Chapter 1

Nonholonomically constrained motions

1.1 Newton's equations

An English translation[1] of Newton's formulation of his *second law* reads: "A change in motion is proportional to the motive force impressed and takes place along a straight line in which that force is impressed." The current formulation reads: mass times acceleration is equal to the force. It uses the term *acceleration* for change of motion.

We will use a suitably generalized version of Newton's second law as our starting point in our presentation of the dynamics of nonholonomically constrained systems on manifolds. The main point of our generalization of Newton's equations is giving an appropriate definition of the notion of acceleration.

When configuration space is $\mathbb{R}^3$, the motion of a particle is described by giving its position vector $\mathbf{q}$ as a function of time, namely, $t \mapsto \mathbf{q}(t)$. In this case, its velocity $\mathbf{v}(t) = \frac{\mathrm{d}\mathbf{q}(t)}{\mathrm{d}t}$ and acceleration $\mathbf{a}(t) = \frac{\mathrm{d}\mathbf{v}(t)}{\mathrm{d}t}$ are well defined vectors in $\mathbb{R}^3$. If one interprets force $\mathbf{F}$ as a vector in $\mathbb{R}^3$, then the usual formulation of the second law reads

$$m\,\mathbf{a}(t) = \mathbf{F}, \tag{1}$$

where m is the mass of the particle. The force $\mathbf{F}$ acting on the particle may depend on $\mathbf{q}(t)$, $\mathbf{v}(t)$, and t. The kinetic energy of a particle with mass m and velocity $\mathbf{v}$ is $k = \frac{1}{2}\,m\,\langle \mathbf{v}, \mathbf{v}\rangle$, where $\langle\ ,\ \rangle$ is the standard Euclidean metric on $\mathbb{R}^3$. Thus k is one half of length of $\mathbf{v}$ with respect to the *kinetic energy metric* $\mathrm{k} = m\,\langle\ ,\ \rangle$.

For a system of n interacting particles in $\mathbb{R}^3$, the configuration space is $(\mathbb{R}^3)^n = \mathop{\times}\limits_{\alpha=1}^{n} \mathbb{R}^3_\alpha$. The position $\mathbf{q}$ of the system, its velocity $\mathbf{v}$ and acceleration

[1] See page 417 of [83].

$\mathbf{a}$ are vectors in $(\mathbb{R}^3)^n$. In this case the second law reads

$$m_\alpha \, \mathbf{a}_\alpha(t) = \mathbf{F}_\alpha, \quad \text{for } \alpha = 1, \ldots, n. \tag{2}$$

Here the subscript α labels the individual particles having mass m_α, acceleration $\mathbf{a}_\alpha$, and subjected to a force $\mathbf{F}_\alpha$. In this case the kinetic energy of the system is one half the square of the length of the velocity with respect to the kinetic energy metric

$$\mathrm{k} = \sum_{\alpha=1}^{n} m_\alpha \, \langle \ , \ \rangle_\alpha, \tag{3}$$

where $\langle \, , \, \rangle_\alpha$ is the standard Euclidean metric on ${\mathbb{R}^3}_\alpha$. If we interpret the force $\mathbf{F}$ imposed on the system as a covector on $(\mathbb{R}^3)^n$, then the left hand side of (2) can be interpreted as the covector given by evaluating the kinetic energy metric k, given by (3), on the vector $\mathbf{a}(t)$. Since the coefficients of k are *constant*, the derivative of the velocity $\mathbf{v}(t)$ coincides with the covariant derivative of $\mathbf{v}(t)$ with respect to the *Levi-Civita connection* ∇ associated to the kinetic energy metric k. This interpretation enables us to extend Newton's equations to manifolds.

Consider now a mechanical system with configuration space Q, which is a smooth manifold. The kinetic energy of the system defines a Riemannian metric k on Q such that for each $v \in T_qQ$, the kinetic energy of the motion of the system with velocity v is $k(v) = \frac{1}{2}\,\mathrm{k}(v,v)$. We shall refer to k as the *kinetic energy metric on the manifold* Q of the system. The metric k gives rise to the vector bundle isomorphisms $\mathrm{k}^\flat : TQ \to T^*Q$ and $\mathrm{k}^\sharp : T^*Q \to TQ$ where

$$\langle \mathrm{k}^\flat(v) \mid u \rangle = \mathrm{k}(v,u) \quad \text{for every } u, v \in T_qQ, \tag{4}$$

and $\mathrm{k}^\sharp = (\mathrm{k}^\flat)^{-1}$.

Let $t \mapsto q(t)$ be a smooth curve on Q which describes a motion of our system. Let $t \mapsto v(t) = \dot{q}(t)$ be its tangent lift to TQ. In other words, $v(t)$ is the *velocity* of the system at time t. Define the *acceleration* of the system as the covariant derivative of its velocity with respect to the Levi-Civita connection ∇ associated to the kinetic energy metric k. In other words, the acceleration $a(t)$ of the system at time t is

$$a(t) = \nabla_v v(t) = \ddot{q}(t). \tag{5}$$

A connection is needed to interpret acceleration as a tangent vector. Throughout the rest of this chapter we will use the Levi-Civita connection associated to the kinetic energy metric.

Let $\pi_Q : T^*Q \to Q$ and $\tau_Q : TQ \to Q$ be the cotangent and tangent bundle projection maps, respectively and observe that $\tau_Q = \pi_Q \circ \mathrm{k}^\flat$. A *force* acting on a mechanical system, which is in the configuration $q \in Q$, is a *covector* in T_q^*Q. Consider a map $\varphi : \mathbb{R} \times TQ \to T^*Q$ such that $\pi_Q \circ \varphi = \tau_Q \circ \mathrm{pr}_2$, where $\mathrm{pr}_2 : \mathbb{R} \times TQ \to TQ : (t, v_q) \mapsto v_q$. Then φ describes the dependence of the force on time, position and velocity of the system. In other words, if $t \in \mathbb{R}$ and $v_q \in T_qQ$, then $\varphi(t, v_q)$ is the force acting on the system at time t, position q, and velocity v_q. *Newton's equations* for the motion $t \mapsto q(t)$ of the system are

$$\mathrm{k}^\flat(\ddot{q}(t)) = \varphi(t, \dot{q}(t)). \tag{6}$$

In what follows we shall show that our formulation of Newton's equations motion for a mechanical system subject to a linear nonholonomic constraint is equivalent to all other standard formulations of its equations of motion.

1.2 Constraints

Constraints in dynamics are restrictions on positions and velocities of the system. Phenomenological constraints are introduced instead of unknown forces to describe observed motions. For example, a rigid body is a system of material points with fixed distances between each pair of points. Another example is the no slip condition in the motion of a rolling body. This constraint requires that the relative velocity of the point of contact of the rolling body vanishes. In the first example, the constraint depends only on the position of the material point. Such constraints are called *holonomic*. In the second example, the no slip condition is a linear relation on the velocity of the motion. Such conditions are called *linear nonholonomic constraints*. Nonlinear nonholonomic constraints appear in control theory, but they will not be considered here, see [74]. Neither shall we consider inhomogeneous linear nonholonomic constraints, which appear in the problem of a sphere rolling on a turntable when regarded in rotating coordinates [89].

There are two ways of dealing with holonomic constraints. First, we can modify the configuration space by imposing the holonomic constraints. Alternatively, we can present the holonomic constraints as linear nonholonomic constraints by extending the tangent bundle of the constraint manifold to a distribution on configuration space. Recall that a *distribution* on a smooth manifold M is a smooth vector subbundle of the tangent bundle TM of M. The original (holonomic) constraint manifold is then an *integral manifold* of this distribution.

Let D be a distribution on Q describing the linear nonholonomic constraints under consideration. A motion $t \mapsto q(t)$ of the system is *dynamically allowed* if its velocity $\dot{q}(t)$ at time t lies in $D_{q(t)}$. If D is involutive,[2] then we are dealing with a family of holonomically constrained systems. Here each integral manifold M of D describes a holonomic constraint.

Our nonholonomically constrained mechanical system is acted on by an *external force* φ_{ext} and a *reaction force* of the constraints $\varphi_{\mathrm{constr}}$. We have hypothesized that the constraints do not depend on time. In particular, the reaction force of the constraint depends only on the velocity and has no explicit dependence on time. We assume that the external force is *conservative*, that is, there is a *potential function* $V \in C^{\infty}(Q)$ such that $\varphi_{\mathrm{ext}} = -\,\mathrm{d}V$. Then *Newton's equation for the nonholonomically constrained system* is

$$\mathrm{k}^{\flat}(\ddot{q}(t)) = -\pi_Q^* \,\mathrm{d}V(q(t)) + \varphi_{\mathrm{constr}}(\dot{q}(t)). \tag{7}$$

In order to be able to solve equation (7) we have to make an assumption on the nature of the reaction force of the constraints. Let $D^0 \subseteq T^*Q$ be the set of all covectors which annihilate the corresponding vector in the constraint distribution D. In other words, for each $q \in Q$,

$$D_q^0 = D^0 \cap T_q^*Q = \{p \in T_q^*Q \,\big|\, \langle p \mid v \rangle = 0 \quad \text{for every } v \in D_q\}. \tag{8}$$

Interpreting covectors in T^*Q as forces, we see that D^0 is the set of all forces which do no work on virtual displacements in D. We now make the following

Hypothesis 1.2.1. *The work of the reaction force of the constraints on virtual displacements in D vanishes. In other words,*

$$\varphi_{\mathrm{constr}}(v) \in D^0_{\tau_Q(v)} \quad \textit{for every } v \in D.$$

Constraints satisfying this hypothesis are called *perfect*. A motion given by a curve $t \mapsto q(t)$ in Q with $\dot{q}(t) \in D_{q(t)}$, which satisfies Newton's equation (7) with constraint force $\varphi_{\mathrm{constr}}$, is called *dynamically admissible*. We summarize the above discussion as

[2]Frobenius' theorem states that if D is involutive, then D is integrable.

Theorem 1.2.2. (d'Alembert principle) *A smooth curve* $q : (t_0, t_1) \subseteq \mathbb{R} \to Q : t \mapsto q(t)$ *is a dynamically admissible motion of a mechanical system with kinetic energy* $k(v) = \frac{1}{2}\,\mathrm{k}(v, v)$ *and potential energy* V *subject to the linear nonholonomic constraint* $D \subseteq TQ$ *if and only if*

$$\mathrm{k}^\flat(\ddot{q}(t)) + \mathrm{d}V(q(t)) \in D^0_{q(t)} \;\; \text{and} \;\; \dot{q}(t) \in D_{q(t)}, \tag{9}$$

for all $t \in (t_0, t_1)$.

1.3 Lagrange-d'Alembert equations

The formulation of the dynamics of constrained motion given in (9) can be expressed in terms of the *Lagrangian* $\ell = k - \tau_Q^* V : TQ \to \mathbb{R}$ for the unconstrained system. Here $\tau_Q^* V = V \circ \tau_Q$ is the pull back of V by the tangent bundle projection map τ_Q. Let $c : \mathbb{R} \to Q : t \mapsto c(t)$ be a smooth curve on Q and let $\dot{c} : \mathbb{R} \to TQ : t \mapsto Tc(t)$ be its tangent lift to TQ. The Lagrange derivative of ℓ at $\dot{c}(t)$ is a 1-form on $T_{c(t)}Q$, which is defined as follows. Let $\{q^i\}$ be local coordinates on Q and let $\{q^i, v^i\}$ be the corresponding local coordinates on TQ. Then the curves c and $\dot{c}$ are given by $t \mapsto q(t) = (q^i(t))$ and $t \mapsto (q(t), \dot{q}(t)) = (q^i(t), v^i(t))$, respectively. The Lagrangian ℓ along $\dot{c}$ is a function of $(q(t), \dot{q}(t))$. Its *Lagrange derivative* $\delta\ell$ at $(q(t), \dot{q}(t))$ is the 1-form on $T_{q(t)}Q$

$$\delta\ell(q(t), \dot{q}(t)) = \left\{ \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \ell}{\partial \dot{q}^i} - \frac{\partial \ell}{\partial q^i} \right\} \mathrm{d}q^i, \tag{10}$$

where summation is performed over repeated indices. The Lagrange derivative is a natural operator which takes the same form in all coordinate systems, see [65], [60].[3]

Lemma 1.3.3. *For every smooth curve* $c : \mathbb{R} \to Q : t \mapsto q(t)$ *on* Q*, the Lagrange derivative of the kinetic energy* $k(v) = \frac{1}{2}\,\mathrm{k}(v, v)$ *along* $\dot{c}$ *is*

$$\delta k(q(t), \dot{q}(t)) = \mathrm{k}(q(t))^\flat(\nabla_{\dot{q}} \dot{q}(t)), \tag{11}$$

for all t.

[3] What we mean by this is the following. Let Q and $\widetilde{Q}$ be configuration spaces which are diffeomorphic by the diffeomorphism φ. The map φ induces a diffeomorphism $T^*\varphi$ of the cotangent bundle $T^*\widetilde{Q}$ onto the cotangent bundle T^*Q, which is defined as follows. For every $\widetilde{q} \in \widetilde{Q}$ and every $\widetilde{\alpha}_{\widetilde{q}} \in T^*_{\widetilde{q}}\widetilde{Q}$, we have $T^*\varphi(\widetilde{\alpha}_{\widetilde{q}})$ is the cotangent vector $(T_{\widetilde{q}}\varphi^{-1})^T(\widetilde{\alpha}_{\widetilde{q}})$ to Q at q. Here the superscript T denotes transpose. If $\ell : TQ \to \mathbb{R}$ and $\widetilde{\ell} : T\widetilde{Q} \to \mathbb{R}$ are Lagrangians, where $\widetilde{\ell} = \ell \circ T\varphi$, then $\delta\widetilde{\ell} = \delta\ell \circ T^*\varphi$. As Lagrange [65] observed, the preceding formula follows from the interpretation of the Lagrange derivative $\delta\ell$ as the variational derivative of $\int \ell \, \mathrm{d}t$.

Proof. We compute

$$\begin{aligned}\delta k(q(t),\dot q(t)) &= \left\{ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial k}{\partial \dot q^i} - \frac{\partial k}{\partial q^i}\right\}\mathrm{d}q^i = \left\{ \frac{\mathrm{d}}{\mathrm{d}t}(k_{ij}\dot q^j) - \tfrac{1}{2}\,\frac{\partial k_{j\ell}}{\partial q^i}\dot q^j \dot q^\ell \right\}\mathrm{d}q^i \\ &= \left\{ k_{ij}\frac{\mathrm{d}}{\mathrm{d}t}\dot q^j + \frac{\partial k_{ij}}{\partial q^\ell}\dot q^\ell \dot q^j - \tfrac{1}{2}\,\frac{\partial k_{j\ell}}{\partial q^i}\dot q^j \dot q^\ell \right\}\mathrm{d}q^i \\ &= \left\{ k_{ij}\frac{\mathrm{d}}{\mathrm{d}t}\dot q^j + \tfrac{1}{2}\left(\frac{\partial k_{jn}}{\partial q^\ell} + \frac{\partial k_{j\ell}}{\partial q^n} - \frac{\partial k_{n\ell}}{\partial q^j}\right)\dot q^n \dot q^\ell \right\}\mathrm{d}q^i \\ &= k_{ij}\left\{ \frac{\mathrm{d}}{\mathrm{d}t}\dot q^j + \Gamma^j_{n\ell}\dot q^n \dot q^\ell \right\}\mathrm{d}q^i = k_{ij}\ (\nabla_{\dot q}\dot q)^j.\end{aligned}$$

Here

$$k_{ij}\Gamma^j_{n\ell} = \tfrac{1}{2}\left(\frac{\partial k_{jn}}{\partial q^\ell} + \frac{\partial k_{j\ell}}{\partial q^n} - \frac{\partial k_{n\ell}}{\partial q^j}\right) \tag{12}$$

are the *Christoffel symbols* of the Levi-Civita connection ∇ of the kinetic energy metric k on Q. □

From (9) it follows that if $\ell = k - \tau^* V$, then

$$\begin{aligned}\delta\ell(q(t),\dot q(t)) &= \delta(k - \tau_Q^* V)(q(t),\dot q(t)) = \delta k(q(t),\dot q(t)) + \mathrm{d}V(q(t)) \\ &= \mathrm{k}^\flat(\ddot q(t)) + \mathrm{d}V(q(t)).\end{aligned}$$

Hence, we are led to the Lagrangian form of the equations of motion for a nonholonomically constrained system.

