Hamiltonian theory of constrained impulsive motion

Jorge Cortés
Department of Applied Mathematics and Statistics
University of California, Santa Cruz
1156 High Street, Santa Cruz, CA 95064, USA
jcortes@ucsc.edu

Alexandre M. Vinogradov
Dipartimento di Matematica e Informatica
Università di Salerno
Via S. Allende, I-84081 Baronissi, Italy
vinograd@unisa.it

March 7, 2006

Abstract

This paper considers systems subject to nonholonomic constraints which are not uniform on the whole configuration manifold. When the constraints change, the system undergoes a transition in order to comply with the new imposed conditions. Building on previous work on the Hamiltonian theory of impact, we tackle the problem of mathematically describing the classes of transitions that can occur. We propose a comprehensive formulation of the Transition Principle that encompasses the various impulsive regimes of Hamiltonian systems. Our formulation is based on the partial symplectic formalism, which provides a suitable framework for the dynamics of nonholonomic systems. We pay special attention to mechanical systems and illustrate the results with several examples.

Keywords. Mechanical systems, variable nonholonomic constraints, Transition Principle, impulsive motion

1 Introduction

In this paper we consider the problem of mathematically describing impulsive motions (impacts, collisions, reflection, refractions) of Hamiltonian systems subject to nonholonomic constraints. An impulsive behavior takes place when one or more of the basic ingredients of the Hamiltonian dynamical picture undergoes a drastic change. As an example, one may consider the instant of time when the configuration space of the system collapses instantaneously because of an inelastic collision. Another example is given by a ray of light that splits into reflected and refracted rays when passing from one optical media to another, and so on. In situations like these, the phase trajectory of the system becomes discontinuous and the problem of how to describe this discontinuity arises.

The problem of describing impulsive motion has been extensively studied in classical books such as [2, 29, 31, 32, 36]. In these references, the emphasis is put on the analysis of mechanical systems subject to impulsive forces, and in particular, the study of rigid body collisions by means of Newton and Poisson laws of impact. Impulsive nonholonomic constraints (i.e. constraints whose reaction force is impulsive) are also considered in [14, 30, 36, 42], and, from a geometric perspective, in more recent works such as [13, 19, 22, 33]. If impulsive constraints and impulsive forces are present at the same time, Newton and Poisson approaches have been revealed to be physically inconsistent in certain cases [8, 38]. This surprising consequence of the impact laws is only present when the velocity along the impact surface stops or reverses during collision due to the friction. Energetically consistent hypothesis for rigid body collisions with slip and friction are proposed in [37, 38]. From a design point of view, the interest in systems subject to impulsive forces is linked to the emergence of nonsmooth and hybrid dynamical systems in Control Theory, i.e., systems where continuous and discrete dynamics coexist, see [7, 8, 9, 40] and references therein. Hybrid mechanical systems that locomote by switching between different constraint regimes and are subject to elastic impacts are studied in [10]. Hyper-impulsive control of mechanical systems is analyzed in [18].

Here, we aim at a comprehensive analysis of the various situations which can occur concerning impulsive regimes of nonholonomic Hamiltonian systems. In particular, we focus on two different but complementary cases. The first one deals with a drastic change in the nonholonomic constraints imposed on the system. The second one concerns a drastic change of the Hamiltonian function and includes, in particular, collisions and impacts of nonholonomic systems. The proposed solution is given in terms of a generalized version of the *Transition Principle*. This principle, sketched for the first time in a series of lectures of the second author [6] for discontinuous Hamiltonians, was recently extended to other non-constrained situations in [34, 35] (see also [17] for a related discussion in an optimal control setting). By its very nature, the Transition Principle is a direct dynamic interpretation of the geometric data of the problem. This feature distinguishes it from other approaches. For instance, in Classical Mechanics, the velocity jumps caused by an impact are traditionally derived from some assumptions on the nature of the impulsive forces (see, for instance, [2, 24]). However, these assumptions are not logical consequences of the fundamental dynamical principles and therefore one should really consider them as additional principles for impulsive problems. The distinguishing feature of the Transition Principle is that it gives full credit to the geometry of the nonholonomic Hamiltonian system. This seems reasonable for the impulsive regime, keeping in mind the adequacy of the Hamiltonian description to the dynamical behavior of the system in the absence of impulsive motions. In addition, there are some noticeable advantages deriving from the Transition Principle. First of all, its application gives an exact and direct description of the post-impact state which is of immediate use for both theoretical and computational purposes. Secondly, it is still valid in some situations where standard methods can be hardly applied. In particular, this is the case of Hamiltonian systems describing the propagation of singularities of solutions of partial differential equations (consider, for instance, the example of geometrical optics) [26, 27, 41]. Clearly, no variational or traditional approach can be applied to this very important class of systems.

A second contribution of this paper concerns the formulation of the dynamics of nonholonomic Hamiltonian systems. We make use of the notion of partial symplectic structures introduced in [6] and relate this framework with other modern approaches to nonholonomic systems (see [3, 4, 5, 15, 21, 23, 28, 39] and references therein). One advantage of the partial symplectic formalism is that it allows us to draw clear analogies between the unconstrained and constrained situations. Another advantage is that the treatment of nonlinear constraints can be easily incorporated.

The paper is organized as follows. Section 2 introduces some geometric preliminaries on distributions, constraint submanifolds and partial symplectic structures. In Section 3, we show how any nonholonomic Hamiltonian system possesses an associated partial symplectic structure, and we use this fact to intrinsically formulate the dynamics. We also analyze systems with instantaneous nonholonomic constraints and systems exhibiting discontinuities. In Section 4, we develop a new formulation of the Transition Principle for systems with constraints. We present the novel notion of focusing points with respect to a constraint submanifold and we also introduce the concept of constrained characteristic. Decisive points are defined for each impulsive regime resorting to in, out and trapping points. Section 5 presents a detailed study of the concepts introduced in the previous sections in the case of mechanical systems. We compute the focusing points and the characteristic curves, and present various results concerning the decisive points. We also prove an appropriate version for generalized constraints of the classical Carnot's theorem for systems subject to impulsive forces: if the constraints are linear, we show that the Transition Principle always implies a loss of energy. We conclude this section by showing that if the constraints are integrable, then our formulation of the Transition Principle recovers the solution for completely inelastic collisions [35]. Section 6 presents various examples of the application of the above-developed theory. Finally, Section 7 presents our conclusions and directions for future research.

To ease the exposition, below we make use of the standard notation concerning differential geometry and the Hamiltonian formalism without making explicit reference to any work. In particular, we denote by Λ^i (resp., $D(M)$) the $C^\infty(M)$-module of ith order differential forms (resp., of vector fields) on a manifold M. We use $F^*(\varphi)$ to denote the pullback with respect to a smooth map F of a function or differential form φ. If x is a point of M, then the subscript x refers to the value of the corresponding geometric object at x. For instance, X_x stands for the vector field vector $X \in D(M)$ evaluated at x. The interested reader may consult classical books such as [1, 20, 25] for further reference. We also assume smoothness of all the objects we are dealing with.

2 Preliminaries

In this paper we deal with Hamiltonian systems defined on the cotangent bundle T^*M of an n-dimensional manifold M. In the particular case of a mechanical system, M and T^*M are, respectively, the configuration space and the phase space of the system. As usual, $\pi_M : T^*M \to M$ (or simply π) stands for the canonical projection from T^*M to M, $H \in C^\infty(T^*M)$ for the Hamiltonian function and $X_H \in D(M)$ for the corresponding Hamiltonian vector field. The canonical symplectic structure on T^*M is denoted by $\Omega = \Omega_M$. In canonical coordinates (q^a, p_a), $a = 1, \ldots, n$ of T^*M, the symplectic form reads $\Omega = dq^a \wedge dp_a$.

We say that the Hamiltonian system (M, H) *comes from a Lagrangian system* (M, L) *on* TM if $H = (\mathcal{L}_L^*)^{-1}(E_L)$, where $E_L \in C^\infty(TM)$ is the energy function corresponding to the (hyper-regular) Lagrangian function $L \in C^\infty(TM)$ and $\mathcal{L}_L : TM \to T^*M$ is the associated Legendre map.

If X is a vector field on T^*M, then the map $\alpha_X : T^*M \to TM$ defined by

$$\alpha_X(\theta) = d_\theta\pi(X_\theta) \in T_{\pi(\theta)}M\,, \quad \theta \in T^*M\,,$$

denotes the *anti-Legendre map associated with* X. In standard coordinates, if $X = A^a(q,p)\frac{\partial}{\partial q^a} + B^a(q,p)\frac{\partial}{\partial p_a}$, then α_X reads $\alpha_X(q^a, p_a) = (q^a, A^a(q,p))$. For the Hamiltonian vector field $X = X_H$, we write α_H instead of α_{X_H}, so that

$$\alpha_H : (q,p) \mapsto \left(q, v = \frac{\partial H}{\partial p}\right)\,.$$

It is not difficult to see that if the Hamiltonian system (M, H) comes from a Lagrangian system (M, L), then $\alpha_H = \mathcal{L}_L^{-1}$.

2.1 Distributions and codistributions

Recall that a *distribution* (resp., *codistribution*) on a manifold M is a vector subbundle of TM (resp., of T^*M). The *annihilator* of a distribution $\mathcal{D}$ on M is the codistribution $\mathrm{Ann}(\mathcal{D})$ defined by

$$\mathrm{Ann}(\mathcal{D})_x = \{\theta \in T_x^*M \mid \theta(\xi) = 0, \forall \xi \in \mathcal{D}_x\}\,, \quad x \in M\,.$$

If $\mathcal{D}$ is $(n-m)$-dimensional, the codistribution $\mathrm{Ann}(\mathcal{D})$ is m-dimensional. The *dual bundle* $\mathcal{D}^*$ *of* $\mathcal{D}$ is canonically identified with the cotangent bundle T^*M modulo $Ann(\mathcal{D})$. We will also denote by $\mathcal{D}^\perp$ the orthogonal complement of a distribution $\mathcal{D}$ on T^*M with respect to the symplectic form Ω, i.e.,

$$\mathcal{D}_y^\perp = \{\xi \in T_y(T^*M) \mid \Omega_y(\xi,\eta) = 0\,,\ \forall \eta \in \mathcal{D}_y\}\,, \quad y \in T^*M\,.$$

A vector field $X \in D(M)$ *belongs* to $\mathcal{D}$ if $X_x \in \mathcal{D}_x$ for all $x \in M$. Vector fields belonging to $\mathcal{D}$ constitute a $C^\infty(M)$-module, denoted by $D_{\mathcal{D}}(M)$, which is a submodule of $D(M)$. In the partial symplectic formalism (see Section 2.3 below), they are interpreted as "constrained" vector fields. Dually, denote by $\Lambda^1_{\mathcal{D}}(M)$ the $C^\infty(M)$-module of sections of the bundle $\mathcal{D}^*$ and by $\Lambda^i_{\mathcal{D}}(M)$ its ith exterior product. These are interpreted as "constrained" differential i-forms. We denote the natural restriction map from $\Lambda^i(M)$ to $\Lambda^i_{\mathcal{D}}(M)$ by $r_{\mathcal{D}} : \Lambda^i(M) \to \Lambda^i_{\mathcal{D}}(M)$.

The geometric description of nonholonomic systems in the framework of the partial symplectic formalism [6] requires a slight "affine" generalization of these standard notions. Namely, an *affine distribution* on a manifold M is an affine subbundle Δ of TM. This means that the fiber Δ_x of Δ over $x \in M$ is an affine subspace in T_xM. Therefore, Δ_x can be represented in the form $\Delta_x = v + \Delta^0_x$ with $v \in T_xM$ and Δ^0_x being the vector subspace of T_xM canonically associated with Δ_x. In this representation, the *displacement vector* v is unique modulo Δ^0_x. The union $\cup_{x\in M}\Delta^0_x$ constitutes a linear distribution of the tangent bundle TM, denoted by Δ^0, canonically associated with Δ. It is not difficult to see that there always exist a vector field $Y \in D(M)$ such that Y_x is a displacement vector for Δ_x. Such vector fields are called *displacement vector fields of* Δ. Obviously, displacement vector fields differ by another vector field belonging to Δ. In coordinate terms, an $(n-m)$-dimensional affine codistribution is described by a system of linear equations $\Phi_i = 0$ with respect to the variables p_a, i.e., $\Phi_i(q,p) = \Phi_{ia}(q)p_a + \Phi_{i0}(q)$, $i = 1,\ldots,m$.

Similarly, an *affine codistribution* on M is an affine subbundle $C \subset T^*M$ of the cotangent bundle. As above, one has $C = \Upsilon + C^0$, where C^0 is the unique codistribution on M canonically associated with C, and $\Upsilon \in \Lambda^1(M)$ is a *displacement form.* Point-wisely this means that $C_x = \Upsilon_x + C^0_x$, for all $x \in M$.

2.1.1 Linear constraints

In the case of linear constraints, the analogy between free and constrained systems is particularly clear. In fact, it is natural to interpret an affine distribution (resp., codistribution) on a manifold M as the "constrained" tangent (resp., cotangent) bundle of M. A linearly constrained Hamiltonian system is then a triple (M,H,C), with $H \in C^\infty(T^*M)$ and C an affine codistribution on M. Similarly, a triple (M,L,Δ), with $L \in C^\infty(TM)$ and Δ an affine distribution on M, is a linearly constrained Lagrangian system. The anti-Legendre map allows one to pass from a constrained Hamiltonian system to the corresponding Lagrangian system and vice versa. More precisely, if (M,H,C) is a linearly constrained Hamiltonian system, the map α_H is linear and (M,H) comes from a Lagrangian system (M,L), then the corresponding linearly constrained Lagrangian system is (M,L,Δ), with $\Delta = \alpha_H(C)$. To go in the opposite direction, one must use the Legendre map $\mathcal{L}_L$ instead of α_H.

Throughout the paper, we distinguish the class of mechanical systems subject to linear constraints because of two reasons. First, classically they have been intensively studied. Second, one can extract from them the motivations for the basic constructions which will be discussed below.

2.1.2 Nonlinear constraints

In the Hamiltonian setting, the *nonholonomic constraints* are given by a submanifold (not necessarily a vector subbundle) $C \subset T^*M$. Similarly, *nonholonomic Lagrangian or kinematic constraints* are given by a submanifold $C' \subset TM$. If the Hamiltonian system (M, H) comes from a Lagrangian system (M, L), then $C = \mathcal{L}_L(C')$ if and only if $C' = \alpha_H(C)$. In mechanics, these two approaches correspond to two possible descriptions of nonholonomic constraints: either as limitations imposed on the momenta or as limitations imposed on the velocities, respectively. The fact that C (resp., C') represents limitations imposed only on the momenta (resp., velocities), but not on the configurations of the system, implies that the projection π must send C (resp., C') surjectively onto M. However, the assumption of "infinitesimal surjectivity" of $\pi|_C$ is more adequate in this context. This means that $\pi|_C$ is a *submersion*, i.e., $d_y(\pi|_C) : T_yC \to T_{\pi(y)}M$ is surjective for all $y \in C$. With this motivation, we adopt the following definition.

Definition 2.1 *A set of* nonholonomic constraints *imposed on a Hamiltonian system (M, H) is a submanifold $C \subset T^*M$ such that $\pi|_C$ is a submersion. The constrained Hamiltonian system is denoted by (M, H, C).*

Since C and $\alpha_H(C)$ are, respectively, interpreted as the constrained cotangent and tangent bundle of the system (M, H, C), we will always assume that they have equal dimensions. It is worth stressing that the above definition makes also sense for manifolds with boundary. In such a case, the boundary of T^*M is $\pi^{-1}(\partial M)$ and the boundary of C is $\partial C = C \cap \pi^{-1}(\partial M)$.

Remark 2.2 Similarly, *nonholonomic Lagrangian constraints* are represented by submanifolds of TM that project regularly onto M.

In what follows, $\Phi_i(q, p) = 0$, $i = 1, ..., m$, will denote a set of local equations defining C. For a point $x \in M$, we denote by C_x the fiber of C at x,

$$C_x = C \cap T_x^*M = \{y \in C \mid \pi(y) = x\} = (\pi_{|C})^{-1}(x)\,.$$

2.1.3 Instantaneous nonholonomic constraints

Let N be a hypersurface in M. Consider the induced hypersurface $T_N^*M = \pi^{-1}(N)$ of T^*M. Let $C \subset T^*M$ be a set of nonholonomic constraints on M. Instantaneous constraints may be thought as limitations on the momenta (resp., velocities) of the system that are imposed only at the instant when a trajectory passes through a point of N. Therefore, they are represented by a submanifold C^{inst} of $C \cap T_N^*M$. These constraints are assumed to be additional to the ones already prescribed by C. In order to admit an adequate mechanical interpretation, we also assume that the projection π restricted to C^{inst} is a submersion onto N. From the Lagrangian point of view, instantaneous kinematic constraints are naturally interpreted as a submanifold $C^{\mathrm{inst}\prime}$ of TN. Based on these considerations, we take the following definition.

