(v) + \sum_{i=1}^m \mu_i\left(\tfrac{(k-1)T}{2^N}\right) \varphi_i(v).$$

It is obvious that, for all $N \in \mathbb{N}$, $u_N \in W^{1,\infty}(0, T; \mathbb{R}^n)$. Also, by use of the second part of theorem 51, we obtain, after a tedious but easy calculation:

$$\forall N \in \mathbb{N}, \quad \|\dot{u}_N\|_{L^\infty} \leq C_1, \tag{1.B.3}$$

where C_1 is the real constant defined in the statement of proposition 52. Now, we are given two arbitrary integers $M \geq N$. A standard but tedious calculation yields:

$$\forall t \in [0, T], \quad \frac{1}{2} |u_M(t) - u_N(t)|^2 \leq \frac{C_2}{2^N} \int_0^t |u_M(s) - u_N(s)| \, \mathrm{d}s$$

where C_2 is a real constant which does not depend on M and N. Actually, we may take:

$$C_2 = \frac{T}{\lambda_G^{\min}} e^{\|\dot{G}\|_{L^\infty} T/\lambda_G^{\min}} \left\{ \left\|\dot{f}\right\|_{L^\infty} + L \sum_{i=1}^m \|\dot{\mu}_i\|_{L^\infty} \right\}$$

Applying lemma 5, we obtain that the sequence u_N converges in the Banach space $C^0([0, T]; \mathbb{R}^n)$ towards a limit u. Coming back to uniform estimate (1.B.3), we can conclude that $u \in W^{1,\infty}(0, T; \mathbb{R}^n)$ and also that a subsequence of $(\dot{u}_N)$ converges towards $\dot{u}$ in L^∞ weak-*. This yields estimate (1.B.1), but there remains to prove that u is a solution of the considered evolution problem. Using the lower-semi-continuity of φ_0, the Fatou lemma and the convergence properties of the sequence (u_N), we easily establish that, for all $v \in L^1(0, T; \mathbb{R}^n)$,

$$-\int_0^T (G \cdot \dot{u} + f) \cdot (v - u) \leq \int_0^T \left\{ \varphi_0(v) - \varphi_0(u) + \sum_{i=1}^m \mu_i(t) \left(\varphi_i(v) - \varphi_i(u)\right) \right\}.$$

Next, there remains to use propositions 49 and 50 to deduce that, for almost every $t \in [0,T]$,

$$-G(t)\cdot\dot{u}(t) - f(t) \in \partial\varphi_0(u(t)) + \sum_{i=1}^{m} \mu_i(t)\partial\varphi_i(u(t))$$

and, so, that u solves the considered evolution problem.

To prove the last estimate of proposition 52, we take the sum of the inequalities:

$$-\left(G\cdot\dot{u}+f\right)\cdot(\tilde{u}-u) \leq \varphi_0(\tilde{u}) - \varphi_0(u) + \sum_{i=1}^{m} \mu_i\left(\varphi_i(\tilde{u}) - \varphi_i(u)\right),$$

and

$$-\left(\tilde{G}\cdot\dot{\tilde{u}}+\tilde{f}\right)\cdot(u-\tilde{u}) \leq \varphi_0(u) - \varphi_0(\tilde{u}) + \sum_{i=1}^{m} \mu_i\left(\varphi_i(u) - \varphi_i(\tilde{u})\right).$$

We obtain:

$$\begin{aligned}\left(\dot{\tilde{u}}-\dot{u}\right)\cdot\tilde{G}\cdot(\tilde{u}-u) \quad &\leq \quad \left(f-\tilde{f}\right)\cdot(\tilde{u}-u) + \sum_{i=1}^{m}\left(\mu_i-\tilde{\mu}_i\right)\left(\varphi_i(\tilde{u})-\varphi_i(u)\right)\\ &\quad + \dot{u}\cdot\left(G-\tilde{G}\right)\cdot(\tilde{u}-u),\end{aligned}$$

which yields:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{2}(\tilde{u}-u)\cdot\tilde{G}\cdot(\tilde{u}-u)\right] \quad &\leq \quad \left[(L+1)\left(|\tilde{f}-f| + \sum_{i=1}^{m}|\tilde{\mu}_i-\mu_i|\right)\right.\\ &\quad \left. + C_1|\tilde{G}-G| + \|\dot{\tilde{G}}\|_{L^\infty}|\tilde{u}-u|\right]|\tilde{u}-u|.\end{aligned}$$