Theorem 1.3.4. (Lagrange-d'Alembert principle) *A smooth curve $c : (t_0, t_1) \to Q : t \mapsto q(t)$ is a dynamically admissible motion of a nonholonomically constrained mechanical system with constraint distribution D and Lagrangian $\ell = k - \tau_Q^* V$ if and only if for every $t \in (t_0, t_1)$*

$$\delta\ell(q(t),\dot q(t)) \in D^0_{q(t)} \ \text{ and } \ \dot q(t) \in D_{q(t)}. \tag{13}$$

Corollary 1.3.5. *The total energy function $h = k + \tau_Q^* V$ is constant on all dynamically admissible motions.*

Proof. Since $\langle \frac{\partial \ell}{\partial \dot q} \mid \dot q\rangle = \mathrm{k}(q)(\dot q, \dot q) = k(q)(\dot q)$, we may write

$$h(q,\dot q) = \langle \frac{\partial \ell}{\partial \dot q} \mid \dot q\rangle - \ell(q,\dot q). \tag{14}$$

Let $t \mapsto q(t)$ be an admissible motion. Then differentiating (14) gives

$$\begin{aligned}\frac{\mathrm{d}h(q(t),\dot q(t))}{\mathrm{d}t} &= \langle \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \ell}{\partial \dot q}\right) \mid \dot q\rangle + \langle \frac{\partial \ell}{\partial \dot q} \mid \ddot q\rangle - \langle \frac{\partial \ell}{\partial q} \mid \dot q\rangle - \langle \frac{\partial \ell}{\partial \dot q} \mid \ddot q\rangle \\ &= \langle \delta\ell(q(t),\dot q(t)) \mid \dot q(t)\rangle = 0,\end{aligned}$$

using (13). □

1.4 Lagrange derivative in a trivialization

We now show how to compute the Lagrange derivative in a trivialization of TQ.

Suppose that Q is an n-dimensional smooth manifold. We say that the tangent bundle $\tau_Q : TQ \to Q$ of Q is *trivial* if there is a smooth mapping $\kappa : TQ \to Q \times \mathbb{R}^n$ such that $\kappa|T_qQ$ is a linear isomorphism with $\mathbb{R}^n$. The map κ is called a *trivialization*. Recall that the tangent bundle of Q is locally trivial, that is, about every point of Q there is an open neighborhood U such that the bundle $\tau|\tau_Q^{-1}(U) : \tau_Q^{-1}(U) \to U$ is trivial. Let $\lambda : Q \times \mathbb{R}^n \to TQ$ be the inverse of the trivialization κ. Using the standard basis $\{e_i\}_{i=1}^n$ of $\mathbb{R}^n$, we see that

$$X_i : Q \to TQ : q \mapsto \lambda(q, e_i)$$

is a smooth vector field on Q such that $\{X_i(q)\}_{i=1}^n$ is a basis of T_qQ for every $q \in Q$. In other words, $\{X_i\}_{i=1}^n$ is a *moving frame* on Q. Conversely, suppose that $\{X_i\}_{i=1}^n$ is a moving frame on Q. For each $c \in \mathbb{R}^n$ define a vector field c^Q on Q by

$$c^Q(q) = \sum_{i=1}^n c_i\, X_i(q). \tag{15}$$

Then $X_i = e_i^Q$ and $c \mapsto c^Q$ is a linear mapping.[4] Therefore the map

$$\lambda : Q \times \mathbb{R}^n \to TQ : (q, c) \mapsto c^Q(q) = c_q^Q \tag{16}$$

is the inverse of a trivialization of the tangent bundle. In this way we see that moving frames on Q and trivializations of TQ are equivalent objects.

Let $\ell : TQ \to \mathbb{R}$ be a smooth function, called a *Lagrangian*, and let $\lambda : Q \times \mathbb{R}^n \to TQ$ be the inverse of a trivialization of TQ. Define the *Lagrangian in the trivialization* λ^{-1} to be $\mathcal{L} = \ell \circ \lambda : Q \times \mathbb{R}^n \to \mathbb{R}$. The following proposition gives a formula for the Lagrange derivative of $\mathcal{L}$.[5]

Proposition 1.4.6. *Let γ be a smooth curve on Q. Define a smooth curve $c : \mathbb{R} \to \mathbb{R}^n : t \mapsto c(t)$ by requiring that $\lambda(\gamma(t), c(t)) = \dot{\gamma}(t)$, which lies in $T_{\gamma(t)}Q$. For any $a \in \mathbb{R}^n$ we have*

[4] In §30 of chapter II Whittaker [118] calls the c_i quasi-coordinates.
[5] Formula (17) is due to Poincaré [92], see next page.

$$\langle \delta\mathcal{L}(\gamma(t), c(t)) \mid a\rangle$$
$$= \langle \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial c}(\gamma(t), c)\Big|_{c=c(t)} \mid a\rangle - \langle \frac{\partial \mathcal{L}}{\partial q}(q, c(t))\Big|_{q=\gamma(t)} \mid a^Q_{\gamma(t)}\rangle$$
$$- \langle \frac{\partial \mathcal{L}}{\partial c}(\gamma(t), c)\Big|_{c=c(t)} \mid \lambda^{-1}_{\gamma(t)}\big([c(t)^Q,\, a^Q]_{\gamma(t)}\big)\rangle. \qquad (17)$$

Here $[c(t)^Q,\, a^Q]$ *denotes the Lie bracket of the vector fields* $c(t)^Q$ *and* a^Q *on* Q*.*

Proof. We compute

$$\langle \frac{\partial \mathcal{L}}{\partial c}(q, c) \mid a\rangle = \langle \frac{\partial \ell}{\partial c}(q, c^Q_q) \mid a^Q_q\rangle = \langle \frac{\partial \ell}{\partial \dot{q}}(q, \dot{q})\Big|_{\dot{q}=c^Q_q} \mid a^Q_q\rangle. \qquad (18)$$

We now use local coordinates on Q. Substituting $q = \gamma(t)$, $c = c(t)$, $\dot{q} = \dot{\gamma}(t) = c^Q_{\gamma(t)}$ into (18), and then differentiating the result with respect to t gives

$$\langle \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial c}(\gamma(t), c)\Big|_{c=c(t)} \mid a\rangle$$
$$= \langle \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \ell}{\partial \dot{q}}(\gamma(t), \dot{q})\Big|_{\dot{q}=\dot{\gamma}(t)} \mid a^Q_{\gamma(t)}\rangle + \langle \frac{\partial \ell}{\partial \dot{q}}(\gamma(t), \dot{q})\Big|_{\dot{q}=\dot{\gamma}(t)} \mid \mathrm{D}a^Q_{\gamma(t)}\, c(t)^Q_{\gamma(t)}\rangle.$$

But

$$\frac{\partial \mathcal{L}}{\partial q}(q, c) = \frac{\partial \ell}{\partial q}(q, c^Q_q) = \frac{\partial \ell}{\partial q}(q, \dot{q})\Big|_{\dot{q}=c^Q_q} + \frac{\partial \ell}{\partial \dot{q}}(q, \dot{q})\Big|_{\dot{q}=c^Q_q} \circ \mathrm{D}c^Q_q.$$

Consequently,

$$\langle \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial c}(\gamma(t), c)\Big|_{c=c(t)} \mid a\rangle - \langle \frac{\partial \mathcal{L}}{\partial q}(q, c(t)) \mid a^Q_{\gamma(t)}\rangle = \langle \delta\ell(\gamma(t), \dot{\gamma}(t)) \mid a^Q_{\gamma(t)}\rangle$$
$$+ \langle \frac{\partial \ell}{\partial \dot{q}}(\gamma(t), \dot{q})\Big|_{\dot{q}=\dot{\gamma}(t)} \mid \Big(\mathrm{D}a^Q_{\gamma(t)}\, c(t)^Q_{\gamma(t)} - \mathrm{D}c(t)^Q_{\gamma(t)} a^Q_{\gamma(t)}\Big)\rangle.$$

This implies (17) because $\delta\mathcal{L}(\gamma(t), c(t)) = \delta\ell(\gamma(t), \dot{\gamma}(t))$ and the Lie bracket of two vector fields X and Y in local coordinates is given by $[X, Y](q) = \mathrm{D}Y(q)X(q) - \mathrm{D}X(q)Y(q)$. □

We now look at the special case of proposition 1.4.6 when Q is a Lie group G with Lie algebra $\mathfrak{g}$. The *standard left trivialization* of TG is

$$\kappa : TG \to G \times \mathfrak{g} : (g, \dot{g}) \mapsto (g, \xi = T_e L_{g^{-1}} \dot{g}). \tag{19}$$

Here $L_g : G \to G : h \mapsto gh$ is multiplication on the left by g. Another way to define ξ is to take the derivative of the curve $t \mapsto g(0)^{-1} g(t)$ at $t = 0$, where $t \mapsto g(t)$ is a smooth curve in G with $g(0) = g$ and $\dot{g}(0) = \dot{g}$. The inverse of κ is given by

$$\lambda : G \times \mathfrak{g} \to TG : (g, \xi) \mapsto T_e L_g \xi. \tag{20}$$

Corollary 1.4.7. *Using the standard left trivialization λ^{-1} of TG, the Lagrangian derivative of $\mathcal{L} = \ell \circ \lambda$ is*

$$\begin{aligned} \delta\mathcal{L}(\gamma(t), c(t)) = {}& \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}}{\partial c}(\gamma(t), c)\Big|_{c=c(t)} - \frac{\partial \mathcal{L}}{\partial c}(\gamma(t), c)\Big|_{c=c(t)} \circ \mathrm{ad}_{c(t)} \\ & - \frac{\partial \mathcal{L}}{\partial q}(q, c(t))\Big|_{q=\gamma(t)} \circ \lambda_{\gamma(t)}. \end{aligned} \tag{21}$$

Here $t \mapsto \gamma(t)$ is a smooth curve on G with the curve $c : \mathbb{R} \to \mathfrak{g} : t \mapsto c(t)$ defined by requiring that $\dot{\gamma}(t) = \lambda(\gamma(t), c(t))$. Moreover, $\mathrm{ad}_\xi : \mathfrak{g} \to \mathfrak{g} : \eta \mapsto [\xi, \eta]$ for every $\xi \in \mathfrak{g}$.

1.5 Hamilton-d'Alembert equations

In this section we derive the *Hamilton-d'Alembert* formulation of dynamics of a nonholonomically constrained system. Because the constraints are given by a distribution $D \subseteq TQ$, it is more convenient to work in *velocity space* TQ than in *momentum space* T^*Q.

The *canonical* 1*-form* θ_Q on T^*Q is defined by

$$\langle \theta_Q(p) \mid u_p \rangle = \langle p \mid T\pi_Q(u_p) \rangle \quad \text{for all } u_p \in T_p(T^*Q). \tag{22}$$

The *canonical symplectic form* on T^*Q is $\omega_Q = -\,\mathrm{d}\theta_Q$. We use the notation $\theta = (\mathrm{k}^\flat)^* \theta_Q$ and $\omega = (\mathrm{k}^\flat)^* \omega_Q$. Since $\mathrm{k}^\flat$ is a diffeomorphism, it follows that ω is a symplectic form on TQ. Moreover, $\omega = -\,\mathrm{d}\theta$.

Lemma 1.5.8. *For every $w \in T_u(TQ)$ with $u \in TQ$, we have $\langle \theta(u) \mid w \rangle = \mathrm{k}(u, T\tau_Q(w))$.*

Proof. Because $\pi_Q \circ \mathrm{k}^\flat = \tau_Q$ we get

$$\begin{aligned}\langle \theta(u) \mid w\rangle &= \langle (\mathrm{k}^\flat)^*\theta_Q(u) \mid w\rangle = \langle \theta_Q(\mathrm{k}^\flat(u)) \mid T\mathrm{k}^\flat(w)\rangle \\ &= \langle \mathrm{k}^\flat(u) \mid T\pi_Q(Tk^\flat(w))\rangle = \mathrm{k}(u, T\tau_Q(w)).\end{aligned}$$

□

Since the fiber of $\tau_Q : TQ \to Q$ is a vector space, for each $q \in Q$ and each $u \in T_qQ$, we have a vector space isomorphism $\iota_u : T_qQ \to \ker T_u\tau_Q$ such that for every $v \in T_qQ$ and every $f \in C^\infty(TQ)$, the *fiber derivative* of f in the direction $\iota_u(v)$ is

$$\iota_u(v)f = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} f(u + tv). \tag{23}$$

Lemma 1.5.9. *For every $u, v \in T_qQ$, $w \in T_u(TQ)$ we have*

$$\omega(w, \iota_u(v)) = \mathrm{k}(T\tau_Q(w), v).$$

Proof. For each $q \in Q$, $u \in T_qQ$ and $w \in T_u(TQ)$,

$$\langle \theta(u) \mid w\rangle = \langle k^\flat(u) \mid T\tau_Q(w)\rangle. \tag{24}$$

Every vector $w \in T_u(TQ)$ can be extended to a vector field on TQ which projects to a vector field on Q under τ_Q. Therefore,

$$\begin{aligned}\omega(w, \iota_u(v)) &= \langle \iota_u(v) \,\lrcorner\, \mathrm{d}\theta(u) \mid w\rangle = \iota_u(v)\langle \theta(u) \mid w\rangle \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \mathrm{k}(u + tv, T\tau_Q(w)) = k(v, T\tau_Q(w)).\end{aligned}$$

□

On the right hand sign of the first equality $\lrcorner$ denotes the left interior product (contraction). In other words, for every pair of vector fields X and X' on TQ, $\langle X \,\lrcorner\, \omega \mid X'\rangle = \omega(X, X')$.

The 2-form ω on TQ is *symplectic*. In other words, ω is nondegenerate and closed. For every function $h \in C^\infty(TQ)$, the *Hamiltonian vector field* corresponding to h is the vector field X_h on (TQ, ω) such that

$$X_h \,\lrcorner\, \omega = \mathrm{d}h. \tag{25}$$

Lemma 1.5.10. *The projection from (TQ, ω) to Q under τ_Q of an integral curve of the Hamiltonian vector field X_k of the kinetic energy $k(v) = \frac{1}{2}\mathrm{k}(v, v)$ is a geodesic of the Levi-Civita connection ∇ associated to* k.