Definition 2.3 *Let (M, H, C) be a constrained Hamiltonian system and let N be a hypersurface of M. A set of* instantaneous constraints along N *imposed on (M, H, C) is a submanifold C^{inst} of $C \cap T_N^*M$ such that π restricted to C^{inst} is a submersion onto N.*

It is worth stressing that, in some cases, a set of instantaneous constraints along N additionally verifies the condition $C^{\text{inst}} \subset \alpha_H^{-1}(TN)$ (here TN is thought to be naturally embedded into TM). In an inelastic scenario, where the nonholonomic motion in M is forced to take place in N after the impact, this latter condition formalizes the parity between the Hamiltonian and Lagrangian approaches: if the Hamiltonian system in question comes from a Lagrangian one, then $C^{\text{inst}} = \mathcal{L}_L(C^{\text{inst}\prime})$, with $C^{\text{inst}\prime}$ being the instantaneous kinematic constraints.

2.2 Dynamics of Hamiltonian systems

As is well-known, in the absence of constraints, the dynamics of the Hamiltonian system (M, H) is given by the Hamiltonian vector field X_H, whose coordinate description is

$$\frac{dq^a}{dt} = \frac{\partial H}{\partial p_a}, \quad \frac{dp_a}{dt} = -\frac{\partial H}{\partial q^a}, \quad a = 1, \ldots, n\,.$$

In the presence of constraints, the "free" Hamiltonian vector field X_H must be modified along the constraint manifold C in order to become tangent to C. In the traditional approach this goal is achieved by adding to X_H another vector field along C, say, R, interpreted as the *reaction of constraints*. From a purely geometrical point of view, the choice of a vector field that makes X_H tangent to C is far from being unique. Therefore, a new principle must be invoked to select the one that merits to be called the "reaction of constraints". The history of this problem (see, for instance, [30]) shows that its solution is not straightforward. By applying, for instance, the Lagrange-d'Alembert principle (see [4, 15, 30]), one gets the following equations of motion

$$\frac{dq^a}{dt} = \frac{\partial H}{\partial p_a}, \quad \frac{dp_a}{dt} = -\frac{\partial H}{\partial q^a} + \lambda_i \frac{\partial \Phi_i}{\partial p_a}, \quad \Phi_i(q^a, p_a) = 0\,,$$

$a = 1, \ldots, n$, $i = 1, \ldots, m$, where the "Lagrange multipliers" λ_i's are to be duly determined. Shortcomings of such an approach are that it is not manifestly intrinsic and does not reveal clearly the geometric background of the situation. This is why in our further exposition we shall follow a purely geometric approach, which does not require any discussion of reactions of constraints. It is based on the concept of *partial symplectic formalism*, which also appears to be more concise from an algorithmic point of view.

2.3 Partial symplectic structures

The following elementary facts from linear algebra will be most useful. Let V be a vector space, $W \subset V$ a subspace and $b : V \times V \to \mathbb{R}$ a bilinear form on V. Denote by $W_b^{\perp}$ the b-orthogonal complement of W,

$$W_b^{\perp} = \{v \in V \mid b(v, w) = 0, \forall w \in W\}\,.$$

Note that $W \cap W_b^{\perp} = 0$ if and only if the restriction $b\,|_W$ of b to W is nondegenerate. The form b is said to be *nondegenerate* on an affine subspace U of V, $U = p_0 + W$, $p_0 \in V$, if it is nondegenerate on its associated vector space W. In such a case, U can be uniquely represented in the form $U = p_1 + W$ with $p_1 \in W_b^{\perp}$ due to the fact that $U \cap W_b^{\perp} = \{p_1\}$. The vector p_1 is called the *canonical displacement* of U with respect to b. Consider the associated map

$$\top_{W,b} : W \longrightarrow W^*, \qquad \top_{W,b}(w) = b(w, \cdot)\,,\ w \in W.$$

In other words, $\top_{W,b}(w)(w') = b(w,w')$, for all $w' \in W$. Obviously, $\top_{W,b}$ is an isomorphism if and only if $b_{|W}$ is nondegenerate.

If b is skew-symmetric and nondegenerate on V, and W is a subspace of V with codimension one, then the kernel of the restricted form $b|_W$, $\ker b|_W$ is a 1-dimensional subspace, i.e., a *line* in V contained in W. Therefore, $\ker b|_W = W_b^{\perp}$.

Let now Δ be an affine distribution on a manifold Q. A form $\omega \in \Lambda^2(Q)$ is called *nondegenerate* on Δ if $b = \omega_x$ is nondegenerate on $U = \Delta_x$, for all $x \in Q$. In such a case, there exists a unique vector field $Y \in D(Q)$ such that Y_x is the canonical displacement of Δ_x with respect to ω_x, for all $x \in Q$. The vector field $Y = Y_{\Delta,\omega}$ is called the *canonical displacement* of Δ with respect to ω. If $\omega \in \Lambda^2(Q)$ is nondegenerate on Δ, then one has the isomorphism of vector bundles

$$\gamma = \gamma_{\Delta^0,\omega} : \Delta^{0*} \to \Delta^0, \qquad \gamma_x = -(\top_{\Delta^0_x,\omega_x})^{-1} : \Delta^{0*}_x \to \Delta^0_x.$$

Passing to sections of these bundles, one gets the isomorphism of $C^\infty(Q)$-modules $\Gamma_{\Delta^0,\omega} : \Lambda^1_{\Delta^0}(Q) \to D_{\Delta^0}(Q)$ defined by

$$\Gamma = \Gamma_{\Delta^0,\omega}(\varrho)(x) = \gamma(\varrho(x)), \quad \varrho \in \Lambda^1_{\Delta^0}(Q). \tag{1}$$

Definition 2.4 *A* partial symplectic structure *on a manifold Q is a pair (Δ, ω) consisting of an affine distribution Δ on Q and a closed 2-form $\omega \in \Lambda^2(Q)$ which is nondegenerate on Δ.*

Given a partial symplectic structure $\Theta = (\Delta, \omega)$, we will use the subscript Θ to denote the associated objects: $\Delta_\Theta = \Delta, \omega_\Theta = \omega, Y_\Theta = Y_{\Delta,\omega}$ and Δ^0_Θ for the distribution canonically associated to Δ. We also write

$$r_\Theta = r_{\Delta^0_\Theta}, \quad D_\Theta = D_{\Delta^0_\Theta}(Q), \quad \Lambda^1_\Theta = \Lambda^1_{\Delta^0_\Theta}(Q), \quad \Gamma_\Theta = \Gamma_{\Delta^0_\Theta,\omega} : \Lambda^1_\Theta \to D_\Theta.$$

In the partial symplectic formalism, the elements of $C^\infty(Q)$-modules D_Θ and Λ^1_Θ may be viewed as "constrained" vector fields and differential forms, respectively. The *constrained Hamiltonian vector field* associated with a Hamiltonian function $H \in C^\infty(Q)$ is defined as

$$X^\Theta_H = \Gamma_\Theta(r_\Theta(dH)) + Y_\Theta. \tag{2}$$

The *almost-Poisson* bracket associated to the partial symplectic structure Θ is

$$\{f,g\}_\Theta = \Gamma_\Theta(r_\Theta(df))(g) = X^\Theta_f - Y_\Theta(g)\,, \quad f,g \in C^\infty(Q)\,.$$

The wording "almost" here refers to the fact that this bracket does not satisfy in general the Jacobi identity. However, it is still skew-symmetric and a bi-derivation.

Definition 2.5 *Let $\Theta = (\Delta, \omega)$ be a partial symplectic structure on a manifold Q. A hypersurface $B \subset Q$ is* transversal *to Θ (or to Δ) if the affine subspaces T_yB and Δ_y of T_yQ are transversal for any $y \in C$.*

If B is *transversal* to Θ, then $T_yB \cap \Delta_y$ is of codimension 1 in Δ_y. If Θ is a partial symplectic structure on $C \subset T^*M$, we shall extend this terminology by saying that Θ is *transversal* to a hypersurface $\tilde{B}$ in T^*M if $\tilde{B}$ is transversal to C, so that $B = \tilde{B} \cap C$ is a hypersurface in C, and B is transversal to Θ.

3 Dynamics of nonholonomic Hamiltonian systems

In this section, we formulate the dynamics of nonholonomic Hamiltonian systems using the partial symplectic formalism. We show how, under some technical conditions, any Hamiltonian system subject to nonholonomic constraints possesses an associated partial symplectic structure. Then, we analyze the cases of systems with instantaneous nonholonomic constraints, and systems exhibiting discontinuities.

3.1 The partial symplectic structure associated with a constrained Hamiltonian system

Let (M, H, C) be a constrained Hamiltonian system. Our goal is to associate with it a partial symplectic structure Θ on the "constrained" cotangent bundle C in such a way the corresponding constrained Hamiltonian field X_H^Θ gives the desired nonholonomic dynamics. With this purpose, consider the *constrained symplectic form* defined by the restriction of the "free" symplectic form Ω_M to C

$$\omega_\Theta = j^*(\Omega_M)\,, \tag{3}$$

with $j : C \hookrightarrow T^*M$ the canonical inclusion. The next step is to construct a suitable affine distribution Δ_Θ on C. A natural non-singularity requirement on C is asking for the regularity of the map $\alpha_H|_C$. This is the reason why we assume that $\alpha_H|_C$ is an immersion, i.e., that the differential $d_y\alpha_H$ is nonsingular for any $y \in C$. Since $\alpha_H : T^*M \to TM$ is fibered, this assumption implies that the map $(\alpha_H)_x : C_x \to TM$ is an immersion for any $x \in M$ and vice versa.

Let $y \in C$ and $x = \pi(y)$. Let Π_y be the affine subspace of T_xM tangent to $\alpha_H(C_x)$ at $z = \alpha_H(y)$. Since by the above assumption $\alpha_H|_C$ is an immersion, $\dim \Pi_y = \dim C_x = n - m$. Consider the affine distribution Δ_Θ on C defined by

$$\Delta_{\Theta y} = \{\xi \in T_yC \mid d_y\pi(\xi) \in \Pi_y\} \subset T_yC. \tag{4}$$

Since $d_y(\pi|_C)$ is surjective, the codimension of $\Delta_{\Theta y}$ in T_yC is equal to the codimension of Π_y in T_xM, i.e., to m. Therefore, $\dim \Delta_{\Theta y} = 2(n - m)$. It is not difficult to see now that if the form ω_Θ is nondegenerate on the distribution Δ_Θ, then $\alpha_H|_C$ is an immersion.

Proposition 3.1 *Let $S = (M, H, C)$ be a constrained Hamiltonian system. Then, $\alpha_H|_C$ is an immersion if the pair $(\Delta_\Theta, \omega_\Theta)$ defined by equations* (3) *and* (4) *is a partial symplectic structure.*

The converse, however, is in general not true. Since the partial symplectic structure associated with $S = (M, H, C)$ is determined by H and C, we will simply denote it by $\Theta(H, C) = (\Delta_{H,C}, \omega_C)$.

For most Hamiltonian systems (including those coming from Mechanics), the anti-Legendre map α_H is regular not only when restricted to C, but on the whole space T^*M. In this is the case, and the Hamiltonian system comes from a Lagrangian system, one can indeed show that the condition of ω_Θ being nondegenerate on the distribution Δ_Θ is equivalent to the so-called *compatibility condition* [3, 23]. Therefore, Proposition 3.1 establishes a link between the classical partial symplectic formalism introduced in [6] and more recent approaches as explained, for instance, in [15]. Also note that the class of mechanical systems automatically verifies the compatibility condition, therefore admitting both formulations. Indeed, for mechanical systems, the conditions in Proposition 3.1 are equivalent.

Definition 3.2 *A* nonholonomic Hamiltonian system *on a manifold M is a constrained system (M, H, C), $H \in C^\infty(T^*M)$, $C \subset T^*M$ such that $\Theta(H, C) = (\Delta_{H,C}, \omega_C)$ is a partial symplectic structure.*

The dynamics of a nonholonomic Hamiltonian system is given by the constrained Hamiltonian vector field X_H^{Θ} with respect to the partial symplectic structure $\Theta = \Theta(H, C)$ (cf. equation (2)). This vector field will be denoted by $X_{H,C}$. Under regularity of the map α_H, $X_{H,C}$ reads in canonical coordinates

$$X_{H,C} = \frac{\partial H}{\partial p_a}\frac{\partial}{\partial q^a} - \left(\frac{\partial H}{\partial q^a} + \mathcal{C}_{ij}\left(\frac{\partial H}{\partial p_b}\frac{\partial \Phi^j}{\partial q^b} - \frac{\partial H}{\partial q^b}\frac{\partial \Phi^j}{\partial p_b}\right)\frac{\partial \Phi^i}{\partial p_c}\mathcal{H}_{ca}\right)\frac{\partial}{\partial p_a}\,,$$

where the matrices $(\mathcal{H}_{ab})$ and $(\mathcal{C}_{ij})$ are defined by

$$(\mathcal{H}_{ab}) = \left(\frac{\partial^2 H}{\partial p_a \partial p_b}\right)^{-1}, \quad (\mathcal{C}_{ij}) = \left(\frac{\partial \Phi^i}{\partial p_a}\mathcal{H}_{ab}\frac{\partial \Phi^j}{\partial p_b}\right)^{-1}.$$

Observe that the *force of reaction of nonholonomic constraints* (see Section 2.2) in the partial symplectic framework is defined *a posteriori* as the difference between the constrained and the free Hamiltonian vector fields, $X_{H,C} - X_H$. Also, note that the almost-Poisson bracket associated with the partial symplectic structure $\Theta_{H,C}$ coincides with the so-called nonholonomic bracket [15, 11, 39].

Transversality It is convenient to adapt the terminology related to the notion of transversality discussed in Section 2.3 to the context of nonholonomic Hamiltonian systems. First, we shall say that a nonholonomic Hamiltonian system $S = (M, H, C)$ is *transversal* to a hypersurface B in T^*M if the underlying partial symplectic structure $\Theta(H, C)$ is transversal to B. Second, if N is a hypersurface in M, we shall say that S is *transversal to* N if S is transversal to the hypersurface $T^*_N M$. The following result follows from the definition of the partial symplectic structure $\Theta(H, C)$.

Proposition 3.3 *A nonholonomic Hamiltonian system (M, H, C) is transversal to a hypersurface $N \subset M$ if and only if $\Pi_y \subset T_{\pi(y)}M$, the affine subspace of $T_{\pi(y)}M$ tangent to $\alpha_H(C_{\pi(y)})$ at $z = \alpha_H(y)$, is transversal to $T_{\pi(y)}N \subset T_{\pi(y)}M$ for all $y \in C$.*

3.2 Instantaneous partial symplectic structures

It is intuitive to think that when a trajectory of a Hamiltonian system (M, H, C) crosses a *critical* hypersurface N in the configuration manifold M, its phase space reduces to T^*N. Moreover, it could possibly be subject to additional instantaneous constraints along N. In the language of our approach, this idea is naturally expressed by saying that all such *critical* states constitute a nonholonomic Hamiltonian system on N. Since T^*N is not naturally embedded into T^*M, a realization of this idea is not completely straightforward. What one really needs is a partial symplectic structure on the manifold of instantaneous constraints C^{inst} which, by definition, is a submanifold of T^*M

Namely, let C^{inst} be a set of instantaneous constraints along N imposed on (M, H, C) (cf. Definition 2.3). Take $y \in C^{\mathrm{inst}}$. Let $x = \pi(y)$ and denote by Π_y^{inst} the affine subspace of $T_xN \subset T_xM$ tangent to $\alpha_H(C_x^{\mathrm{inst}})$ at $\alpha_H(y)$. Consider the 2-form $\omega_{\Theta^{\mathrm{inst}}}$ and the affine distribution $\Delta_{\Theta^{\mathrm{inst}}}$ on C^{inst} defined by

$$\omega_{\Theta^{\mathrm{inst}}} = j^*(\Omega_M)\,, \quad \Delta_{\Theta^{\mathrm{inst}}y} = \{\xi \in T_yC^{\mathrm{inst}} \mid d_y\pi(\xi) \in \Pi_y^{\mathrm{inst}}\} \subset T_yC^{\mathrm{inst}}, \tag{5}$$

with $j : C^{\mathrm{inst}} \hookrightarrow T^*M$ the canonical inclusion. We then have the following definition.