To reach the desired conclusion, it is enough to integrate over $[0,t]$ and to apply successively lemma 5 and lemma 4. □

References

[1] R. ABRAHAM & J.E. MARSDEN (1985), *Foundations of Mechanics*, Addison-Wesley.

[2] P. BALLARD (2000), The dynamics of discrete mechanical systems with perfect unilateral constraints, *Archive for Rational Mechanics and Analysis*, **154**, pp 199–274.

[3] A. BRESSAN (1960), Incompatibilità dei Teoremi di Esistenza e di Unicità del Moto per un Tipo molto Comune e Regolare di Sistemi Meccanici, *Annali della Scuola Normale Superiore di Pisa*, Serie III, Vol. XIV, pp 333–348.

[4] H. BREZIS (1973), *Opérateurs Maximaux Monotones et Semi-groupes de Contractions dans les Espaces de Hilbert*, North-Holland Publishing Company.

[5] I. CHAVEL (1993), *Riemannian Geometry: a Modern Introduction*, Cambridge University Press.

[6] E.A. CODDINGTON & N. LEVINSON (1955), *Theory of Ordinary Differential Equations*, McGraw-Hill Book Company.

[7] P. LÖTSTEDT (1981), Coulomb Friction in Two-Dimensional Rigid Body Systems, *Z. Angew. Math. u. Mech.*, **61**, pp 605–615.

[8] P. LÖTSTEDT (1982), Mechanical Systems of Rigid Bodies subject to Unilateral Constraints, *SIAM J. Appl. Math.*, **42**, no 2, pp 281–296.

[9] M.D.P. MONTEIRO MARQUES (1993), *Differential Inclusions in Nonsmooth Mechanical Problems*, Birkhaüser Verlag, Basel-Boston-Berlin.

[10] J.J. MOREAU (1983), Standard inelastic shocks and the dynamics of unilateral constraints, in *Unilateral problems in structural analysis* (G. Del Piero and F. Maceri Eds), Springer-Verlag, Wien, New-York, pp 173–221.

[11] J.J. MOREAU (1988a), Unilateral contact and dry friction in finite freedom dynamics, in *Nonsmooth Mechanics and Applications*, CISM Courses and Lectures No 302 (J.J. Moreau and P.D. Panagiotopoulos Eds), Springer-Verlag, Wien, New-York, pp 1–82.

[12] J.J. MOREAU (1988b), Bounded variation in time, in *Topics in Nonsmooth Mechanics* (J.J. Moreau, P.D. Panagiotopoulos, G. Strang, Eds.), Birkhaüser Verlag, Basel-Boston-Berlin, pp 1–74.

[13] D. PERCIVALE (1985), Uniqueness in the Elastic Bounce Problem, I, *Journal of Differential Equations*, **56**, pp 206–215.

[14] D. PERCIVALE (1991), Uniqueness in the Elastic Bounce Problem, II, *Journal of Differential Equations*, **90**, pp 304–315.

[15] R.T. ROCKAFELLAR (1970), *Convex Analysis*, Princeton University Press.

[16] W. RUDIN (1966), *Real and complex analysis*, McGraw-Hill.

[17] M. SCHATZMAN (1978), A Class of Nonlinear Differential Equations of Second Order in Time, *Nonlinear Analysis, Theory, Methods & Applications*, **2**, No 2, pp 355–373.

[18] M. SCHATZMAN (1998), Uniqueness and continuous dependence on data for one dimensional impact problems, *Mathematical and Computational Modelling*, **28**, No. 4–8, pp 1–18.