Proof. In terms of local coordinates $\{q^i\}$ on Q and induced local coordinates $\{q^i, v^j = \dot{q}_j\}$ on TQ, we have

$$\omega = -\,\mathrm{d}(k_{ij}v^j) \wedge \mathrm{d}q^j = -k_{ij}\ \mathrm{d}v^i \wedge \mathrm{d}q^j - \frac{\partial k_{ij}}{\partial q^\ell} v^i\, \mathrm{d}q^\ell \wedge \mathrm{d}q^j$$

and

$$\mathrm{d}k = k_{ij}\, v^i\, \mathrm{d}v^j + \tfrac{1}{2}\, \frac{\partial k_{ij}}{\partial q^\ell}\, v^i v^j\ \mathrm{d}q^\ell.$$

If $X_k = \dot{q}^i \frac{\partial}{\partial q^i} + \dot{v}^i \frac{\partial}{\partial v^i}$, then

$$\begin{aligned} X_k \mathbin{\lrcorner} \omega &= -k_{ij}\dot{v}^i\ \mathrm{d}q^j + k_{ij}\dot{q}^j\ \mathrm{d}v^i - \frac{\partial k_{ij}}{\partial q^\ell} v^i \dot{q}^\ell\ \mathrm{d}q^j + \frac{\partial k_{ij}}{\partial q^\ell} v^i \dot{q}^j\ \mathrm{d}q^\ell \\ &= k_{ij}\dot{q}^j\ \mathrm{d}v^i - \left\{ k_{ij}\dot{v}^i + \frac{\partial k_{ij}}{\partial q^\ell} v^i \dot{q}^\ell - \frac{\partial k_{i\ell}}{\partial q^j} v^i \dot{q}^\ell \right\} \mathrm{d}q^j. \end{aligned}$$

Hence, $X_k \mathbin{\lrcorner} \omega = \mathrm{d}k$ implies that $\dot{q}^j = v^j$ and

$$k_{ij}\dot{v}^i + \frac{\partial k_{ij}}{\partial q^\ell} v^i \dot{q}^\ell - \frac{\partial k_{i\ell}}{\partial q^j} v^i \dot{q}^\ell = -\tfrac{1}{2}\, \frac{\partial k_{i\ell}}{\partial q^j} v^i v^\ell.$$

Therefore,

$$\begin{aligned} \dot{v}^j &= k^{im} \left\{ -\tfrac{1}{2}\, \frac{\partial k_{n\ell}}{\partial q^m} v^n v^\ell + \frac{\partial k_{n\ell}}{\partial q^m} v^n v^\ell - \frac{\partial k_{nm}}{\partial q^\ell} v^n v^\ell \right\} \\ &= k^{im} \left\{ \tfrac{1}{2}\, \frac{\partial k_{n\ell}}{\partial q^m} v^n v^\ell - \frac{\partial k_{nm}}{\partial q^\ell} v^n v^\ell \right\} = -\Gamma^i_{n\ell} v^n v^\ell, \end{aligned}$$

where $k^{im}k_{mj} = \delta_{ij}$ and Γ^i_{nk} is the Christoffel symbol (12). Consequently,

$$X_k(v) = v^i \frac{\partial}{\partial q^i} - \Gamma^i_{n\ell} v^n v^\ell \frac{\partial}{\partial v^i}. \tag{26}$$

If $t \mapsto v(t)$ is an integral curve of X_k and $t \to q(t) = \tau_Q(v(t))$ is its projection to Q, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\dot{q}^i(t) + \Gamma^i_{n\ell}(q(t))\dot{q}^n(t)\dot{q}^\ell(t) = 0.$$

Thus $t \to q(t)$ is a geodesic of the Levi-Civita connection ∇ associated to the kinetic energy metric k. □

Lemma 1.5.11. *Let $(t_0, t_1) \to Q : t \mapsto q(t)$ be a smooth integral curve of X_k and let $t \mapsto u(t) = \dot{q}(t)$ be its tangent lift to TQ, that is, $\tau_Q(u(t)) = q(t)$. Then*

$$\dot{u}(t) - X_k(\dot{q}(t)) = \iota_{\dot{q}(t)}(\nabla_{\dot{q}}\dot{q}(t)) \tag{27}$$

for all $t \in (t_0, t_1)$.

Proof. Since $u(t) = \dot{q}(t)$ it follows that $\dot{u}(t) = v^i(t)\frac{\partial}{\partial q^i} + \dot{v}^i(t)\frac{\partial}{\partial v^i}$. Equation (26) implies that

$$\begin{aligned}\dot{u}(t) - X_k(\dot{q}(t)) &= \left(\dot{v}^i + \Gamma^i_{nk} v^n v^k\right)\frac{\partial}{\partial v^i} \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}t}\dot{q}^i(t) + \Gamma^i_{nk}(q(t))\dot{q}^n(t)\dot{q}^k(t)\right)\frac{\partial}{\partial v^i} \\ &= \nabla_{\dot{q}}\dot{q}(t)\frac{\partial}{\partial v^i} = \iota_{\dot{q}(t)}(\nabla_{\dot{q}}\dot{q}(t)).\end{aligned}$$

□

The set

$$F = \{w \in T(TQ) \mid T\tau_Q(w) \in D\} \tag{28}$$

is a distribution on Q. To see this let $u \in D$. Since the map $T_u\tau_Q : T_u(TQ) \to T_{\tau_Q(u)}Q$ is surjective, $(T_u\tau_Q)^{-1}D_{\tau_Q(u)}$ is a linear subspace of $T_u(TQ)$, which has constant dimension and depends smoothly on u. Therefore $\coprod_{u\in D}(T_u\tau_Q)^{-1}D_{\tau_Q(u)}$ is a smooth vector subbundle of $T_D(TQ)$, the bundle of tangent vectors to TQ with base point in D.

Denote by F^0 the *annihilator* of F, that is, for every $u \in TQ$,

$$F^0_u = \{p \in T^*_u(TQ) \mid \langle p \mid w\rangle = 0 \quad \text{for all } w \in F_u\}. \tag{29}$$

Theorem 1.5.12. (Hamilton-d'Alembert principle) *A curve* $(t_0, t_1) \to Q : t \mapsto q(t)$ *with lift* $t \mapsto u(t) = \dot{q}(t) \in D$ *is a dynamically admissible motion corresponding to the Hamiltonian* $h = k + \tau_Q^* V$ *subject to the nonholonomic constraint* D *if and only if*

$$\dot{u}(t) \,\lrcorner\, \omega - \mathrm{d}h(\dot{q}(t)) \in F^0_{\dot{q}(t)} \tag{30}$$

for every $t \in (t_0, t_1)$.

Proof. For each $w \in F_{\dot{q}(t)}$, from lemma 1.5.11 we obtain

$$\begin{aligned}\langle \dot{u}(t) \,\lrcorner\, \omega \mid w\rangle &= \omega(\dot{u}(t), w) = \omega(X_k(\dot{q}(t)), w) + \iota_{\dot{q}(t)}(\nabla_{\dot{q}}\dot{q}), w) \\ &= \omega(X_k(\dot{q}(t)), w) + \omega(\iota_{\dot{q}(t)}(\nabla_{\dot{q}}\dot{q}), w) \\ &= \langle \mathrm{d}k \mid w\rangle - \mathrm{k}(\nabla_{\dot{q}}\dot{q}, T\tau_Q(w)) \\ &= \langle \mathrm{d}h \mid w\rangle - \langle \mathrm{d}V \mid T\tau_Q(w)\rangle - \mathrm{k}(\nabla_{\dot{q}}\dot{q}, T\tau_Q(w)) \\ &= \langle \mathrm{d}h \mid w\rangle - \langle \mathrm{k}^\flat(\nabla_{\dot{q}}\dot{q}) + \mathrm{d}V \mid T\tau_Q(w)\rangle.\end{aligned}$$

Hence, $\dot{u}(t) \,\lrcorner\, \omega - \mathrm{d}h(\dot{q}(t)) \in F^0_{\dot{q}(t)}$ if and only if $\mathrm{k}^\flat(\nabla_{\dot{q}}\dot{q}(t)) + \mathrm{d}V(q(t)) \in D^0_{q(t)}$. By hypothesis $\dot{q}(t) \in D$. The theorem follows from d'Alembert's principle 1.2.2. □

Corollary 1.5.13. *The Hamiltonian function h is constant on every admissible motion.*

Proof. Because $\dot{u}(t) \in F_{\dot{q}(t)}$ we get

$$0 = \dot{u}(t) \lrcorner (\dot{u}(t) \lrcorner \omega - \mathrm{d}h) = \omega(\dot{u}, \dot{u}) - \mathrm{d}h(\dot{u}(t)) = -\mathrm{d}h(\dot{u}(t)).$$

□

1.6 Distributional Hamiltonian formulation

In this section we give a distributional Hamiltonian formulation of the equations of motion of a nonholonomically constrained system.

1.6.1 *The symplectic distribution* (H, ϖ)

Since D is a distribution on Q, it is a submanifold of TQ. We denote by $T^\omega D$ the *symplectic annihilator* of TD. In other words, for every $u \in D$,

$$T^\omega_u D = \{w \in T_u(TQ) \mid \omega(w, v) = 0 \quad \text{for all } v \in T_u D\}. \tag{31}$$

$T^\omega D$ is a distribution on D being a vector subbundle of $T_D(TQ)$, which is the bundle of tangent vectors to TQ whose base point lies in D. Similarly, we denote the symplectic annihilator of the distribution F (28) by F^ω, that is, for every $u \in D$,

$$F^\omega_u = \{w \in T_u(TQ) \mid \omega(w, v) = 0 \quad \text{for all } v \in F_u\}. \tag{32}$$

Theorem 1.6.1.14. *The vector bundle $T_D(TQ)$ has two direct sum decompositions*

$$T_D(TQ) = F^\omega \oplus TD = F \oplus T^\omega D. \tag{33}$$

Proof. First we show that $F^\omega \cap TD = \{0\}$. Because $\ker T\tau_Q$ is a *Lagrangian distribution* on (TQ, ω), that is, $\omega | \ker T\tau_Q = 0$ and $\dim \ker T\tau_Q = \frac{1}{2} \dim T(TQ)$, and $\ker T\tau_Q \subseteq F$, it follows that $F^\omega \subseteq \ker T\tau_Q$. Hence,

$$F^\omega \cap TD = F^\omega \cap (TD \cap \ker T\tau_Q).$$

Let $u \in D$ and $w \in T_u D \cap F^\omega_u$. Then $w \in TD \cap \ker T\tau_Q$ and is of the form $w = \iota_u(v)$ for some $v \in D_{\tau_Q(u)}$. Moreover, $w \in F^\omega_u$ which implies that $\omega(\iota_u(v), w') = \omega(w, w') = 0$ for every $w' \in F_u$. Using lemma 1.5.9 we get $\mathrm{k}(v, T\tau_Q(w')) = 0$ for all $w' \in F_u$. Since $T\tau_Q(F) = D$, we can choose w' so that $T\tau_Q(w') = v$. Thus $\mathrm{k}(v, v) = 0$, which implies that $v = 0$ and $w = \iota_u(v) = 0$. Therefore, $F^\omega \cap TD = \{0\}$.

If d is the *rank* of the distribution D and $n = \dim Q$, we have $\dim F_u = n + d$, $\dim F_u^\omega = 2n - \dim F_u = n - d$, and $\dim T_u D = n + d$. Hence, $\dim F_u^\omega + \dim T_u D = 2n = \dim T_p(TQ)$. Since $F_u^\omega \cap T_u D = \{0\}$, it follows that $T_u(TQ) = F_u^\omega \oplus T_u D$. Taking symplectic annihilators we get $T_u(TQ) = F_u \oplus T_u^\omega D$. □

Let H be the intersection of F (28) and the tangent bundle of D, that is,

$$H = F \cap TD. \tag{34}$$

H is a distribution on D.

Proposition 1.6.1.15. *For each $u \in D$,*

$$H_u \cap \ker T\tau_Q = \{\iota_u(v) \mid v \in D_{\tau_Q(u)}\}.$$

Proof. For every $u, v \in D_q$, $u + sv \in D_q$ for all $s \in \mathbb{R}$. Hence, $\iota_u(v) \in TD$. Since $\iota_u(v) \in \ker T\tau_Q \subseteq F$, it follows that $\iota_u(v) \in H_u$. Consequently,

$$\{\iota_u(v) \mid v \in D_{\tau_Q(u)}\} \subseteq H_u \cap \ker T\tau_Q.$$

Conversely, suppose that $w \in H_u \cap \ker T\tau_Q$. Then $T\tau_Q(w) = 0$ which implies that there exists $v \in T_{\tau_Q(u)}Q$ such that $w = \iota_u(v)$. If $v \notin D_{\tau_Q(u)}$, then there exists a 1-form η on Q with values in D^0 such that $\langle \eta \mid v \rangle \neq 0$. The hypothesis that η has values in D^0 implies that the function $f : TQ \to \mathbb{R} : v' \mapsto \langle \eta \mid v' \rangle$ vanishes on D. But

$$\langle \mathrm{d}f \mid w \rangle = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} \langle \eta \mid u + sv \rangle = \langle \eta \mid v \rangle \neq 0$$

which contradicts the assumption that $w = \iota_u(v) \in H \subseteq TD$. □

For each $u \in D$ write

$$\varpi_u = \omega_u \Big|_{H_u \times H_u}. \tag{35}$$

Corollary 1.6.1.16. *(H, ϖ) is a* symplectic distribution *on D, that is, for every $u \in D$ the 2-form ϖ_u is nondegenerate.*

Proof. Let $\{v_1, ..., v_d\}$ be a k-orthonormal basis of D_q. For each $u \in D_q$, let $w_i \in T_u D$ be a lift of v_i, that is $T\tau_Q(w_i) = v_i$. Then $\{w_1, ..., w_d, \iota_u(v_1), ..., \iota_u(v_d)\}$ is a basis of H_u.

Suppose that $w = \sum_i (a^i w_i + b^i \iota_u(v_i))$ is in the kernel of the restriction of ω to H_u, that is, $\omega(w, w') = 0$ for all $w' \in H_u$. Taking $w' = \iota_u(v_i)$, from lemma 1.5.9 we obtain

$$0 = \omega(w, \iota_u(v_i)) = \mathrm{k}(v_i, T\tau_Q(\sum_j (a^j w_j + b^j \iota_u(v_j)))) = \mathrm{k}(v_i, \sum_j a^j v_j) = a_i.$$

Hence, $w = \sum_i b^i \iota_u(v_i)$. Similarly, taking $w' = w_i$, we get

$$0 = \omega(w, w_i) = \omega(\sum_j b^j \iota_u(v_j), w_i) = \sum_j b^j \omega(\iota_u(v_j), w_i)$$
$$= -\sum_j b^j \mathrm{k}(v_j, T\tau_Q(w_i)) = -\sum_j b^j \mathrm{k}(v_j, v_i) = -b_i.$$

Therefore $w = 0$, that is, the restriction of ω to H_u is nondegenerate. In other words, (H, ϖ) is a symplectic distribution on D. □

In order to get a clearer idea of the distributions D, F, and H defined above, we express them in local coordinates. Let $q = \{q^i\}$ be local coordinates on Q. Then $(q, v = \dot{q}) = \{q^i, v^i = \dot{q}^i\}$ and $(q, v, \dot{q}, \dot{v}) = \{q^i, v^i, \dot{q}^i, \dot{v}^i\}$ are induced local coordinates on TQ and $T(TQ)$, respectively. The tangent bundle projection $\tau_Q : TQ \to Q$ is $\tau_Q(q, v) = q$ and its tangent $T\tau_Q : T(TQ) \to TQ$ is $T\tau_Q(q, v, \dot{q}, \dot{v}) = (q, \dot{q})$.