Definition 3.4 *Let (M, H, C) be a nonholonomic Hamiltonian system and let C^{inst} be a set of instantaneous constraints along a hypersurface $N \subset M$. The pair $(\Delta_{\Theta^{inst}}, \omega_{\Theta^{inst}})$ defined by* (5) *is called the* instantaneous partial symplectic structure along N *if $\omega_{\Theta^{inst}}$ is not degenerate on $\Delta_{\Theta^{inst}}$. If this is the case, C^{inst} is called a* regular *set of instantaneous constraints.*

Note that this structure is defined by H, C^{inst} and N. To highlight this fact, we denote $\Theta^{\mathrm{inst}} = \Theta^{\mathrm{inst}}(H, C^{\mathrm{inst}}, N)$. Accordingly, we denote by $X_{(H,C^{\mathrm{inst}},N)}$ the constrained Hamiltonian vector field $X^{\Theta^{\mathrm{inst}}}_{H^{\mathrm{inst}}}$, with $H^{\mathrm{inst}} = H|_{C^{\mathrm{inst}}}$.

In general, since $C^{\mathrm{inst}} \subset C$ by definition, one has that $\alpha_H|_{C^{\mathrm{inst}}}$ is an immersion. For mechanical systems, this implies that the 2-form $\omega_{\Theta^{\mathrm{inst}}}$ is nondegenerate on $\Delta_{\Theta^{\mathrm{inst}}}$, and therefore any set of instantaneous constraints is regular.

In what follows, we shall only deal with regular instantaneous nonholonomic constraints. A natural class of instantaneous structures arises in the following situation of particular interest. Assume that the nonholonomic Hamiltonian system $S = (M, H, C)$ is transversal to N and that α_H is regular. Then $\alpha_H^{-1}(TN)$ is transversal to C and, hence,

$$C_{(N,H)} = \alpha_H^{-1}(TN) \cap C$$

is a submanifold of codimension 2 in C. Note that $C_{(N,H)}$ is a set of instantaneous nonholonomic constraints on S along N. By construction, the codimension of $\Delta_{Ny} = \Delta_y \cap T_y(C_{(N,H)})$ in Δ_y is also 2 and Ω_y is nondegenerate when restricted to Δ_{Ny}. Therefore, the affine distribution Δ_N and the 2-form $\Omega|_N$ endow $C_{(N,H)}$ with a partial symplectic structure, which is an instantaneous partial symplectic structure along N. We call it the *trace of S on N* and denote it by $S_{(N,H)}$. In the special case $N = \partial M$, we call it the *boundary of S*, and denote it by ∂S, i.e., $\partial S = S_{(\partial M,H)}$. We will denote the constrained Hamiltonian vector field with respect to the trace (resp., boundary) as $X^{\mathrm{tr}} = X^{\mathrm{tr}}_{(H,C,N)}$ (resp., $X^{\partial} = X^{\partial}_{(H,C,\partial M)}$).

3.3 Discontinuous nonholonomic systems

An impulsive behavior of a Hamiltonian system occurs when its trajectory "tries" to go across a *critical* hypersurface N in the configuration space M. In such an instant, the system may be forced to drastically change its constraints, to pass under the control of another Hamiltonian and/or to be eventually subject to additional instantaneous constraints. Such situations may be interpreted as *discontinuities* on both the constraints and the Hamiltonian of the system. Below, we formalize these concepts properly via the notion of cutting-up.

Definition 3.5 *Let $N \subset M$ be a hypersurface of M with $N \cap \partial M = \emptyset$. A pair $(\hat{M}, \varsigma)$, $\varsigma : \hat{M} \to M$, is called a* cutting-up of M along N *if*

(i) $\hat{N} = \varsigma^{-1}(N) \subset \partial \hat{M}$;

(ii) *ς maps $\hat{M} \setminus \hat{N}$ diffeomorphically onto $M \setminus N$;*

(iii) *$\varsigma|_{\hat{N}} : \hat{N} \to N$ is a double covering of N.*

Note that, by definition, ς is a local diffeomorphism. Cuttings-up for a given N exist and are equivalent one to each other. If N divides M into two parts, say, M_+ and M_-, i.e., $M = M_+ \cup M_-$, $M_+ \cap M_- = N$, then $\hat{M}$ may be viewed as the disjoint union of M_+ and M_-, and ς as the map that matches them together along the common border N. Locally any cutting-up is of this form.

For our purposes, it is important to realize that, if $(\hat{M}, \varsigma)$ is a cutting-up of M along N, then $(T^*\hat{M}, T^*\varsigma)$ is a cutting-up of T^*M along the hypersurface $T^*_N M$. Here, $T^*\varsigma$ denotes the dual of the inverse of the isomorphism $d_z\varsigma : T_z\hat{M} \to T_{\varsigma(z)}M$, for all $z \in \hat{M}$. In the following definition, we introduce the class of Hamiltonian systems we shall be dealing with throughout this paper.

Definition 3.6 *Let $N \subset M$ be a hypersurface of M with $N \cap \partial M = \emptyset$ and let $(\hat{M}, \varsigma)$ be a cutting-up of M along N. A* nonholonomic Hamiltonian system discontinuous along N, *denoted $S = (M, H, C \mid N)$, is the direct image with respect to $T^*\varsigma$ of a nonholonomic Hamiltonian system $(\hat{M}, \hat{H}, \hat{C})$. Such system is called* regular *if $(\hat{M}, \hat{H}, \hat{C})$ is transversal to $\hat{N}$.*

A system of instantaneous nonholonomic constraints on S along N *is the direct image with respect to $T^*\varsigma$ of a set of instantaneous constraints $\hat{C}^{inst}$ along $\hat{N}$ on the associated system $\hat{S} = (\hat{M}, \hat{H}, \hat{C} \mid \hat{N})$. The* trace of S on N *is the direct image with respect to $T^*\varsigma$ of the trace of $\hat{S}$ along $\hat{N}$.*

According to Definition 3.6, $\hat{H}$ is a smooth function on $T^*\hat{M}$ and $\hat{C}$ is a submanifold of $T^*\hat{M}$. Therefore, the *direct image* of $\hat{H}$ along the matching map $T^*\varsigma : T^*\hat{M} \to T^*M$ may be viewed as a function on T^*M, which is 1-valued and smooth outside of $T^*_N M$ and 2-valued and smooth on $T^*_N M$. We will continue to use the notation H for this function and will refer to it as a *discontinuous Hamiltonian along* N. Similarly, the direct image $C = T^*\varsigma(\hat{C})$ of $\hat{C}$ will be referred to as *discontinuous nonholonomic constraints along* N. Outside of $T^*_N M$, C is a "good" smooth submanifold of T^*M, whose boundary is an immersed submanifold of $T^*_N M$.

The previous discussion becomes particularly simple when N divides M into two parts, M_+ and M_-, as mentioned above. In such a case, $T^*_N M$ also divides T^*M into two parts, T^*M_+ and T^*M_-, whose common boundary is $T^*_N M$. Then, a discontinuous Hamiltonian H along N may be naturally seen as a pair of Hamiltonians, say, H_+ and H_-, defined on T^*M_+ and T^*M_-, respectively. Similarly, a set of discontinuous nonholonomic constraints along N is regarded as a pair of sets of nonholonomic constraints $C_\pm \subset T^*M_\pm$. Since N always divides M locally, this description constitutes a local picture of a discontinuous nonholonomic Hamiltonian system along N.

We will continue to use the notation $C^{\rm inst}$ (resp., $S^{\rm tr}$) for instantaneous nonholonomic constraints (resp., the trace of S) in the case of discontinuous nonholonomic systems. As before, one may interpret $C^{\rm inst}$ as a 2-valued system of instantaneous nonholonomic constraints along N. In the case when N divides M into two parts, we will distinguish between the two branches using the notation $C^{\rm inst}_\pm$, and write also $X_{(H, C^{\rm inst}_\pm, N)}$ (resp., $X^{\rm tr}_{(H, C_\pm, N)}$).

Remark 3.7 The impulsive behavior of a Hamiltonian system is not necessarily related to some discontinuity. This type of phenomena occurs, for instance, each time that one of its trajectories "strikes" against the boundary ∂M of the configuration space M. Various kinds of collisions, impacts, etc, in mechanical systems are described in this way. Otherwise said, impulsive behavior is characteristic of Hamiltonian systems *with boundary.* Moreover, systems with boundary may be viewed as a "limit" case of discontinuous systems by dropping the requirement $N \cap \partial M = \emptyset$ and choosing $N = \partial M$, $M_- = \emptyset$, $M_+ = M$. This allows a unified approach to both situations.

4 The Transition Principle

In this section we discuss the formulation of the Transition Principle for systems subject to nonholonomic constraints. We first introduce the notions of focusing points, constrained characteristics, and in, out and decisive points. The Transition Principle builds on these elements to prescribe the behavior of the Hamiltonian system when one or more of its ingredients undergoes a drastic change.

4.1 Focusing points

The following simple linear result will be key for the subsequent discussion.

Lemma 4.1 *Let $y \in T^*M$, and let W be an affine subspace in $T_y(T^*M)$ such that Ω_y is nondegenerate on W (hence, $\dim W = 2l$ for certain l) and $\dim d_y\pi(W) = l$. Denote by $W^0 \subset T_y(T^*M)$ and $d_y\pi(W)^0 = d_y\pi(W^0) \subset T_{\pi(y)}M$ the linear subspaces associated with the affine spaces W and $d_y\pi(W)$, respectively. Then the affine subspaces $W^\bullet = y + Ann(d_y\pi(W)^0)$ and $W_\bullet = y + W^0 \cap T_y(T^*_{\pi(y)}M)$ in $T^*_{\pi(y)}M$ passing through y are transversal.*

Proof: Since, by hypothesis, $\dim d_y\pi(W) = l$, one has

$$\dim W^0 \cap T_y(T^*_{\pi(y)}M) = l \quad \text{and} \quad \dim d_y\pi(W)^0 = l\,.$$

Now, the dimension of $\mathrm{Ann}(d_y\pi(W)^0) \subset T^*_{\pi(y)}M$ is $n - l$. Moreover, $W^0 \cap T_y(T^*_{\pi(y)}M)$ is transversal to $\mathrm{Ann}(d_y\pi(W)^0)$ if one identifies the spaces $T^*_{\pi(y)}M$ and $T_y(T^*_{\pi(y)}M)$. The result now follows. ∎

Consider now a nonholonomic Hamiltonian system (M, H, C). Let $y \in C$. Denote by $\Delta = \Delta_{(H,C)}$ be the affine distribution of the corresponding partial symplectic structure $\Theta(H, C)$ (cf. Section 3.1). By Definition 3.2, the affine subspace $W = \Delta_y$ satisfies the assumptions of Lemma 4.1 on W (observe that $d_y\pi(W)$ is precisely Π_y in equation (4)). Therefore, the subspace $W^\bullet = \Delta^\bullet_y$ is well-defined and we put

$$K_y = K_y(H, C) = \Delta^\bullet_y \subset T^*_{\pi(y)}M\,.$$

Moreover, it is not difficult to see that the subspace $W_\bullet = (\Delta_y)_\bullet$ is identical to $T_yC_{\pi(y)}$. This shows that K_y is transversal to C at y, and that $\dim K_y = m$. The *crown of the nonholonomic Hamiltonian system* (M, H, C) is the map

$$\kappa = \kappa_{H,C} : C \longrightarrow A_m(T^*M)\,, \quad y \mapsto K_y\,,$$

where $A_k(T^*M)$ denotes the manifold whose elements are k-dimensional affine submanifolds contained in the fibers of the cotangent bundle T^*M. One can see that the graph of the crown κ,

$$\mathrm{Graph}(\kappa) = \{(y, v) \in C \times T^*M \mid v \in K_y\}$$

is a $2n$-dimensional smooth submanifold of $C \times T^*M$. Note that $\mathrm{Graph}(\kappa)$ is a fiber bundle over C with projection

$$p = p_{(H,C)} : \mathrm{Graph}(\kappa) \longrightarrow C\,, \quad (y, v) \mapsto y\,.$$

The fiber over y of this bundle is precisely K_y. Since $y \in K_y$, the map

$$\sigma : C \longrightarrow \mathrm{Graph}(\kappa)\,, \quad y \mapsto (y, y)\,,$$

is a section of $p_{(H,C)}$. Since the fibers of $p_{(H,C)}$ are affine spaces, the bundle $\mathrm{Graph}(\kappa) \to C$ has a natural vector bundle structure whose zero section is σ. Moreover, this vector bundle is canonically isomorphic to the normal bundle of C in T^*M. This is due to the fact that, for any $y \in C$, the fiber K_y is transversal to C at y. The same argument also guarantees that the map

$$\Xi = \Xi_{(H,C)} : \mathrm{Graph}(\kappa) \longrightarrow T^*M\,, \quad (y, v) \mapsto v\,,$$

induces a diffeomorphism of a neighborhood of the “zero” section $\sigma(C)$ in $\mathrm{Graph}(\kappa)$ onto its image.

Definition 4.2 *Let (M, H, C) be a nonholonomic Hamiltonian system. Given a point $u \in T^*M$, its (H, C)-*focusing locus *$F_{(H,C)}(u)$ is the set of all points $y \in C$ such that $u \in K_y$. In other words,*

$$F_{(H,C)}(u) = p_{(H,C)}\left(\Xi^{-1}_{(H,C)}(u)\right) \subset C_{\pi(u)}\,.$$

A point in $F_{(H,C)}(u)$ is called focusing *for u.*

Standard arguments show that $\Xi_{(H,C)}$ is regular, i.e., of maximal rank $2n$ almost everywhere, that is, with the exception of a closed subset without interior points. Therefore, for a generic point $u \in T^*M$, the subset $\Xi^{-1}_{(H,C)}(u)$ is discrete, and so is $F_{(H,C)}(u)$ as well. Note also that if $u \in C$, then $u \in F_{(H,C)}(u)$.

Remark 4.3 Focusing points can be understood as *nonintegrable* analogs of the notion of *reducing points* considered in [35] in connection with the Transition Principle for inelastic collisions.

Remark 4.4 It is worth noticing that the concept of a focusing point makes also sense in the absence of constraints. Obviously, in this case $F_{(H,C)}(u) = \{u\}$. Therefore there is no need to distinguish between the constrained and non-constrained cases in the statement of the Transition Principle.

If the constraints are linear, i.e., $C = \Upsilon + C^0$ with C^0 a linear codistribution and $\Upsilon \in \Lambda^1(M)$ (the displacement form), then for each $y \in T^*M$,

$$T^*_{\pi(y)}M = C^o_{\pi(y)} \oplus \mathrm{Ann}(d_y\pi(\Delta_y))\,,$$

where $\Delta = \Delta_{(H,C)}$. Denote the corresponding projectors by $\mathcal{P} : T^*_{\pi(y)}M \to C^o_{\pi(y)}$ and $\mathcal{Q} : T^*_{\pi(y)}M \to \mathrm{Ann}(d_y\pi(\Delta_y))$. Given $u \in T^*M$, one has that $z \in F_{(H,C)}(u)$ if and only if $z \in C$ and $\mathcal{P}(z) = \mathcal{P}(u)$. Since $z = \mathcal{P}(z) + \mathcal{Q}(z) = \mathcal{P}(u) + \mathcal{Q}(\Upsilon_y)$, one has the following result.

Proposition 4.5 *Let (M, H, C) be a nonholonomic Hamiltonian system with linear constraints. Then, for $u \in T^*_yM$, there is a unique focusing point given by $F_{(H,C)}(u) = \{\mathcal{P}(u) + \mathcal{Q}(\Upsilon_y)\}$.*

4.2 Instantaneous focusing points

We will also need an instantaneous version of the notion of a focusing point introduced in the previous section. For this purpose, it is sufficient to apply the above construction to instantaneous constraints instead of to the "usual" ones. Namely, let C^{inst} be a system of regular instantaneous constraints along N (see Section 2.1.3) and $\Delta^{\mathrm{inst}} = \Delta_{\Theta^{\mathrm{inst}}}$ be the corresponding affine distribution (see Section 3.2). Following the same reasoning as above, the affine subspace $W = \Delta^{\mathrm{inst}}_y \subset T_y(T^*M)$ satisfies the assumptions of Lemma 4.1. Therefore, the affine subspace $K^{\mathrm{inst}}_y = (\Delta^{\mathrm{inst}})^\bullet$ of $T_y(T^*M)$ is well-defined, and we have all the ingredients to define the notion of instantaneous crown and instantaneous focusing point of a system subject to instantaneous nonholonomic constraints. For completeness, we state the latter.

Definition 4.6 *Let (M, H, C) be a nonholonomic Hamiltonian system and let C^{inst} be a set of instantaneous constraints along a hypersurface $N \subset M$. Given a point $u \in T^*_NM$, its (H, C^{inst}, N)-*instantaneous focusing locus *$F_{(H,C^{inst},N)}(u)$ is the set of all points $y \in C^{inst}$ such that $u \in K^{inst}_y$. In other words,*

$$F_{(H,C^{inst},N)}(u) = p_{(H,C^{inst})}\left(\Xi^{-1}_{(H,C^{inst},N)}(u)\right) \subset C^{inst}_{\pi(u)}\,.$$

A point in $F_{(H,C^{inst},N)}(u)$ is called instantaneous focusing *for u.*

As before, if the instantaneous nonholonomic constraints are linear $C^{\text{inst}} = \Upsilon^{\text{inst}} + C^{\text{inst}^o}$, then for each $y \in T^*_N M$,

$$T^*_{\pi(y)}M = C^o_{\pi(y)} \oplus \operatorname{Ann}(d_y\pi(\Delta^{\text{inst}}_y))\,,$$

where $\Delta^{\text{inst}} = \Delta_{(H,C^{\text{inst}},N)}$. Denoting the corresponding projectors by $\mathcal{P}^{\text{inst}} : T^*_{\pi(y)}M \to {C^{\text{inst}^o}}_{\pi(y)}$ and $\mathcal{Q}^{\text{inst}} : T^*_{\pi(y)}M \to \operatorname{Ann}(d_y\pi(\Delta^{\text{inst}}_y))$, one has the following result.

Proposition 4.7 *Let (M, H, C) be a nonholonomic Hamiltonian system and let C^{inst} be a set of instantaneous affine constraints along a hypersurface $N \subset M$. Then, for $u \in T^*_y M$, there is a unique instantaneous focusing point given by $F_{(H,C^{inst},N)}(u) = \{\mathcal{P}^{inst}(u) + \mathcal{Q}^{inst}(\Upsilon^{inst}_y)\}$.*

4.3 Constrained characteristics

Consider then a partial symplectic structure $\Theta = (\Delta, \omega)$ on a manifold C which is transversal to a hypersurface $B \subset C$ (cf. Definition 2.5). Let Δ^0 denote the linear distribution associated with Δ. For each $y \in B$, consider the linear space $V = \Delta^0_y$, the hyperplane $W = \Delta^0_y \cap T_yB$ of V and the nondegenerate skew-symmetric form $b = \omega_y|_{\Delta^0_y}$. The *characteristic direction at $y \in B$* is defined as

$$l_y = l_y(\Theta, B) = \ker b|_W = \ker(\omega_y|_{\Delta^0_y \cap T_yB}) \subset \Delta^0_y \cap T_yB\,.$$

The proof of the following result is straightforward.

Lemma 4.8 *Given a partial symplectic structure $\Theta = (\Delta, \omega)$ on a manifold C and a hypersurface $B \subset C$ transversal to it, the distribution $y \mapsto l_y(\Theta, B)$ is one-dimensional.*

Definition 4.9 *Given a partial symplectic structure $\Theta = (\Delta, \omega)$ on C and a hypersurface $B \subset C$ transversal to it, $y \mapsto l_y(\Theta, B)$ is called the* characteristic distribution with respect to (Θ, B)*, and its integral curves, denoted by ζ, are the* (Θ, B)-characteristics.

We are particularly interested in the case when we have a nonholonomic Hamiltonian system $S = (M, H, C)$, the partial symplectic structure Θ is $\Theta_{H,C}$, N is a hypersurface in M and $\tilde{B} = T^*_N M$, $B = T^*_N M \cap C$. We will use the terminology $(S|N)$- or $(H, C|N)$-*characteristic* as a substitute for (Θ, B)-characteristic. It should be emphasized that $(H, C|N)$-characteristics are only defined when S is transversal to N (see Section 3.1).

In the absence of constraints, i.e., when $(C = T^*M, \omega = \Omega)$ is a symplectic manifold, and Δ is the trivial distribution $y \mapsto T_yC$ on C, the characteristic curves are precisely the characteristics introduced in [6]. We will refer to *non-constrained characteristics* and *constrained characteristics* when it is necessary to distinguish between the unconstrained and the constrained cases.

Remark 4.10 Just as non-constrained characteristics play a key role in describing holonomic elastic collisions, and reflection and refraction phenomena of rays of light [6, 34], the constrained characteristics will be fundamental in describing the "elastic part" of nonholonomic impulsive phenomena. What is meant by "elastic part" will become clear in Section 4.5 when describing decisive points.

If the constraints C are affine, then the $(H, C|N)$-characteristic passing through a point $y \in C$, $\pi(y) \in N$, is described in a particularly simple way. Namely, following Proposition 3.3, it is not difficult to see that the $(H, C|N)$-characteristic passing through y is given by

$$\zeta_y = y + C^0_{\pi(y)} \cap \operatorname{Ann}(d_y\alpha_H(C^0) \cap T_{\pi(y)}N)\,,$$

with C^0 being the linear codistribution associated to C. In particular, in the absence of constraints, $C = T^*M$ and the characteristics are straight lines in T^*_xM parallel to $\operatorname{Ann}(T_xN)$, $x \in N$.

4.4 In, out and trapping points

Here, we first introduce some concepts concerning the behavior of a vector field in a neighborhood of the boundary of its supporting manifold. We then discuss the notions in, out and trapping points.

Let Q be a manifold with boundary and X a vector field on Q. A point $y \in \partial Q$ is called a *jth order in point for* X if there exists a trajectory of X, $\beta : [0, a] \to Q$, $a > 0$ such that

$$y = \beta(0) \quad \text{and} \quad \beta(t) \notin \partial Q\,, \quad \text{for } 0 < t \leq a\,,$$

and β is jth order tangent to ∂Q at y. A *jth order out point for* X is a jth order in point for $-X$. In the dynamical context we have in mind, in and out points of 0th order are the most important. It is easy to see that $y \in \partial Q$ is a 0th order in point (resp., out point) for X if the vector X_y is transversal to ∂Q and directed inside (resp., outside) of Q. A point that lies on a trajectory of X which is entirely contained in ∂Q is called a *trapping point for* X.

Let $\partial Q^j = \partial Q^j(X)$ denote the subset of all points of ∂Q where X is jth order tangent to ∂Q, and $\partial Q^j_> = \partial Q^j_>(X)$ (resp., $\partial Q^j_< = \partial Q^j_<(X)$) the set of all jth order in points (resp., out points) for X. Note that $\partial Q^j \supset \partial Q^{j+1}$ and

$$\partial Q^j \setminus (\partial Q^j_> \cup \partial Q^j_<) \subset \partial Q^{j+1}. \tag{6}$$

In a generic situation, ∂Q^j is a submanifold (with singularities) of codimension j in ∂Q, which is divided by ∂Q^{j+1} into two parts, $\partial Q^j_> \setminus \partial Q^{j+1}_>$ and $\partial Q^j_< \setminus \partial Q^{j+1}_<$. An analytical description of the previous discussion is obtained by choosing a smooth function f on Q with $f \geq 0$ and $d_z f \neq 0$, for all $z \in \partial Q$ such that $\partial Q = \{f = 0\}$ (which always exists locally). Then

$$\begin{aligned}
\partial Q^j &= \{z \in Q \mid f(z) = 0,\ X(f)(z) = 0, \ldots, X^j(f)(z) = 0\}\,,\\
\partial Q^j_> \setminus \partial Q^{j+1}_> &= \partial Q^j \cap \{z \in Q \mid X^{j+1}(f)(z) > 0\}\,,\\
\partial Q^j_< \setminus \partial Q^{j+1}_< &= \partial Q^j \cap \{z \in Q \mid X^{j+1}(f)(z) < 0\}\,.
\end{aligned}$$

The vector field X is said to be *regular with respect to* ∂Q when the inclusion in equation (6) is an equality for all $j \geq 0$. This is a generic property of vector fields. In such a case, the chain of inclusions

$$\partial Q = \partial Q^0 \supset \partial Q^1 \supset \cdots \supset \partial Q^j \supset \cdots \supset \partial Q^n$$

is a *stratification* of ∂Q whose strata are $\partial Q^j_> \setminus \partial Q^{j+1}_>$ and $\partial Q^j_< \setminus \partial Q^{j+1}_<$. Note also that the set of trapping points precisely corresponds to ∂Q^n.

Consider now a discontinuous nonholonomic Hamiltonian system $(M, H, C \mid N)$ and the corresponding cutting-up $(\hat{M}, \varsigma)$. Then we can resort to the previous discussion with the manifold $Q = \hat{C} \subset T^*\hat{M}$ and the vector field $X = X_{\hat{H},\hat{C}}$. Recall that $\hat{N} \subset \partial\hat{M}$ and $\partial\hat{C} = \hat{C} \cap T^*_{\partial\hat{M}}\hat{M}$.

Definition 4.11 *Let $S = (M, H, C \mid N)$ be a discontinuous nonholonomic system and denote by $(\hat{M}, \varsigma)$ the associated cutting-up. A point $y \in T^*_N M$ is called an* in point *(resp., an* out point*) of S if there exists $z \in T_{\hat{N}}\hat{M}$ such that $y = \varsigma(z)$ and z is an in point (resp., an out point) of $X_{\hat{H},\hat{C}}$ with respect to $\partial\hat{C}$.*

By definition, the map $T^*\varsigma$ restricted to $\partial\hat{C}$ is an immersion. A point in $T^*_N M$ may turn out to be an in and an out point at the same time. To resolve this ambiguity, the branch of $T^*\varsigma$ to which such a point belongs must be taken into consideration. This distinction is easily described in the case when N divides M into two parts. In fact, in this case the system $(M, H, C \mid N)$ may be viewed as a couple of nonholonomic Hamiltonian systems $(M_\pm, H_\pm, C_\pm)$, with the common boundary $\partial M_\pm = N$, and where $H_\pm \in C^\infty(M_\pm)$ and $C_\pm \subset M_\pm$ (cf. Section 3.3). An in (resp., out, or trapping) point of the vector field X_{H_+,C_+} with respect to the boundary ∂C_+ is called an *plus-in* (resp., *plus-out*, or *plus-trapping) point*. Analogous definitions are established for $\varepsilon = -$. In this way, the notions of plus-in point, minus-in point, etc, introduced in [6] for the unconstrained situation are generalized to the constrained case. Finally, we observe that N always divides M locally, and therefore the previous discussion is always valid locally.

4.5 Decisive points

At this point, we are ready to introduce the key notion of decisive point corresponding to an out point. The construction of decisive points depends on two elements: first, the mode (elastic or inelastic) in which the system passes through the critical state and, second, the continuity and differentiability properties of the Hamiltonian. Below, we will limit our discussion to the two most relevant situations, just to avoid not very instructive technicalities arising in the most general context. The first one is the case when the Hamiltonian is smooth and only the constraints are discontinuous along the critical hypersurface. The second one concerns discontinuous Hamiltonians and not necessarily discontinuous constraints. It is worth stressing that the first situation can not be considered as a particular case of the second one, i.e., that the notion of a decisive point is not "continuous" in this sense. In what follows, $\varepsilon \in \{+,-\}$ and $\bar\varepsilon$ stands for the opposite sign to ε. Throughout the section, instantaneous constraints are assumed to be regular.

4.5.1 Elastic mode: change of constraints

Here, we deal with a discontinuous nonholonomic system $(M, H, C|N)$, where the Hamiltonian function is smooth, $H \in C^\infty(M)$.

Definition 4.12 (Decisive points for smooth Hamiltonians and discontinuous constraints) *Let $(M, H, C|N)$ be a regular discontinuous nonholonomic system, with $H \in C^\infty(M)$ and consider a set of instantaneous constraints C^{inst} along N. Let y be an ε-out point of the system. A sequence (y_i, ε_i), $i = 0, 1, \ldots, k$, with $y_i \in C \cap T^*_N M$ is called* (y,ε)-admissible *if it verifies the following conditions:*

(i) $(y_0, \varepsilon_0) = (y, \varepsilon)$;

(ii) *for all $i < k$, y_{i+1} is a focusing point for y_i with respect to either $C^{inst}_{\varepsilon_{i+1}}$ or, if instantaneous constraints are absent, $C_{\varepsilon_{i+1}}$;*

(iii) *y_i is an ε_i-out point for all $i < k$ and y_k is either an ε_k-in point or an ε_k-trapping point;*

(iv) *the sequence of signs $\{\varepsilon_i\}$ alternates, i.e., $\varepsilon_{i+1} = \bar\varepsilon_i$.*

The end *point of an (y,ε)-admissible sequence, (y_k, ε_k), is called* (y,ε)-decisive *and the constrained Hamiltonian vector field $X_{H,C_{\varepsilon_k}}$ is referred to as the vector field* corresponding *to it.*

Remark 4.13 The above formal description of decisive points is equivalent to the following iterative procedure. Take, for instance, a plus-out point y. Then, according to Definition 4.12, all focusing with respect to $C^{\rm inst}_{\bar\varepsilon}$ (resp., to $C_{\bar\varepsilon}$) minus-in and minus-trapping points are decisive. On the other hand, the procedure continues by restarting from any of the remaining focusing points that are minus-out points, and so on. In some situations, this process may turn out to be infinite. At the present time, however, it is not clear whether that kind of phenomena can occur, say in propagation of singularities or similar processes.

4.5.2 Elastic mode: discontinuous Hamiltonians

In this case, decisive points are constructed on the basis of an iterative procedure whose single steps are either of *reflective* or of *refractive* type, as described below. Consider a regular discontinuous nonholonomic system $S = (M, H, C|N)$, which might be subject to additional instantaneous constraints C^{inst} along N. Let $y \in C \cap T^*_N M$ be an ε-out point.

Reflective step

1-st move: $y \Rightarrow z$, where z is a point in the constrained characteristic $\zeta_y(H_\varepsilon, C_\varepsilon)$ such that $H_\varepsilon(z) = H_\varepsilon(y)$.

2-nd move: $z \Rightarrow u$, where u is a focusing point for z with respect to either $C^{\text{inst}}_\varepsilon$ or, if ε-instantaneous constraints are absent, C_ε.

Refractive step

1-st move: $y \Rightarrow z$, where z is a point of the constrained characteristic $\zeta_y(H_\varepsilon, C_\varepsilon)$ and such that $H_{\bar\varepsilon}(z) = H_\varepsilon(y)$.

2-nd move: $z \Rightarrow u$, where u is a focusing point for z with respect to either $C^{\text{inst}}_{\bar\varepsilon}$ or, if $\bar\varepsilon$-instantaneous constraints are absent, $C_{\bar\varepsilon}$.

With a slight abuse of language, we shall say that (y, ε) is the *initial point of the step* and (u, ε) (resp., $(u, \bar\varepsilon)$) is the *end point of the step* if the scenario is reflective (resp., refractive).

Definition 4.14 (Decisive points for discontinuous Hamiltonians) *Consider a regular discontinuous nonholonomic system $(M, H, C|N)$. Let C^{inst} be a set of instantaneous constraints along N. Let y be an ε-out point. A sequence (y_i, ε_i), $i = 0, 1, \ldots, k$, is called* (y, ε)-admissible *if*

(i) *$(y_0, \varepsilon_0) = (y, \varepsilon)$;*

(ii) *(y_i, ε_i) and $(y_{i+1}, \varepsilon_{i+1})$ are the initial and the end points of a step, respectively;*

(iii) *y_i is an ε_i-out point, $0 \leq i < k$, and y_k is an ε_k-in point or an ε_k-trapping point.*

The end point (y_k, ε_k) of an admissible sequence is called (y, ε)-decisive and the constrained Hamiltonian vector field $X_{H, C_{\varepsilon_k}}$ is referred to as the vector field corresponding *to it.*

If the Hamiltonian is discontinuous and the constraints are linear, i.e., $C \subset T^*M$ is a smooth linear submanifold, and the instantaneous constraints are absent, the previous definition of decisive points becomes much simpler, as the following result shows.

Proposition 4.15 *Let $(M, H, C|N)$ be a regular discontinuous nonholonomic Hamiltonian system with smooth linear constraints. Let y be an ε-out point. The* (y, ε)-decisive *points are the in and the trapping points belonging to the intersection of the constrained characteristic ζ_y passing through y with the set $\{z \in C \mid H_\pm(z) = H_\varepsilon(y)\}$.*

Proof: Let y be an ε-out point and denote by $\{z_1, \ldots, z_s\}$ (resp, $\{\overline{z}_1, \ldots, \overline{z}_{\overline{s}}\}$) the points belonging to the intersection of the constrained characteristic ζ_y passing through y with the set $\{z \in C \mid H_\varepsilon(z) = H_\varepsilon(y)\}$ (resp. with $\{z \in C \mid H_{\overline{\varepsilon}}(z) = H_\varepsilon(y)\}$). Since the constraints are smooth, then $u = z$ in the 2nd-move of both a reflective and a refractive step. Now, for any $j \in \{1, \ldots, s\}$, the intersection of the constrained characteristic passing ζ_{z_j} through z_j with the set $\{z \in C \mid H_\varepsilon(z) = H_\varepsilon(z_j)\}$ (resp. with $\{z \in C \mid H_{\overline{\varepsilon}}(z) = H_\varepsilon(y)\}$) is again $\{z_1, \ldots, z_s\}$ (resp, $\{\overline{z}_1, \ldots, \overline{z}_{\overline{s}}\}$). The same observation holds for any $\overline{z}_j$, $j \in \{1, \ldots, \overline{s}\}$. The result now follows from Definition 4.14. ∎

Remark 4.16 The introduced terminology remains valid for nonholonomic systems with boundary (cf. Remark 3.7). In such a case, one has to formally put

$$M_- = \emptyset\,, \quad M_+ = M\,, \quad N = \partial M\,, \quad H_- = \infty\,, \quad H_+ = H\,.$$

This type of geometric data occurs in describing various collision phenomena.