We have $(q, v) \in D_q$ if and only if locally on Q there are ℓ linearly independent 1-forms $\{\alpha_j\}_{j=1}^{\ell}$ such that $\langle \alpha_j(q)|v\rangle = 0$ for every $1 \le j \le \ell$. From the definition (28) of the distribution F we see that

$$F_{(q,v)} = \{(\dot{q}, \dot{v}) \in T_{(q,v)}(TQ) \big| (q, \dot{q}) \in D_q \subseteq T_qQ\}.$$

Consequently, the distribution H (34) is the set of $(\dot{q}, \dot{v}) \in T_{(q,v)}(TQ)$ such that

$$\begin{aligned} 0 &= \langle \alpha_j(q)|\dot{q}\rangle = \langle (\alpha_j(q), 0)|(\dot{q}, \dot{v})\rangle \\ 0 &= \langle \alpha_j(q)|\dot{v}\rangle + \langle D\alpha_j(q)\dot{q}|v\rangle = \langle ((D\alpha_j(q))^t v, \alpha_j(q))|(\dot{q}, \dot{v})\rangle \end{aligned} \tag{36}$$

for $j = 1, \ldots, \ell$, where $D\alpha_j(q)$ denotes the derivative of α_j at q. Clearly, $H_{(q,v)}$ is a linear subspace of $T_{(q,v)}(TQ)$ of codimension 2ℓ and $\coprod_{(q,v)\in D} H_{(q,v)}$ is a smooth vector subbundle of $T_D(TQ)$. In local coordinates the symplectic form ω on TQ is

$$\omega_{(q,v)}((\dot{q}, \dot{v}), (\dot{q}', \dot{v}')) = \langle \mathrm{k}^\flat(\dot{v}')|\dot{q}\rangle - \langle \mathrm{k}^\flat(\dot{v})|\dot{q}'\rangle = \langle (-\mathrm{k}^\flat(\dot{v}), \mathrm{k}^\flat(\dot{q}))|(\dot{q}', \dot{v}')\rangle,$$

where k is a Riemannian metric on Q.

We now give an alternative proof of corollary 1.6.1.16.

Proof of corollary 1.6.1.16. Suppose that $(\dot{q}, \dot{v}) \in H_{(q,v)}$ and that for every $(\dot{q}', \dot{v}') \in H_{(q,v)}$ we have $0 = \omega(q, v)((\dot{q}, \dot{v}), (\dot{q}', \dot{v}'))$. From (36) and the local expression for the symplectic form ω, it follows that for $1 \le j \le \ell$ there are $(\lambda_j, \mu_j) \in \mathbb{R}^2$ such that

$$(-\mathrm{k}^\flat(\dot{v}), \mathrm{k}^\flat(\dot{q})) = \sum_{j=1}^{\ell} \left[\lambda_j(\alpha_j(q), 0) + \mu_j((D\alpha_j(q))^t v, \alpha_j(q))\right],$$

that is,

$$\mathrm{k}^\flat(\dot{q}) = \sum_{j=1}^{\ell} \mu_j \, \alpha_j(q) \;\text{ and }\; -\mathrm{k}^\flat(\dot{v}) = \sum_{j=1}^{\ell} \big[\lambda_j \, \alpha_j(q) + \mu_j \, (D\alpha_j(q))^t\big]. \quad (37)$$

Since $(\dot{q}, \dot{v}) \in H_{(q,v)}$, the first equation in (36) reads

$$0 = \langle \alpha_i(q) | \mathrm{k}^\sharp(\sum_{j=1}^{\ell} \mu_j \, \alpha_j(q))\rangle = \sum_{j=1}^{\ell} \mu_j \, \langle \alpha_i(q) | \mathrm{k}^\sharp(\alpha_j(q))\rangle \quad (38)$$

for every $1 \le i \le \ell$, using the first equation in (37). But k is a nondegenerate symmetric bilinear form. Thus its $\ell \times \ell$ matrix $(\langle \alpha_i(q)|\mathrm{k}^\sharp(\alpha_j(q))\rangle)$ is invertible. Hence (38) implies that $\mu_j = 0$ for every $1 \le j \le \ell$. Since $(\dot{q}, \dot{v}) \in H_{(q,v)}$, using the second equation in (37) and the fact that $\mu_j = 0$ for every $1 \le j \le \ell$, the second equation in (36) reads

$$0 = -\sum_{j=1}^{\ell} \lambda_i \, \langle \alpha_i(q) | \mathrm{k}^\sharp(\alpha_j(q))\rangle, \quad 1 \le i \le \ell.$$

Since k is nondegenerate, we obtain $\lambda_j = 0$ for every $1 \le j \le \ell$. Thus $(\dot{q}, \dot{v}) = (0, 0)$, using (37). □

1.6.2 *H and ϖ in a trivialization*

In this subsection we determine the distribution H and the symplectic form ϖ in a trivialization.

First we define a trivialization of the constraint distribution D on Q. Suppose that $\{X_i\}_{i=1}^d$ is a moving frame on the distribution D. Then

$$\lambda_D : Q \times \mathbb{R}^d \to D \subseteq TQ : (q, c = (c_1, \ldots c_d)) \mapsto \big(q, \sum_{i=1}^{d} c_i \, X_i(q)\big) \quad (39)$$

is the inverse of a trivialization of the distribution D, thought of as a vector subbundle of TQ whose fibers have dimension d.

Second we determine what the distribution H on D is in the trivialization λ_D^{-1}. From (34) we see that the vector subbundle H of TD is defined by $H_u = T_uD \cap (T_u\tau_Q)^{-1}(D_q)$ for every $u \in D$. Here $q = \tau_Q(u)$, where $\tau_Q : TQ \to Q$ is the tangent bundle projection map. Let

$$\pi_1 = \tau_Q \circ \lambda_D : Q \times \mathbb{R}^d \to Q : (q, c) \mapsto q. \quad (40)$$

Then

$$T_{(q,c)}\pi_1 = T_u\tau_Q \circ T_{(q,c)}\lambda_D : T_qQ \times T_c\mathbb{R}^d \to T_qQ : (v_q, \xi) \mapsto v_q,$$

where $u = \lambda_D(q,c)$. From the fact that λ_D is a diffeomorphism, we obtain

$$\begin{aligned}(T_{(q,c)}\lambda_D)^{-1}H_u &= (T_{(q,c)}\lambda_D)^{-1}\big(T_uD \cap (T_u\tau_Q)^{-1}(D_q)\big)\\ &= T_{(q,c)}(Q\times\mathbb{R}^d) \cap (T_{(q,c)}\pi_1)^{-1}(D_q)\\ &= D_q \times T_c\mathbb{R}^d = D_q\times\mathbb{R}^d. \qquad (41)\end{aligned}$$

Thus the distribution H in the trivialization λ_D^{-1} is the distribution $H_{Q\times\mathbb{R}^d}$ defined by

$$(H_{Q\times\mathbb{R}^d})_{(q,c)} = D_q\times\mathbb{R}^d. \qquad (42)$$

Finally we determine what the symplectic form ϖ on D restricted to the distribution H is in the trivialization λ_D^{-1}. Let k be a Riemannian metric on TQ. Then on TQ we have the symplectic form $\omega = (\mathrm{k}^\flat)^*\omega_Q$, where ω_Q is the canonical 2-form $\omega_Q = -\,\mathrm{d}\theta_Q$ on T^*Q and θ_Q is the canonical 1-form, see (22). Pulling ω back by λ_D gives the 2-form $\mu^*\omega_Q$ on $Q\times\mathbb{R}^d$ where

$$\mu = \mathrm{k}^\flat\circ\lambda_D : Q\times\mathbb{R}^d \to T^*Q : (q, c=(c_1,\ldots,c_d)) \mapsto (q, \sum_{i=1}^d c_i\,\alpha_i(q)) \qquad (43)$$

and $\alpha_i(q) = \mathrm{k}^\flat(q)(X_i(q))$. In other words, the 1-forms $\{\alpha_i\}_{i=1}^d$ define a moving coframe on the vector subbundle $D^* = \mathrm{k}^\flat(D)$ of T^*Q.

To determine the 2-form $\omega_{Q\times\mathbb{R}^d} = \mu^*\omega_Q$ on $Q\times\mathbb{R}^d$ we compute

$$\begin{aligned}\mu^*\theta_Q(q,c) &= \theta_Q(q,c)\,T_{(q,c)}\mu\\ &= \sum_{i=1}^d c_i\,\alpha_i(q)\circ T_{\mu(q,c)}\pi_Q\circ T_{(q,c)}\mu, \quad \text{using (43)}\\ &= \sum_{i=1}^d c_i\,\alpha_i(q)\circ T_{(q,c)}\pi_1 = \sum_{i=1}^d c_i\,\pi_1^*\alpha_i(q),\end{aligned}$$

where we view c_i as a function on $Q\times\mathbb{R}^d$. Therefore we get

$$\omega_{Q\times\mathbb{R}^d} = \mu^*(-\,\mathrm{d}\theta_Q) = -\,\mathrm{d}(\mu^*\theta_Q) = -\sum_{i=1}^d\big(\mathrm{d}c_i\wedge\pi_1^*\alpha_i + c_i\,\pi_1^*(\mathrm{d}\alpha_i)\big). \qquad (44)$$

To give an explicit expression for $\omega_{Q\times\mathbb{R}^d}$ restricted to $H_{Q\times\mathbb{R}^d}$, we argue as follows. For $a = (a_1,\ldots,a_d)\in\mathbb{R}^d$ let $a^Q(q) = \sum_{i=1}^d a_i\,X_i(q)$, where $X_i(q) = \lambda_D(q,e_i)$. Then a^Q is a vector field on Q with values in D. For every $a\in\mathbb{R}^d$ define two *vector fields* $a^Q_{\rightarrow}$ and $a^Q_{\uparrow}$ on Q with values in $H_{Q\times\mathbb{R}^d}\subseteq D\times\mathbb{R}^d$ by

$$a^Q_{\rightarrow}(q) = (a^Q(q),0) \text{ and } a^Q_{\uparrow}(q) = (0,a). \qquad (45)$$

From the definition of $a^Q_{\rightarrow}$, $a^Q_{\uparrow}$, and the mapping π_1 it follows that

$$T_{(q,c)}\pi_1 \, a^Q_{\rightarrow}(q) = a^Q(q) \;\text{ and }\; T_{(q,c)}\pi_1 \, a^Q_{\uparrow}(q) = 0. \tag{46}$$

Moreover,

$$\langle \mathrm{d}c_i(q,c) \mid a^Q_{\rightarrow}(q)\rangle = 0 \;\text{ and }\; \langle \mathrm{d}c_i(q,c) \mid a^Q_{\uparrow}(q)\rangle = a_i. \tag{47}$$

Also

$$\begin{aligned}\langle \pi_1^*\alpha_i(q) \mid a^Q_{\rightarrow}(q)\rangle &= \langle \alpha_i(q) \mid T_{(q,c)}\pi_1 \, a^Q_{\rightarrow}(q)\rangle = \langle \alpha_i(q) \mid a^Q(q)\rangle, \quad \text{using (46)} \\ &= \mathrm{k}(q)(X_i(q), a^Q(q)), \quad \text{by definition of } \alpha_i.\end{aligned}$$

We are now in position to calculate $\varpi_{Q\times\mathbb{R}^d}$, which is the restriction of $\omega_{Q\times\mathbb{R}^d}$ to $H_{Q\times\mathbb{R}^d} \times H_{Q\times\mathbb{R}^d}$. Using (44) we find that for every $a, b, \widetilde{a}, \widetilde{b} \in \mathbb{R}^d$ we have

$$\begin{aligned}
-\varpi_{Q\times\mathbb{R}^d}&(q,c)\big(a^Q_{\rightarrow}(q) + b^Q_{\uparrow}(q), \widetilde{a}^Q_{\rightarrow}(q) + \widetilde{b}^Q_{\uparrow}(q)\big) \\
&= \sum_{i=1}^{d} (\mathrm{d}c_i \wedge \pi_1^*\alpha_i)(q,c)\big(a^Q_{\rightarrow}(q) + b^Q_{\uparrow}(q), \widetilde{a}^Q_{\rightarrow}(q) + \widetilde{b}^Q_{\uparrow}(q)\big) \\
&\qquad + \sum_{i=1}^{d} c_i \, \pi_1^*(\mathrm{d}\alpha_i)(q,c)\big(a^Q_{\rightarrow}(q) + b^Q_{\uparrow}(q), \widetilde{a}^Q_{\rightarrow}(q) + \widetilde{b}^Q_{\uparrow}(q)\big) \\
&= -\sum_{i=1}^{d} \big(\langle \mathrm{d}c_i(q,c) \mid \widetilde{b}^Q_{\uparrow}(q)\rangle \, \langle \pi_1^*\alpha_i(q) \mid a^Q_{\rightarrow}(q)\rangle \\
&\qquad\qquad + \langle \mathrm{d}c_i(q,c) \mid b^Q_{\uparrow}(q)\rangle \, \langle \pi_1^*\alpha_i(q) \mid \widetilde{a}^Q_{\rightarrow}(q)\rangle\big) \\
&\qquad\qquad\qquad + \sum_{i=1}^{d} c_i \, \pi_1^*(\mathrm{d}\alpha_i)(q)(a^Q_{\rightarrow}(q), \widetilde{a}^Q_{\uparrow}(q)), \quad \text{using (47)} \\
&= -\sum_{i=1}^{d} \widetilde{b}_i \, \mathrm{k}(q)(X_i(q), a^Q(q)) + \sum_{i=1}^{d} b_i \, \mathrm{k}(q)(X_i(q), \widetilde{a}^Q(q)) \\
&\qquad\qquad + \sum_{i=1}^{d} c_i \, \mathrm{d}\alpha_i(q)(a^Q(q), \widetilde{a}^Q(q)).
\end{aligned}$$

So the symplectic form ϖ on D restricted to the distribution H on D in the trivialization λ_D^{-1} is

$$\begin{aligned}
\varpi_{Q\times\mathbb{R}^d}&(q,c)\big(a^Q_{\rightarrow}(q) + b^Q_{\uparrow}(q), \widetilde{a}^Q_{\rightarrow}(q) + \widetilde{b}^Q_{\uparrow}(q)\big) \\
&= \mathrm{k}(q)\big(a^Q(q), \widetilde{b}^Q(q)\big) - \mathrm{k}(q)\big(\widetilde{a}^Q(q), b^Q(q)\big) \\
&\qquad - \sum_{i=1}^{d} c_i \; \mathrm{d}\alpha_i(q)\big(a^Q(q), \widetilde{a}^Q(q)\big),
\end{aligned} \tag{48}$$

where $\alpha_i(q) = \mathrm{k}^\flat(q)(X_i(q))$. Therefore $\mathrm{d}\alpha_i(q) = \mathrm{d}(\mathrm{k}^\flat(X_i))(q)$, which implies

$$\sum_{i=1}^{d} c_i \, \mathrm{d}\alpha_i(q) = \sum_{i=1}^{d} c_i \, \mathrm{d}(\mathrm{k}^\flat(X_i))(q)$$
$$= \mathrm{d}\mathrm{k}^\flat(\sum_{i=1}^{d} c_i \, X_i)(q) = \mathrm{d}(\mathrm{k}^\flat(c^Q))(q). \tag{49}$$

In other words, with respect to the basis $\{((X_i(q),0),(0,e_i))\}_{1\le i\le d}$ of the space $(H_{Q\times\mathbb{R}^d})_{(q,c)}$, the matrix of $\varpi_{Q\times\mathbb{R}^d}(q,c)$ is $\left(\begin{smallmatrix} B & A^t \\ -A & 0\end{smallmatrix}\right)$. Here $A = A(q)$ is the $d\times d$ positive definite symmetric matrix[6] $(A_{jk}) = \big(\mathrm{k}(q)(X_j(q),X_k(q))\big)$ and $B = B(q)$ is the $d\times d$ antisymmetric matrix $(B_{\ell,m})$, where

$$B_{\ell m} = \sum_{i=1}^{d} c_i \, \mathrm{d}\alpha_i(q)(X_\ell(q),X_m(q)) = L_{X_\ell}\big(\mathrm{k}(c^Q,X_m)\big)(q)$$
$$- L_{X_m}\big(\mathrm{k}(c^Q,X_\ell)\big)(q) - \mathrm{k}(q)\big(c^Q(q),[X_\ell,X_m](q)\big). \tag{50}$$

To obtain the last equality in (50), we have used the fact that for a 1-form β its exterior derivative evaluated on the vector fields X and Y is

$$\mathrm{d}\beta(X,Y) = L_X(Y \,\lrcorner\, \beta) - L_Y(X \,\lrcorner\, \beta) - \beta([X,Y]).$$

Using (49) we can rewrite (50) as

$$\mathrm{d}(\mathrm{k}^\flat(c^Q))(q)(X_\ell,X_m) = L_{X_\ell}(X_m \,\lrcorner\, \mathrm{k}^\flat(c^Q))(q) - L_{X_m}(X_\ell \,\lrcorner\, \mathrm{k}^\flat(c^q))(q)$$
$$- ([X_\ell,X_m] \,\lrcorner\, \mathrm{d}(\mathrm{k}^\flat(c^Q)))(q),$$

which is bilinear in $X_\ell(q)$ and $X_m(q)$. Thus we obtain

$$\sum_{i=1}^{d} c_i \, \mathrm{d}\alpha_i\big(a^Q(q),\widetilde{a}^Q(q)\big) = \mathrm{d}(\mathrm{k}^\flat(c^Q))\big(a^Q(q),\widetilde{a}^Q(q)\big)$$
$$= L_{a^Q}(\widetilde{a}^Q \,\lrcorner\, \mathrm{k}^\flat(c^Q))(q) - L_{\widetilde{a}^Q}(a^Q \,\lrcorner\, \mathrm{k}^\flat(c^Q))(q)$$
$$- ([a^Q,\widetilde{a}^Q] \,\lrcorner\, \mathrm{k}^\flat(c^Q))(q). \tag{51}$$

Consequently

$$\varpi_{Q\times\mathbb{R}^d}(q,c)\big(a^Q_{\rightarrow}(q) + b^Q_{\uparrow}(q), \widetilde{a}^Q_{\rightarrow}(q) + \widetilde{b}^Q_{\uparrow}(q)\big)$$
$$= \mathrm{k}(q)\big(a^Q(q),\widetilde{b}^Q(q)\big) - \mathrm{k}(q)\big(\widetilde{a}^Q(q),b^Q(q)\big)$$
$$- \mathrm{d}(\mathrm{k}^\flat(c^Q))\big(a^Q(q),\widetilde{a}^Q(q)\big) \tag{52}$$
$$= \mathrm{k}(q)\big(a^Q(q),\widetilde{b}^Q(q)\big) - \mathrm{k}(q)\big(\widetilde{a}^Q(q),b^Q(q)\big)$$
$$- L_{a^Q}(\widetilde{a}^Q \,\lrcorner\, \mathrm{k}^\flat(c^Q))(q) + L_{\widetilde{a}^Q}(a^Q \,\lrcorner\, \mathrm{k}^\flat(c^Q))(q)$$
$$+ ([a^Q,\widetilde{a}^Q] \,\lrcorner\, \mathrm{k}^\flat(c^Q))(q). \tag{53}$$

[6]Note that $A = I$ if and only if $\{X_i(q)\}_{i=1}^d$ is a k-orthogonal basis of D_q.

Lemma 1.6.2.17. *For any $u \in D$ the form ϖ_u is a symplectic form on H_u.*

Proof. We have to show that $\varpi_{Q\times\mathbb{R}_d}(q,c)$ is nondegenerate for every $(q,c) \in Q \times \mathbb{R}^d$. This holds if and only if the matrix $\left(\begin{smallmatrix} B & A^t \\ -A & 0 \end{smallmatrix}\right)$ is invertible, that is, if and only if A is invertible. But A is invertible because it is positive definite. □

1.6.3 *Distributional Hamiltonian vector field*

We return to developing the general theory of distributional Hamiltonian systems.

Let H^ω be the symplectic annihilator of H in $(T(TQ), \omega)$, that is, for each $u \in D$,

$$H_u^\omega = \{w \in T_u(TQ) \mid \omega(w,v) = 0 \quad \text{for all } v \in H_u\}. \tag{54}$$

Since H is symplectic, it follows that H^ω is also symplectic. The bundle $T_D(TQ)$ has a direct sum decomposition

$$T_D(TQ) = H \oplus H^\omega. \tag{55}$$

Let ϖ be the restriction of the symplectic form ω on TQ to $H \times H$ and ϖ^ω the restriction of ω to $H^\omega \times H^\omega$. Both of these 2-forms are nondegenerate. Moreover,

$$\omega | T_D(TQ) = \varpi \oplus \varpi^\omega. \tag{56}$$

Similarly, for every $f \in C^\infty(TQ)$, we denote by $\partial_H f$ and $\partial_{H^\omega} f$ the restrictions of $\mathrm{d}f$ to H and H^ω, respectively. Thus we obtain the following decomposition

$$\mathrm{d}f | T_D(TQ) = \partial_H f \oplus \partial_{H^\omega} f. \tag{57}$$

Lemma 1.6.3.18. *$F \cap H^\omega = F^\omega$. Moreover, for every $u \in D$ each $w \in F_u^\omega$ is of the form $w = i_v(u)$, where v is k-orthogonal to D and $\langle \mathrm{d}f \mid w\rangle = 0$ for each $f \in C^\infty(TQ)$.*

Proof. Since $H = F \cap TD$, it follows that $H^\omega = F^\omega + T^\omega D$. From theorem 1.6.1.14, it follows that $F \cap T^\omega D = \{0\}$. Hence,

$$F \cap H^\omega = F \cap (F^\omega + T^\omega D) = F \cap F^\omega + F \cap T^\omega D = F \cap F^\omega = F^\omega \subseteq \ker T\pi_Q.$$

Therefore, each $w \in F_u \cap H_u^\omega$ is of the form $w = \iota_u(v)$ for some $v \in T_u Q$. Since $w \in H_u^\omega$, taking lemma 1.5.9 into account we get

$$0 = \omega(w', w) = \omega(w', \iota_u(v)) = k(v, T\tau_Q(w'))$$

for every $w' \in H_u$. However, $T\tau_Q(H_u) = D_{\tau_Q(u)}$. Therefore v is k-orthogonal to D. This implies that

$$\langle \mathrm{d}f \mid w\rangle = \langle \mathrm{d}f \mid \iota_u(v)\rangle = \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} \big(\tfrac{1}{2}\,\mathrm{k}(u+sv, u+sv) + V(\tau_Q(u))\big) \quad (58)$$

$$= \mathrm{k}(u,v) = 0, \quad \text{because } u \in D.$$

□

We now define *the distributional Hamiltonian vector field of* $h \in C^\infty(D)$ with respect to the *symplectic distribution* (H, ϖ) on D to be the unique *vector field* Y_h on D *with values in* H such that

$$Y_h \,\lrcorner\, \varpi = \partial_H h. \quad (59)$$

Since the skew symmetric bilinear form ϖ on $H \subseteq TD$ is nondegenerate, it follows that Y_h is well defined for every $h \in C^\infty(D)$. Moreover, equation (33) and lemma 1.6.3.18 imply that for every extension $\widetilde{h} \in C^\infty(TQ)$ of $h \in C^\infty(D)$ the H-component of the restriction of the Hamiltonian vector field $X_{\widetilde{h}}$ on (TQ, ω) coincides with Y_h. If the symplectic distribution (H, ϖ) on D is understood, we shall refer to Y_h as the distributional Hamiltonian vector field of h.

Lemma 1.6.3.19. *Let $t \to u(t)$ be an integral curve of the distributional Hamiltonian vector field Y_h of h and $t \to q(t) = \tau_Q(u(t))$. Then $u(t) = \dot{q}(t)$ for all t.*

Proof. By proposition 1.6.1.15, $H_u \cap \ker T\tau_Q = \{\iota_u(v) \mid v \in D_{\tau_Q(u)}\}$. Hence lemma 1.5.9 ensures that $\omega(w, \iota_u(v)) = \mathrm{k}(v, T\tau_Q(w))$ for each $u \in D$ and each $v \in D_{\tau_Q(u)}$. Now

$$0 = \langle Y_h \,\lrcorner\, \varpi - \partial_H h \mid \iota_u(v)\rangle = \omega(Y_h(u), \iota_u(v)) - \langle \mathrm{d}h \mid \iota_u(v)\rangle$$

$$= \mathrm{k}(v, T\tau_Q(Y_h(u))) - \mathrm{k}(v,u) = \mathrm{k}(v, T\tau_Q(Y_h(u)) - u).$$

Thus $T\tau_Q(Y_h(u)) - u$ is k-orthogonal to every vector $v \in D_{\tau_Q(u)}$. But, $T\tau_Q(Y_h(u)) - u \in D_{\tau_Q(u)}$, which gives $T\tau_Q(Y_h(u)) - u = 0$. In other words, $T\tau_Q(Y_h(u)) = u$ for all $u \in D$. This implies that every integral curve $t \mapsto u(t)$ of Y_h is the lift of its projection $t \mapsto q(t)$ to Q by τ_Q, that is, $u(t) = \dot{q}(t)$ for all t. □

Theorem 1.6.3.20. *A curve $t \mapsto q(t)$ in Q is a dynamically admissible motion of a system on (TQ, ω) with Hamiltonian $h = k + \tau_Q^* V$ subject to the linear nonholonomic constraint D if and only if it is the projection to Q under τ_Q of an integral curve $t \mapsto u(t)$ of the distributional Hamiltonian vector field Y_h of h.*

Proof. Suppose that $t \mapsto q(t)$ in Q is a dynamically admissible motion of a system with Hamiltonian $h = k + \tau_Q^* V$ and nonholonomic constraint D. Let $u(t) = \dot{q}(t)$. Since $\dot{q}(t) \in D_{q(t)}$ it follows that $\dot{u}(t) \in TD$. Moreover, the motion $t \mapsto q(t) = \tau_Q(u(t))$ satisfies the constraint $\dot{q} \in D$. But $\dot{q}(t) = T\tau_Q(\dot{u}(t))$ which implies that $\dot{u}(t) \in F$. Hence $\dot{u}(t) \in H$.

Each $w \in F_{\dot{q}(t)}$ can be decomposed as $w = w_H + w_{H^\omega}$ with $w_H \in H$ and $w_{H^\omega} \in H^\omega$. Now

$$\begin{aligned}\langle \dot{u}(t) \lrcorner\, \omega - \mathrm{d}h \mid w\rangle &= \langle \dot{u}(t) \lrcorner\, \omega - \mathrm{d}h \mid w_H + w_{H^\omega}\rangle \\ &= \langle \dot{u}(t) \lrcorner\, \varpi - \partial_H h \mid w_H\rangle - \langle \partial_{H^\omega} h \mid w_{H^\omega}\rangle, \quad (60)\end{aligned}$$

since $\dot{u}(t) \in H_{\dot{q}(t)}$. By the Hamilton-d'Alembert principle 1.5.12, $\dot{u}(t) \lrcorner\, \omega - \mathrm{d}h(\dot{q}(t)) \in F^0_{\dot{q}(t)}$. Because $H \subseteq F$, we obtain

$$\dot{u}(t) \lrcorner\, \varpi = \partial_H h, \tag{61}$$

and

$$\langle \mathrm{d}h \mid u\rangle = 0 \quad \text{for all } u \in F \cap H^\omega, \tag{62}$$

using (60). Equation (61) just says that $t \to u(t) = \dot{q}(t)$ is an integral curve of the distributional Hamiltonian vector field Y_h of h.

Conversely, if $t \to u(t)$ is an integral curve of the distributional Hamiltonian vector field Y_h of h, then equation (61) is satisfied. Moreover, by lemma 1.6.3.18, equation (62) is also satisfied. Hence, equation (60) implies that $\dot{u}(t) \lrcorner\, \omega - \mathrm{d}h \in F^0_{u(t)}$. In addition, if $t \to q(t) = \tau_Q(u(t))$ is the projection of $t \to u(t)$ to Q, then $\dot{q}(t) = u(t)$ for all t by lemma 1.6.3.19. Hence, the Hamilton-d'Alembert principle 1.5.12 ensures that the curve $t \to q(t)$ is a dynamically admissible motion. □

An equation for integral curves $t \to u(t)$ of the distributional Hamiltonian vector field Y_h of h can be written in the form

$$\dot{u}(t) \lrcorner\, \varpi = \partial_H h(u(t)). \tag{63}$$

We shall refer to (63) as the *distributional form* of Hamilton's equations of motion. We shall refer to the quadruple (D, H, ϖ, h), where D is a manifold, (H, ϖ) is a symplectic distribution on D and $h \in C^\infty(D)$ as a *distributional Hamiltonian system.* The vector field Y_h is determined by the restriction of $\mathrm{d}h$ to the distribution H, which implies that it is determined by the restriction of h to the constraint distribution D. However, this last statement uses the symplectic form ϖ on H, which is *not* determined by the restriction of the Lagrangian ℓ, the kinetic energy, or the total energy to D.

We now give an alternative proof of theorem 1.6.3.20 using the local coordinates that we introduced in proving corollary 1.6.1.16.

Proof of theorem 1.6.3.20. Suppose that the curve $t \mapsto q(t)$ is an admissible motion of the nonholonomic system on (TQ, ω) with Hamiltonain h and constraint distribution D. Recall that in our local coordinates

1. (q,v) lies in D if and only if locally on Q there are linearly independent 1-forms $\{\alpha_j\}_{j=1}^{\ell}$ such that $\langle \alpha_j(q)|v\rangle = 0$ for $1 \le j \le \ell$.
2. $(\dot{q}, \dot{v})$ lies in $F_{(q,v)}$ if and only if $T\tau_Q(q,v,\dot{q},\dot{v}) = (q,\dot{q}) \in D_q$.
3. $(\dot{q}, \dot{v})$ lies in $H_{(q,v)}$ if and only if $\langle \alpha_j(q)|v\rangle = 0$ and $\langle \alpha_j(q)|\dot{v}\rangle + \langle D\alpha_j(q)\dot{q}|v\rangle = 0$ for $1 \le j \le \ell$.

Because $t \mapsto q(t)$ is an admissible motion, the curve $t \mapsto u(t) = (q(t), \dot{q}(t)) \in D$. Since

$$T\tau_Q(u(t), \dot{u}(t)) = T\tau_Q(q(t), \dot{q}(t), \dot{q}(t), \ddot{q}(t)) = (q(t), \dot{q}(t))$$

lies in D, it follows that $\dot{u}(t) \in F_{(q(t),\dot{q}(t))}$. Now $\dot{u}(t) = (\dot{q}(t), \ddot{q}(t)) \in H_{(q(t),\dot{q}(t))}$. To see this note that

$$0 = \langle \alpha_j(q(t))|\dot{q}(t)\rangle, \tag{64}$$

since $(q(t), \dot{q}(t)) \in D$. Differentiating (64) with respect to t gives

$$0 = \langle \alpha_j(q(t))|\ddot{q}(t)\rangle + \langle D\alpha_j(q(t))\dot{q}(t)|\dot{q}(t)\rangle. \tag{65}$$

By item 3 above, equations (64) and (65) imply that $\dot{u}(t) = (\dot{q}(t), \ddot{q}(t)) \in H_{(q(t),\dot{q}(t))}$. Since $t \mapsto q(t)$ is an admissible motion, the Hamilton-d'Alembert principle 1.5.12 holds, that is, for every $(\dot{q}', \dot{v}') \in F_{(q,\dot{q})}$ we have

$$((\dot{q}, \ddot{q}) \,\lrcorner\, \omega)(\dot{q}', \dot{v}') - \mathrm{d}h(q,\dot{q})(\dot{q}', \dot{v}') = 0. \tag{66}$$

But $H_{(q,\dot{q})} \subseteq F_{(q,\dot{q})}$. So (66) holds for all $(\dot{q}', \dot{v}') \in H_{(q,\dot{q})}$. Thus (66) reads

$$\dot{u} \,\lrcorner\, \varpi = \partial_H h. \tag{67}$$

If $(\dot{q}', \dot{v}') \in H^{\omega}_{(q,\dot{q})} \cap F_{(q,\dot{q})}$, then (66) reads $\mathrm{d}h(q,\dot{q})(\dot{q}', \dot{v}') = 0$ because $(\dot{q}, \ddot{q}) \in H_{(q,\dot{q})}$. In other words,

$$\langle \mathrm{d}h|u\rangle = 0 \quad \text{for all } u \in H^{\omega} \cap F. \tag{68}$$

This verifies the $\Longrightarrow$ implication of theorem 1.6.3.20.