4.5.3 Inelastic mode: change of constraints

As in the elastic case, we first deal with the case when the Hamiltonian H is smooth. We treat an inelastic behavior of the system as the passage under the control of either the instantaneous discontinuous nonholonomic system or, if instantaneous constraints are absent, the discontinuous boundary system. In this and subsequent sections, the following shorthand notation will be used (cf. Sections 3.2 and 3.3)

$$\begin{array}{rclrcl} C_\varepsilon^{\mathrm{inst,tr}} & = & C_\varepsilon^{\mathrm{inst}} \cap \alpha_{H_\varepsilon}^{-1}(TN)\,, & X_\varepsilon^{\mathrm{inst,tr}} & = & X_{(H_\varepsilon, C_\varepsilon^{\mathrm{inst,tr}}, N)}\,, \\ C_\varepsilon^{\mathrm{tr}} & = & C_{\varepsilon(N,H_\varepsilon)}\,, & X_\varepsilon^{\mathrm{tr}} & = & X^{\mathrm{tr}}_{(H_\varepsilon, C_\varepsilon, N)}\,. \end{array}$$

We also use this notation when the Hamiltonian H is smooth, i.e., $H_\pm = H$.

Definition 4.17 (Decisive points for smooth Hamiltonians and discontinuous constraints) *Consider a regular discontinuous nonholonomic system $(M, H, C|N)$. Let C^{inst} be a set of instantaneous constraints along N. Let y be an ε-out point. An (y, ε)-*decisive point *is a focusing point for y with respect to either $C_{\overline{\varepsilon}}^{inst,tr}$ or, if the instantaneous constraints are absent, $C_{\overline{\varepsilon}}^{tr}$. The constrained Hamiltonian vector field $X_{\overline{\varepsilon}}^{inst,tr}$, respectively, $X_{\overline{\varepsilon}}^{tr}$ is referred to as the* corresponding *vector field.*

4.5.4 Inelastic mode: discontinuous Hamiltonians

As in the elastic case, decisive points are constructed on the basis of *reflective* or *refractive* steps, as we now describe.

Reflected falling step

1-st move: $y \Rightarrow z$, where z is a point of the constrained characteristic $\zeta_y(H_\varepsilon, C_\varepsilon)$ such that $H_\varepsilon(z) = H_\varepsilon(y)$.

2-nd move: $z \Rightarrow u$, where u is a focusing point for z with respect to either $C_\varepsilon^{\mathrm{inst,tr}}$ or, if ε-instantaneous constraints are absent, $C_\varepsilon^{\mathrm{tr}}$.

Refracted falling step

1-st move: $y \Rightarrow z$, where z is a point of the constrained characteristic $\zeta_y(H_\varepsilon, C_\varepsilon)$ such that $H_{\bar{\varepsilon}}(z) = H_\varepsilon(y)$.

2-nd move: $z \Rightarrow u$, where u is a focusing point for z with respect to $C_{\bar{\varepsilon}}^{\mathrm{inst,tr}}$ or, if $\bar{\varepsilon}$-instantaneous constraints are absent, $C_{\bar{\varepsilon}}^{\mathrm{tr}}$.

We shall refer to (u, ε) (resp., $(u, \bar{\varepsilon})$) as a *reflected* (resp. *refracted*) *falling point.*

Definition 4.18 (Decisive points for discontinuous Hamiltonians) *Consider a regular discontinuous nonholonomic system $(M, H, C|N)$. Let C^{inst} be a set of instantaneous constraints along N. Let y be an ε-out point. An (y, ε)*-decisive *point is a falling point for y. The vector field $X_\varepsilon^{inst,tr}$ (resp., X_ε^{tr} if $C_\varepsilon^{inst} = \emptyset$) is called the vector field* corresponding *to a reflected falling point. The vector field $X_{\bar{\varepsilon}}^{inst,tr}$ (resp., $X_{\bar{\varepsilon}}^{tr}$ if $C_{\bar{\varepsilon}}^{inst} = \emptyset$) is called the vector field* corresponding *to a refracted falling point.*

4.6 Transition Principle

From a physical point of view, the Transition Principle formulated below is an explicit description of the discontinuity of a trajectory of a regular nonholonomic Hamiltonian system S that occurs when it traverses a *critical state*. Such a discontinuity is interpreted as an impact, collision, reflection, refraction, etc, depending on the physical situation modeled by the system S. From a mathematical point of view, the Transition Principle corresponds to the definition of the trajectory of a regular discontinuous nonholonomic Hamiltonian system.

The elastic or inelastic character of the impulsive motions of an specific physical system must be taken into account when defining the trajectories. Accordingly, there are two different versions of the Transition Principle that distinguish between the two situations. Let $S = (M, H, C|N)$ stand for a regular discontinuous nonholonomic system and let C^{inst} be eventual instantaneous constraints imposed on S along N. Let $(\hat{M}, \varsigma)$ be the associated cutting-up of M along N (cf. Section 3.3). The *regular part of a trajectory of the system* $\hat{S} = (\hat{M}, \hat{H}, \hat{C})$ is the part of the trajectory of the Hamiltonian vector field $X_{\hat{H},\hat{C}}$ that lies outside $\partial\hat{M}$. The *regular part of a trajectory of* S is the image by ς of the regular part of the corresponding trajectory of $\hat{S}$. At least locally, the regular part may be viewed as a piece of the trajectory of the vector field $X_{H_\varepsilon, C_\varepsilon}$ that lies outside the hypersurface T_N^*M.

Transition Principle. *Let $S = (M, H, C|N)$ be a regular discontinuous nonholonomic system and let C^{inst} be eventual instantaneous constraints on S along N. If a regular trajectory of the vector field $X_{H_\varepsilon, C_\varepsilon}$, $\varepsilon = \pm$ reaches the critical hypersurface $T_N M$ at a point y, it then continues its motion from any (y, ε)-decisive point according to the chosen mode, elastic or inelastic, under the control of the corresponding constrained Hamiltonian vector field.*

Some features of the Transition Principle are worth mentioning. First of all, it prescribes a *splitting of the trajectory* when the number of decisive points is greater than one. Of course, it is difficult to imagine that a true mechanical system "goes into pieces" when reaching the critical hypersurface. But it may perfectly happen when a Hamiltonian system describes the propagation of singularities in a fields or a continuum media. A classical example one finds in geometrical optics when a light ray passing from one optic medium to another splits into reflected and refracted rays (see, for instance, [34]). The trajectory may also be *trapped* by the critical hypersurface. This happens when an "impact" state y possesses no y-decisive points.

5 Mechanical systems

In this section, we particularize the previous discussion to mechanical systems subject to affine constraints. Let g be a Riemannian metric on M and $V \in C^\infty(M)$, and consider the mechanical system whose kinetic energy and potential function are $T(q,v) = \frac{1}{2}g(v,v)$ and V, respectively. The corresponding Lagrangian function is $L(q,v) = T(q,v) - V(q)$ and the Hamiltonian one is

$$H(q,p) = \hat{T}(q,p) + V(q)\,, \tag{7}$$

where $\hat{T}(q,p) = \frac{1}{2}\mathcal{G}(p,p)$, and $\mathcal{G}$ is the co-metric, i.e., the metric on the cotangent bundle induced by g. In a local chart q^a on M, the local expressions of g and $\mathcal{G}$ are

$$g = g_{ab}dq^a \otimes dq^b\,, \quad \mathcal{G} = g^{ab}\frac{\partial}{\partial q^a} \otimes \frac{\partial}{\partial q^b}\,.$$

In the mechanical case, the Legendre transform $\mathcal{L}_L : TM \longrightarrow T^*M$ is a linear bundle mapping whose local description is $\mathcal{L}_L(q^a, \dot{q}^a) = (q^a, g_{ab}\dot{q}^b)$.

Consider an affine distribution $C = C^0 + Y$ in T^*M determining some nonholonomic constraints on the system (M,H). The linearity of $\alpha_H = \mathcal{L}_L^{-1}$ implies that the space $\alpha_H(C) = \alpha_H(C^0) + \alpha_H(Y)$ is a distribution of affine spaces on M, or otherwise said, that $\alpha_H(C^0)$ is a linear distribution on M. Throughout this section, we will often resort to the shorthand notation $\mathcal{D} = \alpha_H(C^0)$ and $\Upsilon = \alpha_H(Y)$. Now, it is easy to verify that $T_q^*M = C_q^0 \oplus \mathrm{Ann}(\mathcal{D})_q$, with associated projectors

$$\mathcal{P}_q : T_q^*M \longrightarrow C_q^0\,, \quad \mathcal{Q}_q : T_q^*M \longrightarrow \mathrm{Ann}(\mathcal{D})_q\,, \quad q \in M\,.$$

Let $\mu_1 = \mu_{1a}dq^a, \dots, \mu_m = \mu_{ma}dq^a$ be 1-forms such that (locally) $\mathrm{Ann}(\mathcal{D}) = \mathrm{span}\{\mu_1, \dots, \mu_m\}$. Define the local function $\mu_{i0} : M \to \mathbb{R}$ by $\mu_{i0}(q) = -\mu_i(\Upsilon(q))$. Then $\alpha_H(C)$ is locally defined by the equations

$$\mu_{ia}(q)\dot{q}^a + \mu_{i0}(q) = 0\,, \quad 1 \le i \le m\,.$$

Now, consider the matrices

$$G = (g_{ab})\,, \quad J = (\mu_{ia})\,, \quad \mathcal{B} = JG^{-1}J^t\,. \tag{8}$$

From the discussion after Proposition 3.1, recall that $(\Delta_{H,C}, \omega_C)$ is a partial symplectic structure if and only if $\alpha_H|_C$ is an immersion, or, equivalently, if the compatibility condition is verified. Following [12], the latter is equivalent to the matrix $\mathcal{B}$ being invertible. A direct computation give the following local expression for the projectors $\mathcal{P}$ and $\mathcal{Q}$,

$$\mathcal{P}(x) = x - \mathcal{Q}(x)\,, \quad \mathcal{Q}(x) = J^t\mathcal{B}^{-1}JG^{-1}x\,, \quad x \in T^*M\,.$$

Finally, let $N \subset M$ be a hypersurface and assume that the nonholonomic system $S = (M,H,C)$ is transversal to N. Consider also a set of instantaneous nonholonomic linear constraints $C^{\mathrm{inst}} = ({C^{\mathrm{inst}}}^o, \Upsilon_{C^{\mathrm{inst}}})$ imposed on S along N. Note that $T_q^*M = {C_q^{\mathrm{inst}}}^o \oplus \mathrm{Ann}(\mathcal{D}^{\mathrm{inst}})_q$, with associated projectors

$$\mathcal{P}_q^{\mathrm{inst}} : T_q^*M \longrightarrow {C_q^{\mathrm{inst}}}^o\,, \quad \mathcal{Q}_q^{\mathrm{inst}} : T_q^*M \longrightarrow \mathrm{Ann}(\mathcal{D}^{\mathrm{inst}})_q\,, \quad q \in N\,.$$

5.1 Focusing points

Since the mechanical system is subject to an affine distribution of constraints, Proposition 4.5 implies that for a given $u \in T^*M$, the focusing locus is $F_{(H,C)}(u) = \{\mathcal{P}(u) + \mathcal{Q}(\Upsilon)\}$. Regarding the instantaneous focusing points, according to Proposition 4.7 one has that $F_{(H,C^{\mathrm{inst}},N)}(u) = \{\mathcal{P}^{\mathrm{inst}}(u) + \mathcal{Q}^{\mathrm{inst}}(\Upsilon_{C^{\mathrm{inst}}})\}$.

5.2 Constrained characteristics

Here we give an explicit description of the characteristic curves. Let N be the critical hypersurface, and assume that (locally) $N = f^{-1}(0)$, with $f \in C^\infty(M)$ verifying that $d_q f \neq 0$ for all $q \in N$. Consider the covector field $\mathcal{P}(df)$ along N defined as $q \mapsto \mathcal{P}(df)_q = \mathcal{P}_q(d_q f)$, $q \in N$. The transversality assumption between C and N implies that $\mathcal{P}(df)_q \neq 0$, for all $q \in N$. Clearly $\mathcal{P}(df) \in C^0$. In addition, for $v \in \mathcal{D} \cap TN$,

$$\mathcal{P}(df)(v) = (df - \mathcal{Q}(df))(v) = df(v) = 0\,,$$

and one can conclude that $C^0 \cap \mathrm{Ann}(\mathcal{D} \cap TN) = \mathrm{span}\{\mathcal{P}(df)\}$. Therefore, we have the following result.

Lemma 5.1 *The constrained characteristic of a mechanical system $(M, H, C|N)$ passing through $y \in C \cap T_N^*M$ is given by $\zeta_y = y + \mathrm{span}\{\mathcal{P}(d_{\pi(y)} f)\} \subset C \cap T_N^*M$.*

Note that in the absence of constraints one recovers the standard non-constrained characteristic $\zeta_y = y + \mathrm{span}\{d_{\pi(y)} f\}$ passing through y.

5.3 Decisive points: elastic mode

5.3.1 Change of constraints

Let $C_\pm \subset T^*M$ be two affine constraint submanifolds. Denote by $\mathcal{P}_\pm$ and $\mathcal{Q}_\pm$ the projectors corresponding to $C_\pm$ and the co-metric $\mathcal{G}$. Let $y \in C_\varepsilon \cap T_N^*M$ be a ε-out point, $\varepsilon \in \{+,-\}$. Then, according to Definition 4.12, an y-admissible sequence, (y_i, ε_i), $i = 0, 1, \ldots, k$, is necessarily of the form $y_{i+1} = \mathcal{P}_{\varepsilon_{i+1}}(y_i) + \mathcal{Q}_{\varepsilon_{i+1}}(\Upsilon_{C_{\varepsilon_{i+1}}})$. If instantaneous constraints are present, then one has to use the projectors $\mathcal{P}_\varepsilon^{\mathrm{inst}}$ and $\mathcal{Q}_\varepsilon^{\mathrm{inst}}$ instead of $\mathcal{P}_\varepsilon$ and $\mathcal{Q}_\varepsilon$, respectively.

Remark 5.2 Mechanical systems subject to generalized constraints are also treated in [16] in a somehow different context. The approach taken there makes use of generalized (i.e. non-constant rank) codistributions defining the nonholonomic constraints and a generalized version of Newton's second law. Under appropriate regularity conditions, it can be seen that the 'post-impact' point in [16] is a decisive point of the Hamiltonian system according to Definition 4.12.

5.3.2 Discontinuous Hamiltonian systems

Let $C_\pm \subset T^*M$ be two affine constraint submanifolds. Let $g_\pm$ be a Riemannian metric on $M_\pm$ and $V_\pm \in C^\infty(M_\pm)$ such that

$$H_\pm(q,p) = \hat{T}_\pm(q,p) + V_\pm(q)\,, \quad \hat{T}_\pm(q,p) = \frac{1}{2}\mathcal{G}_\pm(p,p)\,. \tag{9}$$

For simplicity, we only treat the case $V_\pm = V|_{M_\pm}$, $V \in C^\infty(M)$. We denote by $\mathcal{P}_\pm$ and $\mathcal{Q}_\pm$ the projectors corresponding to $C_\pm$ and the co-metric $\mathcal{G}_\pm$. Additionally, let $C_\pm^{\mathrm{inst}} \subset T_N^*M$ be affine constraint submanifolds corresponding to some instantaneous constraints imposed along N. Denote by $\mathcal{P}_\pm^{\mathrm{inst}}$ and $\mathcal{Q}_\pm^{\mathrm{inst}}$ the projectors corresponding to $C_\pm^{\mathrm{inst}}$ and the co-metric $\mathcal{G}_\pm$.

Let $y \in C_\varepsilon \cap T_N^*M$ be an ε-out point. Following Definition 4.14, we first describe the reflective and refractive steps with initial point (y, ε). According to Lemma 5.1, we have to look for points of the form

$$x = y + c\mathcal{P}_\varepsilon(d_q f)\,, \quad q = \pi(y)\,,$$

for some c, which in addition belong to the same H-energy level as y.

Reflective step Concerning the 1-st move, note that $y + c\mathcal{P}_\varepsilon(d_q f)$ and y belong to T_q^*M. Then, the equality $H_\varepsilon(y + c\mathcal{P}_\varepsilon(d_q f)) = H_\varepsilon(y)$ implies that $\hat{T}_\varepsilon(y + c\mathcal{P}_\varepsilon(d_q f)) = \hat{T}_\varepsilon(y)$. Now,

$$\hat{T}_\varepsilon(y + c\mathcal{P}_\varepsilon(d_q f)) = \hat{T}_\varepsilon(y) + c\,\mathcal{G}_\varepsilon(y, \mathcal{P}_\varepsilon(d_q f)) + \frac{c^2}{2}\mathcal{G}_\varepsilon(\mathcal{P}_\varepsilon(d_q f), \mathcal{P}_\varepsilon(d_q f))\,,$$

and, therefore, we have

$$c\left(\mathcal{G}_\varepsilon(y, \mathcal{P}_\varepsilon(d_q f)) + \frac{c}{2}\mathcal{G}_\varepsilon(\mathcal{P}_\varepsilon(d_q f), \mathcal{P}_\varepsilon(d_q f))\right) = 0\,,$$

with solutions

$$c_{\varepsilon,1} = 0\,, \quad c_{\varepsilon,2} = -\frac{2\,\mathcal{G}_\varepsilon(y, \mathcal{P}_\varepsilon(d_q f))}{\mathcal{G}_\varepsilon(\mathcal{P}_\varepsilon(d_q f), \mathcal{P}_\varepsilon(d_q f))}\,. \tag{10}$$

An important property of these points is contained in the following lemma.

Lemma 5.3 *Let $y \in C_\varepsilon \cap T_N^*M$ and $c_{\varepsilon,2}$ be the constant given by* (10). *Then,*

$$\mathcal{G}_\varepsilon(y, d_q f) = \mathcal{G}_\varepsilon(\mathcal{P}_\varepsilon(y) + \mathcal{Q}_\varepsilon(\Upsilon_q), d_q f)\,,$$
$$\mathcal{G}_\varepsilon(y + c_{\varepsilon,2}\mathcal{P}_\varepsilon(d_q f), d_q f) = \mathcal{G}_\varepsilon(-\mathcal{P}_\varepsilon(y) + \mathcal{Q}_\varepsilon(\Upsilon_q), d_q f)\,.$$

Proof: The first statement follows by noting that if $y \in C_q$, then $y = \mathcal{P}_\varepsilon(y) + \mathcal{Q}_\varepsilon(\Upsilon_q)$. For the second one, notice that

$$\begin{aligned}\mathcal{G}_\varepsilon(y + c_{\varepsilon,2}\mathcal{P}_\varepsilon(d_q f), d_q f) &= \mathcal{G}_\varepsilon(y, d_q f) + c_{\varepsilon,2}\mathcal{G}_\varepsilon(\mathcal{P}_\varepsilon(d_q f), \mathcal{P}_\varepsilon(d_q f)) \\ &= \mathcal{G}_\varepsilon(y, d_q f - 2\mathcal{P}_\varepsilon(d_q f)) = -\mathcal{G}_\varepsilon(\mathcal{P}_\varepsilon(y), \mathcal{P}_\varepsilon(d_q f)) + \mathcal{G}_\varepsilon(\mathcal{Q}_\varepsilon(y), d_q f) \\ &= \mathcal{G}_\varepsilon(-\mathcal{P}_\varepsilon(y) + \mathcal{Q}_\varepsilon(\Upsilon_q), d_q f)\,,\end{aligned}$$

which gives the desired result. ■

The 2-nd move simply consists of determining the focusing points for points (10) with respect to $C_\varepsilon^{\text{inst}}$ or, if ε-instantaneous constraints are absent, with respect to C_ε. This is done in terms of the corresponding projectors, exactly as explained in Section 5.1 above.

Refractive step Concerning the 1-st move, the equality $H_{\bar{\varepsilon}}(y + c\mathcal{P}_\varepsilon(d_q f)) = H_\varepsilon(y)$ implies $\hat{T}_{\bar{\varepsilon}}(y + c\mathcal{P}_\varepsilon(d_q f)) = \hat{T}_\varepsilon(y)$. Now,

$$\hat{T}_{\bar{\varepsilon}}(y + c\mathcal{P}_\varepsilon(d_q f)) = \hat{T}_{\bar{\varepsilon}}(y) + c\,\mathcal{G}_{\bar{\varepsilon}}(y, \mathcal{P}_\varepsilon(d_q f)) + \frac{c^2}{2}\mathcal{G}_{\bar{\varepsilon}}(\mathcal{P}_\varepsilon(d_q f), \mathcal{P}_\varepsilon(d_q f))\,.$$

Therefore, one has

$$c\left(\mathcal{G}_{\bar{\varepsilon}}(y, \mathcal{P}_\varepsilon(d_q f)) + \frac{c}{2}\mathcal{G}_{\bar{\varepsilon}}(\mathcal{P}_\varepsilon(d_q f), \mathcal{P}_\varepsilon(d_q f))\right) + \hat{T}_{\bar{\varepsilon}}(y) - \hat{T}_\varepsilon(y) = 0\,,$$

with solutions $i = 1, 2$,

$$c_{\bar{\varepsilon},i} = \frac{1}{\mathcal{G}_{\bar{\varepsilon}}(\mathcal{P}_\varepsilon(d_q f)), \mathcal{P}_\varepsilon(d_q f))}\Big(-\mathcal{G}_{\bar{\varepsilon}}(y, \mathcal{P}_\varepsilon(d_q f)) \pm \sqrt{\mathcal{G}_{\bar{\varepsilon}}(y, \mathcal{P}_\varepsilon(d_q f))^2 - 2\mathcal{G}_{\bar{\varepsilon}}(\mathcal{P}_\varepsilon(d_q f), \mathcal{P}_\varepsilon(d_q f))(\hat{T}_{\bar{\varepsilon}}(y) - \hat{T}_\varepsilon(y))}\Big)\,. \tag{11}$$

As before, the 2-nd move simply consists of computing the focusing points for the solutions (11) with regards to $C_{\bar{\varepsilon}}^{\text{inst}}$ or, if $\bar{\varepsilon}$-instantaneous constraints are absent, $C_{\bar{\varepsilon}}$. This is done in terms of the corresponding projectors according to Section 5.1.

Discontinuous Hamiltonian systems with smooth constraints

In this situation, there is a single constraint submanifold C, and a discontinuous Hamiltonian $H_\pm$ on T^*M. Denote by $\mathcal{P}_\pm$ and $\mathcal{Q}_\pm$ the projectors corresponding to C and the co-metrics $\mathcal{G}_\pm$, respectively. According to Proposition 4.15, the decisive points for a given ε-out point $y \in C \cap T^*_N M$ are simply the in and trapping points belonging to the intersection of the constrained characteristic ζ_y passing through y with the set $\{z \in C \mid H_\pm(z) = H_\varepsilon(y)\}$. Therefore, as candidate ε-decisive points we have the solution corresponding to $c_{\varepsilon,2}$ in (10), and as candidate $\bar{\varepsilon}$-decisive points we have the solutions corresponding to $c_{\bar{\varepsilon},i}$, $i = 1, 2$, in (11).

Proposition 5.4 *Let $y \in C_\varepsilon \cap T^*_N M$ be a ε-out point. If the constraints are linear, $C = C^0$, then the solution corresponding to $c_{\varepsilon,2}$ in* (10) *is a ε-decisive point for y.*

Proof: The basic observation is the second order character of the dynamics, both in the presence and in the absence of nonholonomic constraints. This implies that for any $y \in T^*M$ and any distribution of affine constraints C, we have $X_H(f)(y) = X_{H,C}(f)(y)$, since f is only a function of the configurations. Note that if H is of mechanical type, then $X_H(f)(x) = \mathcal{G}(x, d_q f)$, for any $x \in T^*_q M$. Now, from Lemma 5.3, taking $H = H_\varepsilon$, one gets

$$\mathcal{G}_\varepsilon(y + c_2\mathcal{P}_\varepsilon(d_q f), d_q f) = -\mathcal{G}_\varepsilon(y, d_q f)\,.$$

Since y is a ε-out point, then $\mathcal{G}_\varepsilon(y, d_q f) \neq 0$. Consequently, $X_{H_+}(f)(y + c_2\mathcal{P}_+(d_q f)) = -\mathcal{G}_+(d_q f, y)$ has the opposite sign, and hence it is an in point. ■

5.4 Decisive points: inelastic mode

5.4.1 Change of constraints

Let $C_\pm \subset T^*M$ be two affine constraint submanifolds and let C^{inst} be a set of instantaneous affine constraints. We denote by $\mathcal{P}^{\text{inst}}_\pm$ and $\mathcal{Q}^{\text{inst}}_\pm$ the projectors corresponding to $C^{\text{inst}}_\pm$ and the co-metric $\mathcal{G}$. If the instantaneous constraints are absent, denote by $\mathcal{P}_\pm$ and $\mathcal{Q}_\pm$ the projectors corresponding to $C^{\text{tr}}_\pm$ and the co-metric $\mathcal{G}$. Let $y \in C_\varepsilon \cap T^*_N M$ be a ε-out point, $\varepsilon \in \{+,-\}$. Then, according to Definition 4.17, the unique y-decisive point is $\mathcal{P}^{\text{inst}}_{\bar{\varepsilon}}(y) + \mathcal{Q}^{\text{inst}}_{\bar{\varepsilon}}(\Upsilon_{C_{\bar{\varepsilon}}})$ (or, if there are no instantaneous constraints, $\mathcal{P}_{\bar{\varepsilon}}(y) + \mathcal{Q}_{\bar{\varepsilon}}(\Upsilon_{C_{\bar{\varepsilon}}})$).

5.4.2 Discontinuous Hamiltonian systems

As in Section 5.3.2, let $C_\pm \subset T^*M$ be two affine constraint submanifolds, $g_\pm$ a Riemannian metric on $M_\pm$ and $V_\pm \in C^\infty(M_\pm)$ such that equation (9) is verified. For simplicity, we only treat the case $V_\pm = V|_{M_\pm}$, $V \in C^\infty(M)$. We denote by $\mathcal{P}_\pm$ and $\mathcal{Q}_\pm$ the projectors corresponding to $C_\pm$ and the co-metric $\mathcal{G}_\pm$. Additionally, let $C^{\text{inst}}_\pm \subset T^*_N M$ be affine constraint submanifolds corresponding to some instantaneous constraints imposed along N. We denote by $\mathcal{P}^{\text{inst,tr}}_\varepsilon$ and $\mathcal{Q}^{\text{inst,tr}}_\varepsilon$ the projectors associated with the submanifold $C^{\text{inst,tr}}_\varepsilon$ and the co-metric $\mathcal{G}_\varepsilon$. In the absence of instantaneous constraints, we denote by $\mathcal{P}^{\text{tr}}_\varepsilon$ and $\mathcal{Q}^{\text{tr}}_\varepsilon$ the projectors associated with the submanifold $C^{\text{tr}}_\varepsilon$ and the co-metric $\mathcal{G}_\varepsilon$. In case $N = \partial M$, we denote the latter with the superscript "∂" instead of "tr".

Let $y \in C_\varepsilon \cap T^*_N M$ be an ε-out point. The points associated with y resulting from the 1-st moves in a reflected or a refracted falling step are given, respectively, by equations (10) and (11). As before, the 2-nd move simply consists of computing the focusing points for these solutions with respect to $C^{\text{inst,tr}}_\varepsilon$

for a reflected falling step (respectively, $C_{\bar{\varepsilon}}^{\mathrm{inst,tr}}$ for a refracted falling step) or, if the instantaneous constraints are absent, $C_{\varepsilon}^{\mathrm{tr}}$ (respectively, $C_{\bar{\varepsilon}}^{\mathrm{tr}}$). This is done in terms of the corresponding projectors according to Section 5.1. According to Definition 4.18, this gives all the y-decisive points.

Proposition 5.5 *Let $y \in C_\varepsilon \cap T_N^* M$ be an ε-out point and assume that the constraints are linear. For $N = \partial M$, the unique y-reflected falling point is given by $\mathcal{P}_\varepsilon^{inst,\partial}(y)$ (or, in the absence of ε-instantaneous nonholonomic constraints, $\mathcal{P}_\varepsilon^{\partial}(y)$).*

Proof: From the previous discussion, we know that the points in the constrained characteristic passing through y with the same H_ε-energy level are y itself and $y + c_{\varepsilon,2}\mathcal{P}_\varepsilon(d_q f)$, $q = \pi(y)$ (cf. equation (10)). Now, note that $d_q f$ belongs to the $\mathcal{G}_\varepsilon$-orthogonal complement of $\alpha_{H_\varepsilon}^{-1}(T(\partial M))$, i.e.

$$\mathcal{G}_\varepsilon(d_q f, \beta) = d_q f(\alpha_{H_\varepsilon}(\beta)) = 0\,, \quad \beta \in \alpha_{H_\varepsilon}^{-1}(T(\partial M))\,.$$

Using the equality $C_\varepsilon^{\partial} = C_\varepsilon \cap \alpha_{H_\varepsilon}^{-1}(T(\partial M))$, we have that $d_q f \in \alpha_{H_\varepsilon}^{-1}(T(\partial M))^{\perp_\varepsilon}$ implies $\mathcal{P}_\varepsilon^{\mathrm{inst},\partial}(\mathcal{P}_\varepsilon(d_q f)) = 0$ and $\mathcal{P}_\varepsilon^{\partial}(\mathcal{P}_\varepsilon(d_q f)) = 0$. The result is then a consequence of Definition 4.18. ■

5.5 Energy behavior

In this section, we discuss the consequences regarding the energy behavior of the system that result from the application of the Transition Principle.

Lemma 5.6 *Given $y \in T_N^* M$, let $x = \mathcal{P}(y) + \mathcal{Q}(\Upsilon_q)$, $q = \pi(y)$, be the associated y-focusing point with respect to a submanifold $C \subset T^*M$. Then*

$$\hat{T}(x) \leq \hat{T}(y) + \hat{T}(\mathcal{Q}(\Upsilon_q), \mathcal{Q}(\Upsilon_q))\,,$$

and the equality holds if and only if y belongs to C^0.

Proof: Note that

$$\begin{aligned}
\mathcal{G}(\mathcal{P}(y) + \mathcal{Q}(\Upsilon_q), \mathcal{P}(y) + \mathcal{Q}(\Upsilon_q)) &= \mathcal{G}(\mathcal{P}(y), \mathcal{P}(y)) + \mathcal{G}(\mathcal{Q}(\Upsilon_q), \mathcal{Q}(\Upsilon_q))) \\
&\leq \mathcal{G}(\mathcal{P}(y), \mathcal{P}(y)) + \mathcal{G}(\mathcal{Q}(y), \mathcal{Q}(y)) + \mathcal{G}(\mathcal{Q}(\Upsilon_q), \mathcal{Q}(\Upsilon_q)) \\
&= \mathcal{G}(y, y) + \mathcal{G}(\mathcal{Q}(\Upsilon_q), \mathcal{Q}(\Upsilon_q))\,,
\end{aligned}$$

where we have used that $\mathcal{G}$ is positive-definite, and the fact that C^0 and $\mathrm{Ann}(\mathcal{D})$ are orthogonal spaces with respect to the co-metric $\mathcal{G}$. If the equality holds, then $\mathcal{G}(\mathcal{Q}(y), \mathcal{Q}(y)) = 0$, which is equivalent to $y \in C^0$. ■

As a consequence of this simple lemma we can conclude that in the case of linear constraints the Transition Principle always implies a loss of energy. This is a suitable generalization to constrained systems of the classical Carnot theorem for systems subject to impulsive forces [36].

Theorem 5.7 (Carnot's theorem for generalized linear constraints) *Suppose that the Hamiltonian system is subject to nonholonomic constraints given by a linear distribution. Then the Transition Principle implies always a loss of energy as the result of an "impact".*

Proof: Under linear constraints, note that $\Upsilon = 0$. From Lemma 5.6, we get $\hat{T}(x) \leq \hat{T}(y)$ with $F_{(H,C)}(y) = \{x\}$. The result now follows from the formulation of the Transition Principle and the definitions of decisive points in Section 4.5 (cf. Definitions 4.12-4.18). ■

Under linear constraints, the trajectory of the system maintains the same energy level after the application of the Transition Principle in the following cases:

(i) when the decisive points are determined according to Definitions 4.12 and 4.17 and the impact point $y \in T_N^*M$ belongs to $C_+ \cap C_-$; and

(ii) when the constraints are smooth, and therefore the decisive points are determined according to Proposition 4.15.

If the decisive points are determined according to Definitions 4.14 and 4.18, then nothing can be said in general. The refractive steps will typically imply an energy decrease.