We now prove the converse. Suppose that (67) and (68) hold. Note that

$$F = (F \cap H) \oplus (F \cap H^{\omega}) \;=\; H \oplus (F \cap H^{\omega}). \tag{69}$$

From (67) we see that (66) holds for every $(\dot{q}', \dot{v}') \in H_{(q,\dot{q})}$; while from (68) it follows that (66) holds for every $(\dot{q}', \dot{v}') \in H^{\omega}_{(q,\dot{q})} \cap F_{(q,\dot{q})}$. Thus from (69) we find that equation (66) holds for every $(\dot{q}', \dot{v}') \in F_{(q,\dot{q})}$. In other words, the Hamilton-d'Alembert principle holds for the curve $t \mapsto q(t)$.

Let $t \mapsto u(t) = (q(t), r(t))$. Then by (66) the curve $t \mapsto \dot{u}(t) = (\dot{q}(t), \dot{r}(t))$ satisfies

$$\omega(q,r)((\dot{q}, \dot{r}), (\dot{q}', \dot{v}')) = \mathrm{d}h(q,r)(\dot{q}', \dot{v}') \tag{70}$$

for every $(\dot{q}', \dot{v}') \in F_{(q,r)}$. Using the local expression for the symplectic form ω and differentiating the Hamiltonian $h = k + \tau_Q^* V$, we see that (70) is equivalent to

$$\mathrm{k}(q)(\dot{q}, \dot{v}') - \mathrm{k}(q)(\dot{r}, \dot{q}') = \tfrac{1}{2}\left(D\mathrm{k}(q)\dot{q}'\right)(r,r) + \mathrm{k}(q)(r, \dot{v}') + DV(q)\dot{q} \tag{71}$$

for every $(\dot{q}', \dot{v}') \in F_{(q,r)}$. Set $\dot{q}' = 0$. Then (71) becomes

$$\mathrm{k}(q)(\dot{q}, \dot{v}') = \mathrm{q}(r, \dot{v}') \quad \text{for every } \dot{v}'.$$

This implies that $r = \dot{q}$, since $\mathrm{k}(q)$ is nondegenerate. Hence $u(t) = (q(t), \dot{q}(t))$. But $(u(t), \dot{u}(t)) \in F_{(q(t),\dot{q}(t))}$. So $T\tau_Q(u(t), \dot{u}(t)) = (q(t), \dot{q}(t)) \in D$. In other words, the curve $t \mapsto (q(t), \dot{q}(t))$ lies in the constraint distribution D. Hence $t \mapsto q(t)$ is an admissible motion. □

1.7 Almost Poisson brackets

1.7.1 *Hamilton's equations*

In this section we define an almost Poisson bracket and give a formulation of Hamilton's equations of the distributional Hamiltonian system (D, H, ϖ, h) in terms of this bracket.

Define a bracket $\{\,,\,\}$ on $C^\infty(D)$ by

$$\{f, g\}(u) = \varpi(u)(Y_g(u), Y_f(u)) = \langle \mathrm{d}g(u) \mid Y_f(u) \rangle\,, \tag{72}$$

for every $f, g \in C^\infty(D)$ and every $u \in D$. Here Y_f and Y_g are the distributional Hamiltonian vector fields (59) corresponding to f and g, respectively. Because $\{f, g\} = L_{Y_f} g$, the bracket $\{\,,\,\}$ satisfies Leibniz' rule

$$\{f, g \cdot h\} = \{f, g\} \cdot h + g \cdot \{f, h\}, \tag{73}$$

for every $f, g, h \in C^\infty(D)$. Here $(f \cdot g)(u) = f(u)g(u)$ for every $u \in D$. Therefore the bracket $\{\,,\,\}$ defines an *almost Poisson structure* on $C^\infty(D)$, that is, an antisymmetric bilinear map $\{\,,\,\} : C^\infty(D) \times C^\infty(D) \to C^\infty(D)$.

Because the distribution D need not be integrable, the bracket $\{\,,\,\}$ need *not* satisfy the Jacobi identity. Thus an almost Poisson bracket on $C^\infty(D)$ need not be a Poisson bracket on $C^\infty(D)$.

We can use the almost Poisson structure $\{\,,\,\}$ on $C^\infty(D)$ to rewrite the distributional form of Hamilton's equations (63) as

$$\frac{\mathrm{d}f(u(t))}{\mathrm{d}t} = \{h, f\}(u(t)) \tag{74}$$

for every $f \in C^\infty(D)$.

Proposition 1.7.1.21. *A curve $u : \mathbb{R} \to D : t \mapsto u(t)$ satisfies equation (74) for every $f \in C^\infty(D)$ and all t if and only if it is an integral curve of the distributional Hamiltonian vector field Y_h, that is, $\frac{\mathrm{d}u(t)}{\mathrm{d}t} = Y_h(u(t))$.*

Proof. Suppose that $t \mapsto u(t)$ is an integral curve of Y_h and that $f \in C^\infty(D)$. Then

$$\frac{\mathrm{d}f(u(t))}{\mathrm{d}t} = (L_{Y_h} f)(u(t)) = \langle \mathrm{d}f(u(t)) \mid Y_h(u(t))\rangle = \{h, f\}(u(t)).$$

Conversely, suppose that (74) holds for every $f \in C^\infty(D)$. Recall that every tangent vector w_u to D at u can be identified with the derivation $f \mapsto \langle \mathrm{d}f(u)|w_u\rangle$. From the definition of the bracket $\{\,,\,\}$, the derivation $f \mapsto \{h, f\}$ corresponds to L_{Y_h}. Hence the tangent vector to the curve $t \mapsto u(t)$ at $u(t)$ is $Y_h(u(t))$, that is $\frac{\mathrm{d}u(t)}{\mathrm{d}t} = Y_h(u(t))$. Therefore $t \mapsto u(t)$ is an integral curve of Y_h. □

Corollary 1.7.1.22. *The smooth function $h : D \to \mathbb{R}$ is constant on the integral curves of Y_h.*

Proof. We have

$$L_{Y_h} h = Y_h \,\lrcorner\, \mathrm{d}h = \{h, h\} = 0,$$

since $\{\,,\,\}$ is skew symmetric. □

For each $u \in D$ we have the inclusion map $i_u : H_u \to T_uD$. Because i_u is injective, its transpose $i_u^t : T_u^*D \to H_u^*$ is surjective. Since the skew symmetric form $\varpi_u : H_u \times H_u \to \mathbb{R}$ is bilinear, there is a linear mapping $\varpi_u^\flat : H_u \to H_u^*$ defined by $\langle \varpi_u^\flat(v_u) \mid w_u\rangle = \varpi_u(v_u, w_u)$ for every $v_u, w_u \in H_u$. Because ϖ_u is nondegenerate, the map $\varpi_u^\flat$ is invertible. We denote its inverse by $\varpi_u^\sharp$. For each $u \in D$ define the map $\Pi_u : T_u^*D \times T_u^*D \to \mathbb{R}$ by

$$\Pi_u(\alpha_u, \beta_u) = \langle \beta_u \mid i_u \circ \varpi_u^\sharp \circ i_u^t(\alpha_u)\rangle,$$

for every $\alpha_u, \beta_u \in T^*_u D$. Then $\Pi_u : T^*_u D \times T^*_u D \to \mathbb{R}$ is a skew symmetric bilinear map. Moreover the map $u \mapsto \Pi_u$ is smooth and is called the *almost Poisson structure tensor field* Π associated to the almost Poisson bracket $(\{\, , \,\}, C^\infty(D))$. Because D is a finite dimensional smooth manifold, for every $u \in D$ we have $T^*_u D = \mathrm{span}\{\mathrm{d}f(u) \big| \text{for all } f \in C^\infty(D)\}$. Therefore for every $f, g \in C^\infty(D)$ and every $u \in D$, we have

$$\begin{aligned}\Pi_u(\mathrm{d}f(u), \mathrm{d}g(u)) &= \langle \mathrm{d}g(u) \mid i_u({\varpi_u}^\sharp(i^t_u(\mathrm{d}f(u))))\rangle \\ &= \langle \mathrm{d}g(u) \mid Y_f(u)\rangle = \{f, g\}(u). \end{aligned} \tag{75}$$

Since Π_u is bilinear, it induces a linear map $\Pi^\sharp_u : T^*_u D \to T_u D$ defined by $\langle \beta_u \mid \Pi^\sharp_u(\alpha_u)\rangle = \Pi_u(\alpha_u, \beta_u)$ for every $\alpha_u, \beta_u \in T^*_u D$. The map $u \mapsto \Pi^\sharp_u$ is smooth.

Lemma 1.7.1.23. *For every $u \in D$, we have*

$$H_u = \mathrm{im}\, \Pi^\sharp_u = \mathrm{span}\{Y_f(u) \big| \textit{for every } f \in C^\infty(D)\}. \tag{76}$$

Proof. We show that the first equality in (76) holds. By definition $\Pi^\sharp_u(\mathrm{d}f(u)) = i_u(\varpi^\sharp_u(i^t_u(\mathrm{d}f(u)))) = Y_f(u)$ for every $f \in C^\infty(D)$. But $i^t_u : T^*_u D \to H^*_u$ is surjective and $\varpi^\sharp_u : H^*_u \to H_u$ is bijective. Therefore $\mathrm{im}\, \Pi^\sharp_u = H_u$.

Next we prove the second equality in (76). Suppose that $f \in C^\infty(D)$. Since $Y_f(u) \in H_u \subseteq T_u D$, it follows that $\mathrm{span}\{Y_f(u) \big| \text{for every } f \in C^\infty(D)\} \subseteq H_u$. Conversely, suppose that $v_u \in H_u$. Then $\varpi^\flat_u(v_u) \in H^*_u$. Since $i^t_u : H^*_u \to T^*_u D$ is surjective and $T^*_u D = \mathrm{span}\{\mathrm{d}g(u) \big| \text{for all } g \in C^\infty(D)\}$, there is a smooth function f on D such that $i^t_u(\mathrm{d}f(u)) = \varpi^\flat_u(v_u)$. In other words, $v_u = \varpi^\sharp_u(i^t_u(\mathrm{d}f(u))) = Y_f(u)$. Therefore $H_u \subseteq \mathrm{span}\{Y_f(u) \big| \text{for every } f \in C^\infty(D)\}$. This proves the lemma. □

Now we show how to recover a symplectic generalized distribution starting from an almost Poisson structure tensor field. Suppose that M is a smooth manifold and that $\{\, , \,\}$ is an almost Poisson structure on $C^\infty(M)$, that is, $\{\, , \,\} : C^\infty(M) \times C^\infty(M) \to C^\infty(M)$, which is bilinear, skew symmetric, and satisfies Leibniz' rule

$$\{f, g \cdot h\} = \{f, g\} \cdot h + g \cdot \{f, h\}$$

for every $f, g, h \in C^\infty(M)$. For every $m \in M$ define the map

$$\Pi_m : T^*_m M \times T^*_m M \to \mathbb{R} : (\mathrm{d}f(m), \mathrm{d}g(m)) \mapsto \{f, g\}(m).$$

Recall that the cotangent space $T^*_m M$ to M at m is equal to $\mathrm{span}\{\mathrm{d}f(m) \big| f \in C^\infty(M)\}$. Then Π_m is bilinear and skew symmetric.

Moreover, Π_m depends smoothly on m. We call the map $m \mapsto \Pi_m$ the *almost Poisson tensor field* Π on M associated to the almost Poisson structure $\{\, ,\, \}$ on $C^\infty(M)$. Because Π_m is bilinear, there is an associated linear map $\Pi^\sharp_m : T^*_m M \to T_m M$ defined by $\langle \mathrm{d}g(m) \mid \Pi^\sharp_m(\mathrm{d}f(m))\rangle = \Pi_m(\mathrm{d}f(m), \mathrm{d}g(m))$, which depends smoothly on m. For $f \in C^\infty(M)$ let $P_f(m) = \Pi^\sharp_m(\mathrm{d}f(m))$. Then $m \mapsto P_f(m)$ is a smooth vector field on M called the *almost Poisson vector field* associated to f.

Lemma 1.7.1.24. *For each $m \in M$ let $H_m = \operatorname{im}\Pi^\sharp_m$. Then the map $m \mapsto H_m$ is a* generalized distribution H *on M, in the sense that it is locally spanned by smooth vectors on M.*

Proof. By definition $H_m = \operatorname{span}\{P_f(m) \in T_m M \,\big|\, f \in C^\infty(M)\}$. For each $m \in M$ choose $f_1, \ldots, f_{n_m} \in C^\infty(M)$ such that $\{P_{f_i}(m)\}_{i=1}^{n_m}$ spans H_m. Since the function $M \mapsto \mathbb{Z}_{>0} : m \mapsto n_m$ is lower semi-continuous, there is an open neighborhood U of m in M such that it attains a maximum N. Therefore for every $m \in U$ we have $H_m = \operatorname{span}\{P_{f_i}(m)\}_{i=1}^{N}$. Consequently, $m \to H_m$ is a generalized distribution on M. □

For each $m \in M$ by definition of H_m the linear map $\Pi^\sharp_m : T^*_m M \to H_m$ is surjective. Therefore the induced linear map $\widetilde{\Pi}^\sharp_m : T^*_m M/\ker\Pi^\sharp_m \to H_m$ is a bijective linear map.

Lemma 1.7.1.25. *The linear map*

$$j_m : T^*_m M/\ker\Pi^\sharp_m \to H^*_m : \alpha_m + \ker\Pi^\sharp_m \mapsto \alpha_m | H_m$$

is bijective.

Proof. To see that the map j_m is well defined it suffices to show that $\ker\Pi^\sharp_m | H_m = 0$. Let $\beta_m \in \ker\Pi^\sharp_m$ and let $v_m \in H_m$. Since $\Pi^\sharp_m : T^*_m M \to H_m$ is surjective, there is a $\gamma_m \in T^*_m M$ such that $\Pi^\sharp_m(\gamma_m) = v_m$. Therefore

$$\begin{aligned}\beta_m(v_m) &= \beta_m(\Pi^\sharp_m(\gamma_m)) = \Pi_m(\gamma_m, \beta_m)\\ &= -\Pi_m(\beta_m, \gamma_m),\ \text{since } \Pi_m \text{ is skew symmetric}\\ &= -\gamma_m(\Pi^\sharp_m(\beta_m)) = 0,\ \text{since } \beta_m \in \ker\Pi^\sharp_m. \end{aligned} \tag{77}$$

Therefore j_m is well defined. Clearly it is linear.

Since the inclusion map $i_m : H_m \to T_m M$ is injective, its transpose $i^t_m : T^*_m M \to H^*_m$ is surjective. Let $\beta_m \in H^*_m$. Then there is $\gamma_m \in T^*_m M$ such that $i^t_m(\gamma_m) = \beta_m$. For $v_m \in H_m$, we have

$$\beta_m(v_m) = (i^t_m(\gamma_m))(v_m) = \gamma_m(i_m(v_m)) = \gamma_m(v_m),$$

that is, $\gamma_m | H_m = \beta_m$. So $j_m(\gamma_m + \ker \Pi^\sharp_m) = \gamma_m | H_m = \beta_m$. Therefore j_m is surjective. Since $\widetilde{\Pi}^\sharp_m$ is a bijective linear map, it follows that $\dim T^*_m M / \ker \Pi^\sharp_m = \dim H_m = \dim H^*_m$. Therefore j_m is bijective. □

Let $\varpi^\flat_m = j_m \circ (\widetilde{\Pi}^\sharp_m)^{-1} : H_m \to H^*_m$. Then $\varpi^\flat_m$ is a bijective linear mapping associated to the skew symmetric bilinear map

$$\varpi_m : H_m \times H_m \to \mathbb{R} : (v_m, w_m) \mapsto \langle \varpi^\flat_m(v_m) \mid w_m \rangle.$$

Thus we have proved

Proposition 1.7.1.26. *The generalized distribution (H, ϖ) on M is symplectic.*

Corollary 1.7.1.27. *For every $f \in C^\infty(M)$ the Hamiltonian vector field Y_f with respect to the generalized symplectic distribution (H, ϖ) is equal to the almost Poisson vector field P_f.*

Proof. For every $m \in M$ we have

$$P_f(m) = \Pi^\sharp_m(\mathrm{d}f(m)) = \widetilde{\Pi}^\sharp_m(\mathrm{d}f(m) + \ker \Pi^\sharp_m) \in H_m,$$

by definition of H_m. So

$$\begin{aligned} \varpi^\flat_m(P_f(m)) &= \varpi^\flat_m(\widetilde{\Pi}^\sharp_m(\mathrm{d}f(m) + \ker \Pi^\sharp_m)) = j_m(\mathrm{d}f(m) + \ker \Pi^\sharp_m) \\ &= \mathrm{d}f(m) | H_m = \varpi^\flat_m(Y_f(m)). \end{aligned}$$

Therefore $P_f(m) = Y_f(m)$ because $\varpi^\flat_m$ is invertible. □

1.7.2 *Nonholonomic Dirac brackets*

Next we give a procedure for finding an almost Poisson bracket on D in terms of the Poisson bracket on (TQ, ω), defined by (72). Suppose that the nonholonomic constraint D is a distribution on Q given by $D_q = \bigcap_{i=1}^{\ell} \ker \varphi_i(q)$, where φ_i, $i = 1, \ldots, \ell$, are 1-forms on Q such that $\varphi_i(q)$ are linearly independent on $T^*_q Q$ for every $q \in Q$. Since D is a distribution, it is a submanifold of TQ, defined by the common zeroes of the *constraint functions*

$$c_i : TQ \to \mathbb{R} : u \mapsto \langle \varphi_i(\tau_Q(u)) | u \rangle, \qquad 1 \le i \le \ell. \tag{78}$$

Here $\tau_Q : TQ \to Q$ is the tangent bundle projection. Note that for every $v_u \in T_u(TQ)$

$$c_i(u) = \langle \tau^*_Q \varphi_i(u) \mid v_u \rangle, \tag{79}$$

because

$$\langle \tau_Q^* \varphi_i(u) \mid v_u \rangle = \langle \varphi_i(\tau_Q(u)) \mid T_u \tau_Q v_u \rangle \ = \ \langle \varphi_i(\tau_Q(u)) \mid u \rangle.$$

Now consider the *constraint map* $c : TQ \to \mathbb{R}^\ell : u \mapsto \big(c_1(u), \ldots \ c_\ell(u)\big)$. Then

Fact 1.7.2.28. 0 *is a regular value of the map* c.

Proof. First we show that for every $u \in TQ$ the 1-forms $\{\mathrm{d}c_i(u)\}_{i=1}^\ell$ are linearly independent. Suppose that $0 = \sum_{i=1}^\ell \alpha_i \, \mathrm{d}c_i(u)$. Then $0 = \sum_{i=1}^\ell \alpha_i \, \mathrm{d}\tau_Q^* \varphi_i(u)$, using (79), which implies $0 = \tau_Q^*\big(\sum_{i=1}^\ell \alpha_i \, \mathrm{d}\varphi_i(q)\big)$, where $q = \tau_Q(u)$. Hence $0 = \sum_{i=1}^\ell \alpha_i \, \mathrm{d}\varphi_i(q)$. But then $\alpha_i = 0$ for all $1 \le i \le \ell$ because $\{\mathrm{d}\varphi_i(q)\}_{i=1}^\ell$ are linearly independent. Since $\{\mathrm{d}c_i(u)\}_{i=1}^\ell$ are linearly independent, the derivative of the map c is surjective. □

Next we show that

Fact 1.7.2.29. $c^{-1}(0) = D$.

Proof. To see this note that

$$\begin{aligned} c_i(u) = 0 \quad \text{for all } i &\Longleftrightarrow \langle \varphi_i(\tau_Q(u)) | u \rangle = 0 \quad \text{for all } i \\ &\Longleftrightarrow u \in \ker \varphi_i(q) \quad \text{for all } i, \text{ where } q = \tau(u) \\ &\Longleftrightarrow u \in D_q. \end{aligned}$$

□

We now prove

Fact 1.7.2.30. *For every* $u \in D$ *we have*

$$T_u^\omega D = \mathrm{span}\{X_{c_1}(u), \ldots, X_{c_\ell}(u)\}. \tag{80}$$

Proof. For every $v_u \in T_u D$,

$$\omega_u(X_{c_i}(u), v_u) = \langle \mathrm{d}c_i(u) \mid v_u \rangle \ = \ 0.$$

So $X_{c_i}(u) \in T_u^\omega D$. Therefore $\{X_{c_i}(u)\}_{i=1}^\ell \subseteq T_u^\omega D$. Note that $\{X_{c_i}(u)\}_{i=1}^\ell$ are linearly independent because if

$$0 = \sum_{i=1}^\ell \alpha_i X_{c_i}(u) \ = \ \omega_u^\sharp\big(\sum_{i=1}^\ell \alpha_i \, \mathrm{d}c_i(u)\big),$$

then $0 = \sum_{i=1}^\ell \alpha_i \, \mathrm{d}c_i(u)$. But this implies $\alpha_i = 0$ for $1 \le i \le \ell$. From fact 1.7.2.28 it follows that $(T_u D)^\circ = \mathrm{span}\,\{\mathrm{d}c_i(u)\}_{i=1}^\ell$. Moreover, $\omega_u^\sharp$ is a bijective linear map from $(T_u D)^\circ$ onto $T_u^\omega D$. Therefore $\dim T_u^\omega D = \ell$, which completes the proof. □

By definition, F (28) is the distribution on TQ whose image under $T\tau_Q$ is the distribution D. In our situation we have

Fact 1.7.2.31. *For every $u \in TQ$*

$$F_u^\omega = \mathrm{span}\{\omega_u^\sharp(\tau_Q^*\varphi_i(u)) \big| \, i = 1, \ldots, \ell\}. \tag{81}$$

Proof. To see this we argue as follows. By definition, $\omega_u^\sharp(\tau_Q^*\varphi_i(u))$ is the unique vector in $T_u(TQ)$ such that for every $w \in T_u(TQ)$

$$\langle \varphi_i(q) | T\tau_Q w\rangle = \langle \tau_Q^*\varphi_i(u)|w\rangle \; = \omega(\omega_u^\sharp(\tau_Q^*\varphi_i(u)), w), \tag{82}$$

where $q = \tau_Q(u)$. Now

$$w \in F_u \Leftrightarrow T\tau_Q v \in D_q \Leftrightarrow \langle \tau_Q^*\varphi_i(u) \mid v\rangle = \langle \varphi_i(\tau_Q(u))|T\tau_Q v\rangle = 0 \quad \forall \; i.$$

From (82) it follows that $\omega_u^\sharp(\tau_Q^*\varphi_i(u)) \in F_u^\omega$.

Conversely, suppose that $w' \in F_u^\omega$. Let $q = \tau_Q(u)$. Since $w' \in T_D(TQ)$, lemma 1.6.3.18 implies that there is a $v' \in T_qQ$ k-orthogonal to D_q such that $w' = \iota_u(v')$. Consider the 1-form on Q defined by $\varphi(q) = k(q)^\flat(v')$. Then for every $u' \in D_q$, we have $\langle \varphi(q) \mid u'\rangle = k(v', u') = 0$ because v' is k-orthogonal to D_q. Thus $\varphi(q)$ is a linear combination of the 1-forms $\varphi_i(q)$. For every $w \in T_u(TQ)$ we have

$$\omega(u)(w', w) = k(q)(v', T\tau_Q w) \; = \; \langle \varphi_i(\tau_Q(u))|T\tau_Q w\rangle \; = \; \langle \tau_Q^*\varphi_i(u)|w\rangle,$$

which implies

$$w' \in \mathrm{span}\{\omega_u^\sharp(\tau_Q^*\varphi_i)(u) \big| \, i = 1, \ldots, \ell\}.$$

This proves (81). □

By definition of the symplectic distribution $H = F \cap TD$ (34) on D, using theorem 1.6.3.20 we have $F_u^\omega \cap T_u^\omega D = \{0\}$ for every $u \in D$. Consequently,

$$\begin{aligned} H_u^\omega &= T_u^\omega D \oplus F_u^\omega \\ &= \mathrm{span}\{X_{c_1}(u), \ldots, X_{c_\ell}(u), \omega_u^\sharp(\tau_Q^*\varphi_1(u)), \ldots, \omega_u^\sharp(\tau_Q^*\varphi_\ell)(u)\} \end{aligned} \tag{83}$$

is a ω_u-symplectic subspace of $(T_u(TQ), \omega_u)$ of dimension 2ℓ.

The Poisson bracket $[f_1, f_2]$ of $f_1, f_2 \in C^\infty(TQ)$, with respect to the symplectic form ω, can be expressed in terms of the differentials $\mathrm{d}f_1$ and $\mathrm{d}f_2$ as

$$[f_1, f_2] = \omega(\omega^\sharp(\mathrm{d}f_1), \omega^\sharp(\mathrm{d}f_2)).$$

In the above formula, we can replace differentials by arbitrary 1-forms and extend the notion of the Poisson bracket to one forms which need not be exact. In particular, we can write

$$[\tau_Q^*\varphi_i, \tau_Q^*\varphi_j] = \omega(\omega^\sharp(\tau_Q^*\varphi_i), \omega^\sharp(\tau_Q^*\varphi_j)), \qquad (84)$$
$$\{c_i, \tau_Q^*\varphi_j\} = \omega(\omega^\sharp(\mathrm{d}c_i), \omega^\sharp(\tau_Q^*\varphi_j)).$$

Since $T_D(TQ) = H \oplus H^\omega$, for every $f \in C^\infty(TQ)$, we can express the restriction of X_f to points in D in the form $X_{f|D} = Y_f + Z_f$, where Y_f has values in H and Z_f has values in H^ω. Similarly, for $g \in C^\infty(TQ)$, the restriction $\mathrm{d}g\ |_{T_DTQ}$ of $\mathrm{d}g$ to points in D can be written as $\mathrm{d}g\ |_{T_D(TQ)} = \partial_H g \oplus \partial_{H^\omega} g$, where $\partial_H g$ annihilates H^ω and $\partial_{H^\omega} g$ annihilates H. Hence, restriction of $[f, g]$ to D can be expressed as

$$\begin{aligned}[f,g]\Big|_D &= \langle dg \mid X_f\rangle\Big|_D \\ &= \langle \mathrm{d}g \mid Y_f + Z_f\rangle\Big|_D = \langle \partial_H g \oplus \partial_{H^\omega} g \mid Y_f\rangle + \langle \mathrm{d}g \mid Z_f\rangle \\ &= \langle \partial_H g \mid Y_f\rangle + \langle \mathrm{d}g \mid Z_f\rangle = \{f\Big|_D, g\Big|_D\} + \langle \mathrm{d}g \mid Z_f\rangle,\end{aligned}$$

where $\{f\big|_D, g\big|_D\}$ is the nonholonomic almost Poisson bracket of $f\big|_D$ and $g\big|_D$ in $C^\infty(D)$. From equation (83) that

$$\begin{aligned}Z_f(u) &= a_1(u)X_{c_1}(u) + \ldots + a_\ell(u)X_{c_\ell}(u) \\ &\quad + a_{\ell+1}(u)\,\omega_u^\sharp(\tau_Q^*\varphi_1(u)) + \ldots + a_{2\ell}(u)\,\omega_u^\sharp(\tau_Q^*\varphi_\ell(u))\end{aligned}$$

for every $u \in D$. We can express the coefficients $a_1(u), \ldots, a_\ell(u)$ and $b_1(u), \ldots, b_\ell(u)$ in terms of Poisson brackets of $c_1, ..., c_\ell$ and $\tau_Q^*\varphi_1, ..., \tau_Q^*\varphi_\ell$ with f as follows. Since $\langle \mathrm{d}c_j \mid Y_f\rangle = 0$ and $\langle \tau_Q^*\varphi_j \mid Y_f\rangle = 0$ for all $j = 1, \ldots, \ell$, we get

$$\begin{aligned}[f, c_j] &= \langle \mathrm{d}c_j \mid X_f\rangle = \langle \mathrm{d}c_j \mid Z_f\rangle \\ &= \sum_{i=1}^{\ell} a_i \langle \mathrm{d}c_j \mid X_{c_i}\rangle + \sum_{i=1}^{\ell} a_{\ell+i} \langle \mathrm{d}c_j \mid \omega^\sharp(\tau_Q^*\varphi_i)\rangle, \\ [f, \tau_Q^*\varphi_j] &= \langle \tau_Q^*\varphi_j \mid X_f\rangle = \langle \tau_Q^*\varphi_j \mid Z_f\rangle \\ &= \sum_{i=1}^{\ell} a_i \langle \tau_Q^*\varphi_j \mid X_{c_i}\rangle + \sum_{i=1}^{\ell} a_{\ell+i} \langle \tau_Q^*\varphi_j \mid \omega^\sharp(\tau_Q^*\varphi_i)\rangle.\end{aligned}$$

Using the definition of the extended Poisson brackets (84) we are led to a system of 2ℓ linear equations

$$\begin{aligned}\textstyle\sum_{i=1}^{\ell} a_i\,[c_i, c_j] + \sum_{i=1}^{\ell} a_{\ell+i}\,[\tau_Q^*\varphi_i, c_j] &= [f, c_j] \\ \textstyle\sum_{i=1}^{\ell} a_i\,[c_i, \tau_Q^*\varphi_j] + \sum_{i=1}^{\ell} a_{\ell+i}[\tau_Q^*\varphi_i, \tau_Q^*\varphi_j] &= [f, \tau_Q^*\varphi_j]\end{aligned} \qquad (85)$$

for $a_1, ..., a_{2\ell}$. This system can be written in terms of matrices as $Aa = b$, where

$$A = \left(\begin{array}{c|c} ([c_j, c_i]) & ([c_j, \tau_Q^* \varphi_i]) \\ \hline ([\tau_Q^* \varphi_j, c_i]) & ([\tau_Q^* \varphi_j, \tau_Q^* \varphi_i]) \end{array} \right). \tag{86}$$