Remark 5.8 This type of energy arguments also allows to discard as follows the possibility of chattering when computing the y-decisive points if the constraints change (see Definition 4.12 and Remark 4.13 above). Let $N = \{y \in T^*M \mid f(y) = 0\}$. Assume there is an infinite y-admissible sequence (y_i, ε_i), $i = 0, \dots, \infty$. For each i, we have that $y_i \neq y_{i+1}$, since otherwise

$$X^l_{H,C_{\varepsilon_{i+1}}}(f)(y_{i+1}) = X^l_H(f)(y_{i+1}) = X^l_{H,C_{\varepsilon_i}}(f)(y_i)\,, \quad \text{for all } l\,,$$

which together with the fact that y_i is a ε_i-out point, implies that y_{i+1} is a ε_{i+1}-in point. The latter contradicts the definition of admissible sequence. As a consequence of Lemma 5.6, we then have

$$\hat{T}(y_0) > \hat{T}(y_1) > \hat{T}(y_2) > \cdots > \hat{T}(y_i) \geq 0\,.$$

The limit of this sequence is zero, which implies that the y-decisive point corresponding to such a sequence would be 0, that is, the trajectory would get 'stuck' when reaching N.

5.6 Integrable constraints

The integrability of the constraints simplifies the application of the Transition Principle. Consider, for instance, the situation when the mechanical system is unconstrained on M_- and is subject to some generalized linear constraints $C = C^0$ on M_+ that turn out to be holonomic, i.e., $\alpha_H(C^0) = \mathcal{D}$ is integrable. Denote by $\{S_\alpha\}$, α being an m-dimensional parameter, the foliation of M_+ induced by $\mathcal{D}$. Locally this foliation is described by m functions $f_i \in C_\infty(M)$ such that

$$q \in S_\alpha \Longleftrightarrow f_i(q) = \alpha_i\,,\ 1 \leq i \leq m\,.$$

A similar situation has been treated in [35] in the context of totally inelastic collisions (note, however, that in [35] the integrable distribution is defined only on N, whereas here $\mathcal{D}$ is defined on M_+). The integrable constraints imposed by $\mathcal{D}$ can be interpreted as an abrupt reduction of the phase space of the mechanical system.

By definition, one has that $\operatorname{Ann}(\mathcal{D}) = \operatorname{span}\{df_1, \dots, df_m\}$. The matrix J in (8) is then given by $J = (\partial f_i / \partial q^a)$ and the projector $\mathcal{P}$ is $\mathcal{P}(x) = (1 - J^t \mathcal{B}^{-1} J G^{-1})x$. Let $y \in C_- \cap T_N^*M$ be the impact state of a trajectory $q(t)$ coming from M_-. From the discussion in Section 5.3, we obtain that the unique focusing point associated to y is $x = (1 - J^t \mathcal{B}^{-1} J G^{-1})y$. The trajectory will continue its motion in M_+, M_- or N depending on the in/out/trapping character of the focusing point x. If it evolves in M_+ (more precisely, in $S_\alpha \subset M_+$ with α such that $x \in S_\alpha$), we call it the 'refraction' of the original trajectory. If it evolves in M_-, we call it the 'reflection' of the original trajectory.

6 Examples

In this section we consider four examples to illustrate the theory exposed above. They all present the example of a rolling sphere considered in various constrained situations. The first one is taken from [16] and is treated here in order to provide a further comparison with previous approaches. The second one combines the presence of smooth nonholonomic constraints with discontinuous Hamiltonians and instantaneous constraints acting on the system along a hypersurface. The third one consists of a ball rolling on a rotating surface whose angular velocity is suddenly changed to a different value, and this is modeled via a discontinuous affine distribution of constraints. Finally, the fourth one presents a two-wheeled system with a rod of variable length and illustrates the application of the Transition Principle in both the elastic and the inelastic modes.

6.1 A rolling sphere

Consider a homogeneous sphere rolling on a plane. Assume it has unit mass ($m = 1$) and let k^2 be its inertia about any axis. Let (x, y) denote the position of the center of the sphere and let (φ, θ, ψ) denote the Eulerian angles. The configuration space is therefore $Q = \mathbb{R}^2 \times \mathrm{SO}(3)$. Assume that the plane is smooth if $x < 0$ and absolutely rough if $x > 0$ (see Figure 1). On the smooth half-plane, the motion of the sphere is assumed free, that is, the sphere can slip. On the rough half-plane, the sphere should roll without slipping due to the constraints imposed by the roughness. We are interested in determining the eventual sudden changes in the trajectories of the sphere when it reaches the line separating the smooth and the rough half-planes.

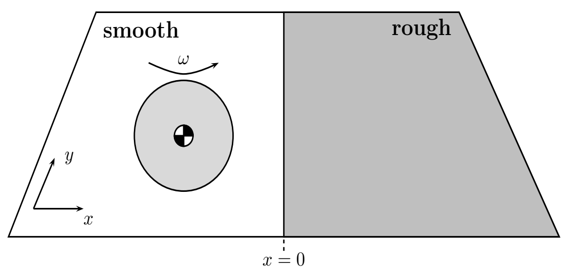

Figure 1: The rolling sphere on a 'special' surface.

The kinetic energy of the sphere is

$$T = \frac{1}{2}\left(\dot{x}^2 + \dot{y}^2 + k^2(\omega_x^2 + \omega_y^2 + \omega_z^2)\right) , \tag{12}$$

where ω_x, ω_y and ω_z are the angular velocities with respect to the inertial frame, given by

$$\omega_x = \dot{\theta}\cos\psi + \dot{\varphi}\sin\theta\sin\psi , \quad \omega_y = \dot{\theta}\sin\psi - \dot{\varphi}\sin\theta\cos\psi , \quad \omega_z = \dot{\varphi}\cos\theta + \dot{\psi} .$$

The condition of rolling without sliding of the sphere when $x > 0$ implies that the point of contact of the sphere and the plane has zero velocity

$$\phi^1 = \dot{x} - r\omega_y = 0 , \quad \phi^2 = \dot{y} + r\omega_x = 0 ,$$

where r is the radius of the sphere.

Following the classical procedure [30], we introduce quasi-coordinates 'q^1', 'q^2' and 'q^3' such that '$\dot{q}^1$'$= \omega_x$, '$\dot{q}^2$'$= \omega_y$ and '$\dot{q}^3$'$= \omega_z$. The latter merely have a symbolic meaning in the sense that in the present example, for instance, the partial derivative operators $\partial/\partial q^i$ should be interpreted as linear combinations of the partial derivatives with respect to Euler's angles. Also to the differential forms dq^i one should attach the appropriate meaning, i.e. they do not represent exact differentials but, instead, we should read them as $dq^1 = \cos\psi\, d\theta + \sin\theta \sin\psi\, d\varphi$, etc.

The singular hypersurface N is defined by $N = \{x = 0\}$. In this case, the constraints are linear and the nonholonomic distribution $\alpha_H(C) = \mathcal{D}$ on M_+ is given by

$$\mathcal{D}_{(x,y,q^1,q^2,q^3)} = \text{span}\left\{ r\frac{\partial}{\partial x} + \frac{\partial}{\partial q^2}, -r\frac{\partial}{\partial y} + \frac{\partial}{\partial q^1}, \frac{\partial}{\partial q^3} \right\}.$$

Here we are dealing with a single distribution which constrains the motion on M_+.

In the following we compute the decisive points for this example. Let $\lambda \in C_- \cap T^*M$ be a minus-out point. A direct computation shows that the expression of the projector $\mathcal{P} : T^*M \to C$ in local coordinates is

$$\mathcal{P} = \begin{pmatrix} \frac{r^2}{r^2+k^2} & 0 & 0 & \frac{r}{r^2+k^2} & 0 \\ 0 & \frac{r^2}{r^2+k^2} & -\frac{r}{r^2+k^2} & 0 & 0 \\ 0 & \frac{-rk^2}{r^2+k^2} & \frac{k^2}{r^2+k^2} & 0 & 0 \\ \frac{rk^2}{r^2+k^2} & 0 & 0 & \frac{k^2}{r^2+k^2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{13}$$

Therefore, the single focusing point for $\lambda \in T^*_N M$ is given by $x = \mathcal{P}(\lambda) \in C \cap T^*_N M$. If we denote $\lambda = (x_0, y_0, q^1_0, q^2_0, q^3_0, (p_x)_0, (p_y)_0, (p_1)_0, (p_2)_0, (p_3)_0)$ and $x = (x, y, q^1, q^2, q^3, p_x, p_y, p_1, p_2, p_3)$, we get

$$\begin{aligned} p_x &= \frac{r^2(p_x)_0 + r(p_2)_0}{r^2+k^2}, & p_1 &= \frac{-rk^2(p_y)_0 + k^2(p_1)_0}{r^2+k^2}, \\ p_y &= \frac{r^2(p_y)_0 - r(p_1)_0}{r^2+k^2}, & p_2 &= \frac{rk^2(p_x)_0 + k^2(p_2)_0}{r^2+k^2}, \\ & & p_3 &= (p_3)_0 . \end{aligned}$$

Note also that the focusing point with respect to $C_- = T^*M$ associated with x is x itself. Therefore, if x is a plus-out point, the only admissible sequence for λ is $\{(\lambda, -), (x, +), (x, -)\}$. If x is either a plus-in or a plus-trapping point, then the only admissible sequence for λ is $\{(\lambda, -), (x, +)\}$. The set of plus-trapping points for the dynamics X_{H,C_+} is $\partial(T^*M)^n = \{\mu \in T^*M \mid x = 0,\ p_x = 0\}$. Consequently, the trajectory is refracted, i.e., the sphere follows its motion on M_+ under the dynamics X_{H,C_+} (rolling without slipping) if $p_x \geq 0$. Otherwise (i.e., if $p_x < 0$), the trajectory is reflected by the "roughness" and continues in M_- under the dynamics X_H starting from x.

6.2 A rolling sphere hitting a wall

This is a classical example [13, 19, 30] that we treat here for the sake of completeness. Consider again a homogeneous sphere of radius r and unit mass. Assume that the sphere rolls without sliding on a horizontal table, and that at a certain instant of time it hits a wall determined by the plane $x = d > 0$ (cf. Figure 2). When this happens, the following constraint is instantaneously imposed on the system,

$$\psi = \dot{y} - r\omega_z = 0\,.$$

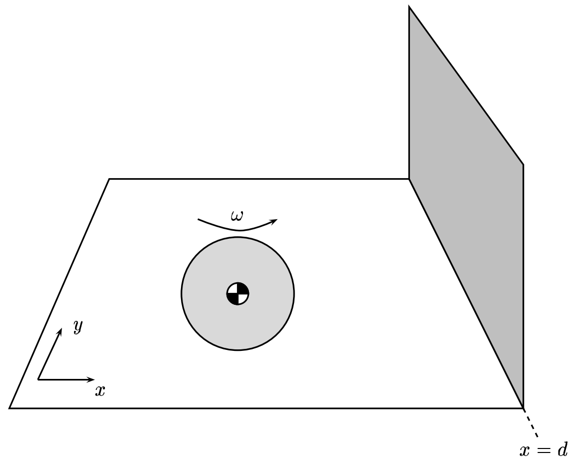

Figure 2: A rolling sphere that eventually hits a wall.

Therefore, we are in the situation explained in Remark 4.16. The configuration space of the system is $M = M_+ = \{x < d\}$, with the boundary $N = \{x = d\}$, and the linear constraint submanifold $C = C_+ \subset T^*M$ is given by $\alpha_H(C) = \mathcal{D}$,

$$\mathcal{D} = \operatorname{span}\left\{ r\frac{\partial}{\partial x} + \frac{\partial}{\partial q^2}, -r\frac{\partial}{\partial y} + \frac{\partial}{\partial q^1}, \frac{\partial}{\partial q^3} \right\}.$$

The expression for the projector $\mathcal{P} : T^*M \to C$ is given by equation (13). The submanifold giving the instantaneous constraints along N is

$$C^{\text{inst}} = \{\lambda \in C \mid \psi(\alpha_H(\lambda)) = 0\}.$$

The projector $\mathcal{P}^{\text{inst}} = \mathcal{P}_+^{\text{inst}} : T^*M \to C^{\text{inst}}$ is

$$\mathcal{P}^{\text{inst}}(\lambda) = \frac{r\lambda_x + \lambda_2}{r^2 + k^2}\left(r dx + k^2 dq^2\right) + \frac{-r\lambda_y + \lambda_1 - \lambda_3}{r^2 + 2k^2}\left(-r dy + k^2 dq^1 - k^2 dq^3\right). \tag{14}$$

Let $\lambda = (x_0, y_0, q_0^1, q_0^2, q_0^3, (p_x)_0, (p_y)_0, (p_1)_0, (p_2)_0, (p_3)_0) \in C_+ \cap T_N^*M \subset T^*M$ be a plus-out point, i.e., $\mathcal{G}(\lambda, dx) < 0$. We first consider an elastic impact. Since $H_- = \infty$, we only compute the outcome of a reflective step. According to (10), the points in the constrained characteristic passing through λ within the same H_+-energy level are

$$\lambda \quad \text{and} \quad \lambda + c_{+,2}\mathcal{P}(dx), \quad \text{with } c_{+,2} = -2\,\frac{r^2 + k^2}{r^2}(p_x)_0.$$

The associated focusing points are given by

$$\mathcal{P}^{\text{inst}}(\lambda) \quad \text{and} \quad \mathcal{P}^{\text{inst}}(\lambda) + c_{+,2}\mathcal{P}^{\text{inst}}(\mathcal{P}(dx)). \tag{15}$$

Note that $\mathcal{P}^{\text{inst}}(\mathcal{P}(dx)) = \mathcal{P}^{\text{inst}}(dx) = \mathcal{P}(dx)$, and therefore the points in (15) belong to the same constrained characteristic and to the same H_+-energy level. Denoting the coordinates of the point

$\mathcal{P}^{\rm inst}(\lambda)$ by $(x, y, q^1, q^2, q^3, p_x, p_y, p_1, p_2, p_3)$, we get

$$p_x = (p_x)_0 \qquad p_y = \frac{(r^2+k^2)(p_y)_0 + r(p_3)_0}{r^2+2k^2}\,, \qquad \begin{aligned} p_1 &= -k^2\frac{(r^2+k^2)(p_y)_0 + r(p_3)_0}{r(r^2+2k^2)} \\ p_2 &= k^2\frac{(p_x)_0}{r} \\ p_3 &= k^2\frac{(r^2+k^2)(p_y)_0 + r(p_3)_0}{r(r^2+2k^2)}\,, \end{aligned}$$

Now, notice that $\mathcal{P}^{\rm inst}(\lambda)$ is a plus-out point, because $\mathcal{G}(\mathcal{P}^{\rm inst}(\lambda), dx) = \mathcal{G}(\lambda, dx) < 0$. Therefore, following Proposition 5.4, we conclude that the sequence $\{(\lambda, +), \mathcal{P}^{\rm inst}(\lambda) + c_{+,2}\mathcal{P}^{\rm inst}(dx), +)\}$ is λ-admissible, and $\mathcal{P}^{\rm inst}(\lambda) + c_{+,2}\mathcal{P}^{\rm inst}(dx)$ is a decisive point. The other possible λ-admissible sequence corresponds to

$$\{(\lambda, +), \mathcal{P}^{\rm inst}(\lambda), \mathcal{P}^{\rm inst}(\lambda) + c_{+,2}\mathcal{P}^{\rm inst}(dx), +)\}\,,$$

but renders the same decisive point.

In the case of an inelastic impact, Proposition 5.5 yields that the unique λ-decisive point is $\mathcal{P}^{{\rm inst},\partial}(\lambda)$. After the impact, the ball continues its motion along the wall under the dynamics specified by the vector field $X_+^{{\rm inst},\partial}$.

6.3 A rolling sphere on a rotating table

Consider again a homogeneous sphere of radius r and unit mass. Assume that the sphere rolls without sliding on a horizontal table which is rotating with certain constant angular velocity about a vertical axis through one of its points (see Figure 3). Let Ω_- and Ω_+ be two angular velocities. Here we consider the following situation: each time the sphere reaches the hypersurface $x = y$, an impulsive force is exerted on the table to put it spinning with a different angular velocity. That is, if the angular velocity of the table was Ω_-, the force applied on its rotation axis changes it to Ω_+ and vice versa. We assume that $\Omega_- < \Omega_+$. This can be modeled as thinking of a system which is subject to two different affine constraint distributions. In order for this model to be consistent, we also assume that the surface of the table is rough enough so that sphere is rolling without slipping at all times.

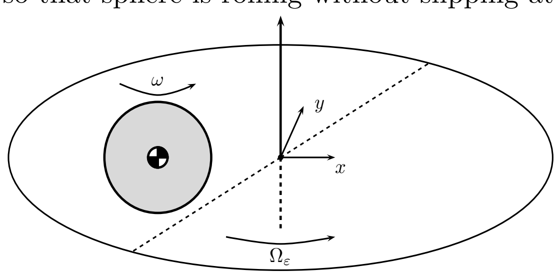

Figure 3: A rolling sphere on a rotating table.

The Lagrangian is again given by equation (12). The nonholonomic constraints are now affine in the velocities,

$$\dot{x} - r\omega_y = -\Omega y, \qquad \dot{y} + r\omega_x = \Omega x\,.$$

The constraint space $\alpha_H(C)$ is then described by

$$\alpha_H(C) = \mathcal{D} + Y = \mathrm{span}\left\{ r\frac{\partial}{\partial x} + \frac{\partial}{\partial q^2}, -r\frac{\partial}{\partial y} + \frac{\partial}{\partial q^1}, \frac{\partial}{\partial q^3} \right\} + Y\,,$$

where Y is the vector field defined by

$$Y = -\Omega y \frac{\partial}{\partial x} + \Omega x \frac{\partial}{\partial y} \, .$$

Note that the projection of $\Upsilon = \mathcal{L}_L(Y)$ to $\mathrm{Ann}(\mathcal{D})$ is given by

$$\mathcal{Q}(\Upsilon) = \frac{\Omega k^2}{r^2 + k^2} \left(-y dx + x dy - x r dq^1 - y r dq^2 \right) \, .$$

Following the discussion for the case of affine constraints, given $y \in T^*M$, the focusing point with respect to $C_\pm$ is given by

$$x = \mathcal{P}(y) + \mathcal{Q}(\Upsilon_\pm) \, ,$$

where $\mathcal{P}$ is the projector in (13).

Assume that the sphere is rolling on the hyperplane $M_- = \{x < y\}$ and that the constant angular velocity of the table is Ω_-. Consider the case when the sphere "hits" the hypersurface $N = \{x = y\}$ with the impact state

$$\lambda = (x_0, y_0, q_0^1, q_0^2, q_0^3, (p_x)_0, (p_y)_0, (p_1)_0, (p_2)_0, (p_3)_0) \in C_- = C_-^o + \Upsilon_- \, .$$

Denote the coordinates of the associated focusing point by

$$x = \mathcal{P}(\lambda) + \mathcal{Q}(\Upsilon_+) = (x, y, q^1, q^2, q^3, p_x, p_y, p_1, p_2, p_3) \, .$$

Then

$$\mathcal{G}(df, x) = p_x - p_y = (p_x)_0 - (p_y)_0 + \frac{k^2}{r^2 + k^2} (x_0 + y_0)(\Omega_- - \Omega_+) \, . \tag{16}$$

Given that λ is an minus-out point, we have that $\mathcal{G}(df, \lambda) = (p_x)_0 - (p_y)_0 > 0$. If $x_0 = y_0 < 0$, then the second term in (16) is also positive, and $\{(\lambda, -), (x, +)\}$ is the unique admissible sequence for λ. In this case, x is the λ-decisive point. On the contrary, for certain values of $x_0 = y_0 > 0$, it might happen that $\mathcal{G}(df, x)$ is negative, i.e., that x is a plus-out point. Now, note that the focusing point associated with x is λ itself, since

$$\mathcal{P}(x) + \mathcal{Q}(\Upsilon_-) = \mathcal{P}(\mathcal{P}(\lambda)) + \mathcal{Q}(\Upsilon_-) = \mathcal{P}(\lambda) + \mathcal{Q}(\Upsilon_-) = \lambda \, .$$

As a consequence, in this case there would not be any λ-decisive point. This problem stems from the fact the modeling of this example as a system subject to affine constraints does not take into account that the jump in the angular velocity of the table takes place *no matter what*. Therefore, after the impact, we should really regard C_+ as the new set of affine constraints acting on the whole configuration manifold. With this interpretation, x would obviously be a plus-in point (and hence decisive). In other words, the trajectory of the ball gets reflected back by the blow of the greater velocity Ω_+.

6.4 A two-wheeled system with a rod of variable length

Consider a system composed of two wheels of different radii, $r_1 < r_2$, connected by a massless rod of variable length ℓ (see Figure 4). For simplicity, assume that the two-wheeled system moves along a line, and that both the masses and the momenta of inertia of the wheels are unitary. The wheels are subject to the standard constraints of non-slipping. Assume that the length ℓ of the rod is constrained

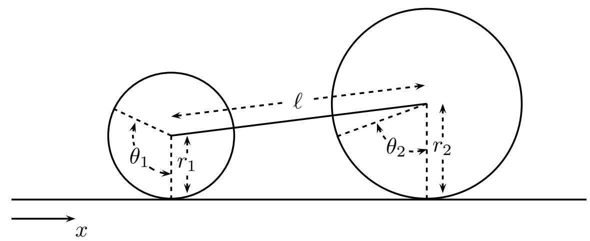

Figure 4: A two-wheeled system with a rod of variable length.

between a minimum length a and a maximum length b. Here we consider the following two situations: (i) when the length ℓ of the rod becomes extreme, an elastic impact occurs; (ii) when the length ℓ of the rod becomes extreme, an arresting device fixes it, and therefore an inelastic impact occurs.

The Lagrangian of the system is given by the kinetic energy of the wheels

$$L = \frac{1}{2}\left(\dot\theta_1^2 + \dot\theta_2^2 + \dot x_1^2 + \dot x_2^2\right).$$

The conditions of rolling without sliding are encoded in the constraints

$$\dot x_1 - r_1\dot\theta_1 = 0\,, \quad \dot x_2 - r_2\,\dot\theta_2 = 0\,,$$

which, since we are considering the motion of the two-wheeled system only along a line, turn out to be holonomic. The constraint on the length of the rod is given by

$$a \leq \ell = \sqrt{(r_2-r_1)^2 + (x_2-x_1)^2} \leq b\,.$$

Following Remark 4.16, we set $M_- = \emptyset$, $M_+ = M = \{(x_1,x_2,\theta_1,\theta_2) \in \mathbb{R}^2 \times \mathbb{S}^1 \times \mathbb{S}^1 \mid a \leq \ell(x_1,x_2,\theta_1,\theta_2) \leq b\}$, with boundary set $N = \partial M = \{(x_1,x_2,\theta_1,\theta_2) \in \mathbb{R}^2 \times \mathbb{S}^1 \times \mathbb{S}^1 \mid \ell(x_1,x_2,\theta_1,\theta_2) = a \text{ or } \ell(x_1,x_2,\theta_1,\theta_2) = b\}$, and linear constraint submanifold $C = C_+ \subset T^*M$ given by $\alpha_H(C) = \mathcal{D}$,

$$\mathcal{D} = \operatorname{span}\left\{r_1\frac{\partial}{\partial x_1} + \frac{\partial}{\partial \theta_1}, r_2\frac{\partial}{\partial x_2} + \frac{\partial}{\partial \theta_2}\right\}.$$

The expression for the projector $\mathcal{P} : T^*M \to C$ in local coordinates is given by the following matrix

$$\mathcal{P} = \begin{pmatrix} \frac{r_1^2}{1+r_1^2} & 0 & \frac{r_1}{1+r_1^2} & 0 \\ 0 & \frac{r_2^2}{1+r_2^2} & 0 & \frac{r_2}{1+r_2^2} \\ \frac{r_1}{1+r_1^2} & 0 & \frac{1}{1+r_1^2} & 0 \\ 0 & \frac{r_2}{1+r_2^2} & 0 & \frac{1}{1+r_2^2} \end{pmatrix}.$$

Let $\lambda \in T^*M_+$ be a plus-out point with $\ell(\lambda) = b$ and $\mathcal{G}(\lambda, d\ell) > 0$. Since $H_- = \infty$, we only compute the outcome of a reflective step. Following equation (10), the points in the constrained characteristic passing through λ with the same H_+-energy level are λ and $\lambda + c_{+,2}\mathcal{P}(d\ell)$, with

$$c_{+,2} = -\frac{2\,\mathcal{G}(\lambda, \mathcal{P}(d\ell))}{\mathcal{G}(\mathcal{P}(d\ell), \mathcal{P}(d\ell))}\,. \tag{17}$$

According to Proposition 5.4, the point $\lambda + c_{+,2}\mathcal{P}(d\ell)$ is $+$-decisive.

Consider now an inelastic impact, i.e., the case when the length ℓ of the rod becomes fixed after the impact. Since there are no additional instantaneous constraints imposed on the system at the impact state, we compute the decisive points with regards to the boundary of the constraint manifold,

$$C^\partial = C \cap \alpha_{H_+}^{-1}(T\partial M) = \{(x_1, x_2, \theta_1, \theta_2, p_{x_1}, p_{x_2}, p_{\theta_1}, p_{\theta_2}) \in T^*M \mid p_{x_1} = p_{x_2}, \\ p_{x_1} = r_1 p_{\theta_1},\, p_{x_2} = r_2 p_{\theta_2},\, \ell(x_1, x_2, \theta_1, \theta_2) = a \text{ or } \ell(x_1, x_2, \theta_1, \theta_2) = b\}\,.$$

As before, we only compute the outcome of a reflective step. Following Proposition 5.5, we deduce that the unique decisive point is $\mathcal{P}^\partial(\lambda)$. After the inelastic impact, the length of the rod is fixed forever after, the velocities of the two wheels of the system are reset according to $\mathcal{P}^\partial(\lambda)$ and evolve according to $X^\partial_{(H,C,N)}$.

7 Conclusions

We have developed a generalization of the Transition Principle to deal with impulsive regimes in general nonholonomic systems, and particularized our discussion to the case of mechanical systems with affine constraints. We have also provided a geometric formulation of the dynamics of nonholonomic Hamiltonian systems via partial symplectic structures. Future work will be devoted to the development of a suitable version of the Transition Principle for optimal control problems, the comparison of quantitative and qualitative predictions made by the Transition Principle in specific examples, and the implementation of the results obtained here in numerical schemes for impulsive nonholonomic systems.

Acknowledgments

J. Cortés was partially supported by faculty research funds granted by the University of California, Santa Cruz. The authors would like to thank the reviewer for his/her comments on the paper.

References

[1] R. Abraham, J.E. Marsden, T.S. Ratiu: *Manifolds, Tensor Analysis and Applications*. 2nd ed., Springer-Verlag, New-York-Heidelberg-Berlin, 1988.

[2] P. Appell: *Traité de Mécanique Rationnelle*, 6th ed, tome II. Gauthier-Villars, Paris, 1953.

[3] L. Bates, J. Śniatycki: Nonholonomic reduction. *Rep. Math. Phys.* **32** (1) (1992), 99-115.

[4] A.M. Bloch: *Nonholonomic Mechanics and Control*. Interdisciplinary Applied Mathematics Series, vol 24, Springer-Verlag, New York, 2003.

[5] A.M. Bloch, P.S. Krishnaprasad, J.E. Marsden, R.M. Murray: Nonholonomic mechanical systems with symmetry. *Arch. Rat. Mech. Anal.* **136** (1996), 21-99.

[6] A.V. Bocharov, A.M. Vinogradov: Appendix II of A.M. Vinogradov, B.A. Kuperschmidt: The structures of Hamiltonian mechanics. *Russ. Math. Surv.* **32** (4) (1977), 177-243.

[7] R.W. Brockett: Hybrid models for motion control systems. In *Essays in Control: Perspectives in the Theory and its Applications*, Birkhäuser, Boston, MA, 1993, pp. 29-53.

[8] B. Brogliato: *Nonsmooth Impact Mechanics: Models, Dynamics and Control*. Lectures Notes in Control and Inform. Sci. 220, Springer, New York, 1996.

[9] B. Brogliato: Some perspectives in the analysis and control of complementarity systems. *IEEE Trans. Automatic Control* **48** (6) (2003), 918-935.

[10] F. Bullo, M. Zefran: Modeling and controllability for a class of hybrid mechanical systems. *IEEE Trans Robotics Automation* **18** (4) (2002), 563-573.

[11] F. Cantrijn, M. de León, D. Martín de Diego: On almost-Poisson structures in nonholonomic Mechanics. *Nonlinearity* **12** (1999), 721-737.

[12] J.F. Cariñena, M.F. Rañada: Lagrangian systems with constraints: a geometric approach to the method of Lagrange multipliers. *J. Phys. A: Math. Gen.* **26** (6) (1993), 1335-1351.

[13] H. Cendra, E.A. Lacomba, W. Reartes: The Lagrange-d'Alembert-Poincaré equations for the symmetric rolling sphere. In *Proceedings of the Sixth "Dr. Antonio A. R. Monteiro" Congress of Mathematics*, E. Fernández, A. Germani, L. Monteiro, H. Cendra, L. Piovan. eds, Univ. Nac. Sur Dep. Mat. Inst. Mat., Bahía Blanca, 2001, pp 19-32.

[14] B. Chen, L.S. Wang, S.S. Chu, W.T. Chou: A new classification of nonholonomic constraints. *Proc. R. Soc. Lond. A* **453** (1997), 631-642.

[15] J. Cortés: *Geometric, control and numerical aspects of nonholonomic systems*. Lecture Notes in Mathematics Series, vol 1793, Springer-Verlag, Berlin-Heidelberg-New York, 2002.

[16] J. Cortés, M. de León, D. Martín de Diego, S. Martínez: Mechanical systems subjected to generalized nonholonomic constraints. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.* **457** (2007) (2001), 651-670.

[17] J. Cortés, S. Martínez: The consistency problem in optimal control: the degenerate case. *Rep. Math. Phys.* **51** (2/3) (2003), 171-186.

[18] M. Favretti: On the use of configuration variables as control variables in mechanical systems. *Proc. IEEE Conf. Decision and Control*, San Diego, CA, 1997, pp. 4210-4215.

[19] A. Ibort, M. de León, E. Lacomba, J.C. Marrero, D. Martín de Diego, P. Pitanga: Geometric formulation of Carnot's theorem. *J. Phys. A: Math. Gen.* **34** (2001), 1691-1712.

[20] S. Kobayashi, K. Nomizu: *Foundations of Differential Geometry*. Interscience Tracts in Pure and Applied Mathematics. Interscience Publishers, Wiley, New-York, 1963.

[21] J. Koiller: Reduction of some classical nonholonomic systems with symmetry. *Arch. Rat. Mech. Anal.* **118** (1992), 113-148.

[22] E.A. Lacomba, W.A. Tulczyjew: Geometric formulation of mechanical systems with one-sided constraints. *J. Phys. A: Math. Gen.*, **23** (1990), 2801-2813.

[23] M. de León, D. Martín de Diego: On the geometry of nonholonomic Lagrangian systems. *J. Math. Phys.* **37** (1996), 3389-3414.

[24] T. Levi-Civita, U. Amaldi: *Lezioni di Meccanica Razionale*, vol.II, parte II, Zanicchelli, Bologna, 1952.

[25] P. Libermann, C.-M. Marle: *Symplectic Geometry and Analytical Mechanics*. D. Reidel Publ. Comp., Dordrecht, 1987.

[26] F. Lizzi, G. Marmo, G. Sparano, A.M. Vinogradov: Eiconal type equations for geometrical singularities of solutions in field theory. *J. Geom. Phys.* **14** (1994), 211-235.

[27] V.V. Lychagin: Geometric theory of singularities of solutions of nonlinear differential equations. *J. Soviet Math.* **51** (1990), 2735-2757.

[28] C.-M. Marle: Reduction of constrained mechanical systems and stability of relative equilibria. *Commun. Math. Phys.* **174** (1995), 295-318.

[29] J.J. Moreau: *Mécanique classique*. Tome II, Masson, Paris, 1971.

[30] J.I. Neimark, N.A. Fufaev: *Dynamics of Nonholonomic Systems*. Translations of the American Mathematical Society, vol. 33, Providence, Rhode Island, 1972.

[31] P. Painlevé: *Cours de Mécanique*. Tome I, Paris, Gauthier-Villars, 1930.

[32] L.A. Pars: *A Treatise on Analytical Dynamics*. Ox Bow, Woodbridge, CT, 1965.

[33] S. Pasquero: Ideality criterion for unilateral constraints in time-dependent impulsive mechanics. *J. Math. Phys.* **46** (2005), 112904.

[34] F. Pugliese, A.M. Vinogradov: On the geometry of singular Lagrangians. *J. Geom. Phys.* **35** (2000), 35-55.

[35] F. Pugliese, A.M. Vinogradov: Geometry of inelastic collisions. *Acta Appl. Math.* **72** (2002), 77-85.

[36] R. Rosenberg: *Analytical Dynamics*. Plenum Press, New York, 1977.

[37] D.E. Stewart: Rigid-body dynamics with friction and impact. *SIAM Review* **42** (2000), 3-39.

[38] W.J. Stronge: Rigid body collisions with friction. *Proc. R. Soc. Lond. A* **431** (1990), 169-181.

[39] A.J. van der Schaft, B.M. Maschke: On the Hamiltonian formulation of nonholonomic mechanical systems. *Rep. Math. Phys.* **34** (1994), 225-233.

[40] A.J. van der Schaft, J.M. Schumacher: *An Introduction to Hybrid Dynamical Systems*. Lecture Notes in Control and Information Sciences Series, vol. 251, Springer-Verlag, New-York, 1999.

[41] A.M. Vinogradov: Geometry of nonlinear differential equations. *J. Soviet Math.* **17** (1981), 1624-1649.

[42] L.S. Wang, W.T. Chou: The analysis of constrained impulsive motion. *J. Appl. Mech.-Trans. of the ASME* **70** (2003) (4), 583-594.