Claim 1.7.2.32. *The matrix A is invertible.*

Proof. We compute each block of A. First,

$$\omega_u(X_{c_j}(u), X_{c_i}(u)) = \langle \mathrm{d}c_j(u) \mid X_{c_i}(u) \rangle = [c_j, c_i]_u.$$

Next

$$\omega_u(\omega^\sharp(\tau_Q^* \varphi_i(u)), X_{c_j}(u)) = \langle \tau_Q^* \varphi_i(u) \mid X_{c_j}(u) \rangle \;=\; [c_j, \tau_Q^* \varphi_i]_u.$$

Finally

$$\omega_u(\omega^\sharp(\tau_Q^* \varphi_i(u)), \omega^\sharp(\tau_Q^* \varphi_i(u))) = \langle \tau_Q^* \varphi_i(u) \mid \omega^\sharp(\tau_Q^* \varphi_i(u)) \rangle \;=\; [\tau_Q^* \varphi_j, \tau_Q^* \varphi_i]_u.$$

Therefore the matrix $A = (A_{ij})$ is the matrix of the symplectic form ω with respect to the basis (83) of H^ω. Hence A is invertible because (H_u^ω, ω_u) is a symplectic subspace of $(T_u(TQ), \omega_u)$. □

Let $A^{-1} = (A^{ij}) = \begin{pmatrix} (b_{ij}) & (c_{ij}) \\ (-c_{ji}) & (d_{ij}) \end{pmatrix}$ be the inverse of the matrix A. Solving (85) gives

$$a_i = \sum_{j=1}^{\ell} \left(b_{ij}\,[f, c_j] + c_{ij}\,[f, \tau_Q^* \varphi_j] \right),$$

$$a_{\ell+i} = \sum_{j=1}^{\ell} \left(-c_{ji}\,[f, c_j] + d_{ij}\,[f, \tau_Q^* \varphi_j] \right),$$

when written out in components. Now

$$\begin{aligned} \langle \mathrm{d}g \mid Z_f \rangle &= \sum_{i=1}^{\ell} \left(a_i \langle \mathrm{d}g \mid X_{c_i} \rangle + a_{\ell+i} \langle \mathrm{d}g \mid \omega^\sharp(\tau_Q^* \varphi_i) \rangle \right) \\ &= \sum_{i=1}^{\ell} \left(a_i\,[g, c_i] + a_{\ell+i}\,[g, \tau_Q^* \varphi_i] \right) \\ &= \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \left(b_{ij}\,[f, c_j] + c_{ij}[f, \tau_Q^* \varphi_j] \right) [g, c_i] + \\ &\qquad \left(-c_{ji}\,[f, c_j] + d_{ij}\,[f, \tau_Q^* \varphi_j] \right) [g, \tau_Q^* \varphi_i] \\ &= \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} [g, c_i]\, b_{ij}\,[f, c_j] + [g, c_i]\, c_{ij}\,[f, \tau_Q^* \varphi_j] \\ &\qquad - [g, \tau_Q^* \varphi_i]\, c_{ji}\,[f, c_j] + [g, \tau_Q^* \varphi_i]\, d_{ij}\,[f, \tau_Q^* \varphi_j]. \end{aligned}$$

This implies that

$$\{f\big|_D, g\big|_D\} = [f,g]\big|_D - \sum_{1\le i,j\le \ell} ([g,c_i],[g,\tau_Q^*\varphi_i])A^{ij}\begin{pmatrix}[f,c_j]\\ [f,\tau_Q^*\varphi_j]\end{pmatrix}. \tag{87}$$

In chapter 5 in order to find the equations of motion for Carathéodory's sleigh, we use the special case of (87) when $\ell = 1$. Written out explicitly, we get

$$\begin{aligned}\{f\big|_D, g\big|_D\} &= [f,g]\big|_D \\ &+ \tfrac{1}{[c_1,\tau_Q^*\varphi_1]}([g,c_1],[g,\tau_Q^*\varphi_1])\begin{pmatrix}0 & [c_1,\tau_Q^*\varphi_1]\\ -[c_1,\tau_Q^*\varphi_1] & 0\end{pmatrix}\begin{pmatrix}[f,c_1]\\ [f,\tau_Q^*\varphi_1]\end{pmatrix}.\end{aligned} \tag{88}$$

1.8 Momenta and the momentum equation

In this section Q is a smooth manifold with a Riemannian metric k.

1.8.1 *Momentum functions*

Let Z be a smooth vector field on Q. The *momentum function* associated to Z is $\mu_Z : T^*Q \to \mathbb{R} : \alpha_q \mapsto \langle \alpha_q \mid Z(q)\rangle$. Let $\varphi_s : Q \to Q$ be the flow of Z. Then the flow of the *cotangent lift* Z_{T^*Q} of Z to T^*Q is

$$\widetilde{\varphi}_s : T^*Q \to T^*Q : \alpha_q \mapsto (T_q\varphi_{-s})^t\alpha_q.$$

Note that $T\pi_Q \circ Z_{T^*Q} = Z\circ\pi_Q$, where $\pi_Q : T^*Q \to Q$ is the cotangent bundle projection map.

Lemma 1.8.1.33. *Z_{T^*Q} is the Hamiltonian vector field X_{μ_Z} on (T^*Q,ω_Q) corresponding to the momentum function μ_Z. Here $\omega_Q = -\,\mathrm{d}\theta_Q$, where θ_Q is the canonical 1-form on T^*Q.*

Proof. First we show that θ_Q is invariant under the flow of Z_{T^*Q}. For $u \in T_qQ$, $\alpha \in T_q^*Q$, and $v_\alpha \in T_\alpha(T^*Q)$ we have

$$\begin{aligned}\langle \widetilde{\varphi}_s^*\theta_Q(\alpha) \mid v_\alpha\rangle &= \langle \theta_Q(\widetilde{\varphi}_s(\alpha)) \mid T_u\widetilde{\varphi}_s v_\alpha\rangle = \langle \widetilde{\varphi}_s(\alpha) \mid T_u(\tau_Q\circ\widetilde{\varphi}_s)v_\alpha\rangle \\ &\qquad\text{where } \tau_Q \text{ is the tangent bundle projection map} \\ &= \langle \widetilde{\varphi}_s(\alpha) \mid T\varphi_s(T_u\tau_Q v_\alpha)\rangle = \langle \alpha \mid u\rangle = \langle \theta_Q(\alpha) \mid v_\alpha\rangle.\end{aligned}$$

Therefore

$$0 = (L_{Z_{T^*Q}}\theta_Q)(\alpha) = (Z_{T^*Q} \,\lrcorner\, \mathrm{d}\theta_Q)(\alpha) + \mathrm{d}\langle \theta_Q(\alpha) \mid Z_{T^*Q}(\alpha)\rangle,$$

that is,

$$(Z_{T^*Q} \,\lrcorner\, \omega_Q)(\alpha) = \mathrm{d}\langle \theta_Q(\alpha) \mid Z_{T^*Q}(\alpha)\rangle$$
$$= \langle \alpha \mid T\pi_Q \, Z_{T^*Q}(\alpha)\rangle = \langle \alpha \mid Z(\pi_Q(\alpha)\rangle = \mu_Z(\alpha).$$

□

Using the vector bundle isomorphism $\mathrm{k}^\flat : TQ \to T^*Q$, we pull back the momentum function μ_Z to a smooth function P_Z on TQ. In other words, $P_Z(v_q) = \mathrm{k}(q)(v_q, Z(q))$ for every $q \in Q$ and every $v_q \in T_qQ$. P_Z is the *momentum function* on TQ associated to the vector field Z with respect to the Riemannian metric k. Let X_{P_Z} be the Hamiltonian vector field corresponding to P_Z on the symplectic manifold (TQ, ω), where $\omega = (\mathrm{k}^\flat)^*\omega_Q$. The *tangent lift* Z_{TQ} of the vector field Z is the vector field on TQ whose flow $\overline{\varphi}_s : TQ \to TQ$ is $\overline{\varphi}_s(v_q) = T_q\varphi_s \, v_q$ for every $q \in Q$. Note that $T\tau_Q \circ Z_{TQ} = Z \circ \tau_Q$.

Lemma 1.8.1.34. *We have*

$$T\tau_Q \circ X_{P_Z} = Z \circ \tau_Q. \tag{89}$$

Also $X_{P_Z} = Z_{TQ}$ *if and only if the flow of* Z *preserves the Riemannian metric* k.

Proof. Pulling back both sides of $X_{\mu_Z} \,\lrcorner\, \omega_Q = \mathrm{d}\mu_Z$ by $\mathrm{k}^\flat$ gives

$$(\mathrm{k}^\flat)^* X_{\mu_Z} \,\lrcorner\, \omega = (\mathrm{k}^\flat)^* X_{\mu_Z} \,\lrcorner\, (\mathrm{k}^\flat)^*\omega_Q = \mathrm{k}^\flat(X_{\mu_Z} \,\lrcorner\, \omega_Q)$$
$$= (\mathrm{k}^\flat)^* \,\mathrm{d}\mu_Z = \mathrm{d}((\mathrm{k}^\flat)^*\mu_Z) = \mathrm{d}P_Z.$$

Therefore $X_{P_Z} = (\mathrm{k}^\flat)^* X_{\mu_Z}$. From $T\pi_Q \circ \mathrm{k}^\flat = T\tau_Q$ it follows that

$$T\tau_Q \circ X_{P_Z} = T\pi_Q \circ \mathrm{k}^\flat \circ (\mathrm{k}^\flat)^* X_{\mu_Z} = T\pi_Q \circ X_{\mu_Z} \circ \mathrm{k}^\flat$$
$$= Z \circ \pi_Q \circ \mathrm{k}^\flat = Z \circ \tau_Q.$$

This proves (89).

Since $X_{P_Z} = (\mathrm{k}^\flat)^* X_{\mu_Z}$ and $X_{\mu_Z} = Z_{T^*Q}$, it follows that $X_{P_Z} = Z_{TQ}$ if and only if $Z_{TQ} = (\mathrm{k}^\flat)^* Z_{T^*Q}$. This equality holds if and only if $T\varphi_s = (\mathrm{k}^\flat)^{-1} \circ (T\varphi_{-s})^t \circ \mathrm{k}^\flat$, that is, for every $q \in Q$ and every $v_q \in T_qQ$ we have

$$\mathrm{k}^\flat(\varphi_s(q))(T_q\varphi_s v_q) = \mathrm{k}^\flat(q)(v_q) T_{\varphi_{-s}(q)}\varphi_s. \tag{90}$$

Evaluating both sides of (90) on $T_q\varphi_s w_q$, where $w_q \in T_qQ$, we get

$$\mathrm{k}(\varphi_s(q))(T_q\varphi_s v_q, T_q\varphi_s w_q) = \mathrm{k}(q)(v_q, w_q),$$

for every $q \in Q$ and every $v_q, w_q \in T_qQ$. In other words, the flow of Z is an isometry. □

1.8.2 Momentum equations

Let (D, H, ϖ, h) be a distributional Hamiltonian system with constraint distribution D on Q and symplectic distribution (H, ϖ) on D.

A smooth curve $\gamma : I \subseteq \mathbb{R} \to TQ$ satisfies the *second order equation condition* for $f \in C^\infty(Q)$ if and only if for every $t \in I$

$$\frac{\mathrm{d}\tau_Q^* f(\gamma(t))}{\mathrm{d}t} = \langle \mathrm{d}f(q(t)) \mid v(t) \rangle, \tag{91}$$

where $q(t) = \tau_Q(\gamma(t))$ and $v(t) = \gamma(t) \in T_{q(t)}Q$. Note that (91) is equivalent to

$$\langle \mathrm{d}f(\gamma(t)) \mid T_{\gamma(t)}\tau_Q(\dot{\gamma}(t)) \rangle = \langle \mathrm{d}f(\gamma(t)) \mid v(t) \rangle \tag{92}$$

for every $t \in I$.

Lemma 1.8.2.35. *For every $f \in C^\infty(Q)$ the following statements are equivalent.*

1. *Y_f defines a second order differential equation. In other words, if $I \subseteq \mathbb{R} \to D \subseteq TQ : t \mapsto (q(t), v(t))$ is an integral curve of Y_f, then $v(t) = \frac{\mathrm{d}q(t)}{\mathrm{d}t}$ for every $t \in I$.*
2. *There is a function $V \in C^\infty(Q)$ such that*

$$f(q, v_q) = \tfrac{1}{2}\,\mathrm{k}(q)(v_q, v_q) + V(q), \tag{93}$$

for every $(q, v_q) \in D$.

Proof. Using a local trivialization of D and the notation of §1.6.3, we may write $Y_f = a_\rightarrow + b_\uparrow$, see (45). Writing $u = (q, v) \in D_q$ we have $b_\uparrow(u) = \iota_u(a^Q(u))$. Therefore

$$\langle \mathrm{d}f(u) \mid \iota_u(z) \rangle = \mathrm{k}(q)(a^Q(u), z), \tag{94}$$

for every $z = b^Q(u) \in D_q$. Because $a^Q(u) = T_u\tau_Q Y_f(u)$, statement 1 is equivalent to $a^Q(u) = v$. Therefore (94) is equivalent to

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(u + tz) = \mathrm{k}(q)(v, z), \tag{95}$$

for every $z \in D_q$. But (95) is equivalent to (93). □

A smooth curve $\gamma : I \subseteq \mathbb{R} \to TQ$ satisfies the *momentum equation* for the smooth vector field Z on Q if and only if

$$\frac{\mathrm{d}P_Z(\gamma(t))}{\mathrm{d}t} = -L_{P_Z} h(\gamma(t)), \tag{96}$$

for every $t \in I$.

Theorem 1.8.2.36. *Every integral curve* $I \subseteq \mathbb{R} \to D : t \mapsto u(t)$ *of the distributional Hamiltonian vector field* Y_h *associated to the distributional Hamiltonian system* (D, H, ϖ, h) *satisfies the second order equation condition for every* $f \in C^\infty(Q)$ *and the momentum equation for every smooth vector field* Z *on* Q*, which takes values in the constraint distribution* D*. Conversely, let* $\gamma : I \subseteq \mathbb{R} \to D \subseteq TQ : t \mapsto u(t) = (q(t), v(t))$ *be a smooth curve. Let* $\mathcal{F}$ *be a collection of smooth functions* f *on* Q *such that* $\mathrm{span}\{\mathrm{d}f_{q(t)}|D_{q(t)}\} = D_{q(t)}$ *for every* $t \in I$ *and let* $\mathcal{Z}$ *be a collection of smooth vector fields on* Q *with values in* D *such that* $\mathrm{span}\{Z(q(t))\} = D_{q(t)}$ *for every* $t \in I$*. If* γ *satisfies the second order equation condition for every* $f \in \mathcal{F}$ *and the momentum equation for every* $Z \in \mathcal{Z}$*, then* γ *is an integral curve of* Y_h*.*

Proof. Suppose that $\gamma : I \subseteq \mathbb{R} \to D \subseteq TQ : t \mapsto u(t) = (q(t), v(t))$ is an integral curve of Y_h. From lemma 1.8.2.35 it follows that $T_{\gamma(t)}\tau_Q(\dot{\gamma}(t)) = v(t)$. Therefore (92) holds for all $t \in I$, that is, the second order equation condition holds for every $f \in C^\infty(Q)$. Let Z be a smooth vector field on Q with values in D. Since $T\tau_Q \circ X_{P_Z} = Z \circ \tau_Q$ and $Z(Q) \subseteq D$, it follows that X_{P_Z} has values in the distribution F (28). Therefore

$$\begin{aligned} L_{Y_h} P_Z = Y_h \mathbin{\lrcorner} \mathrm{d}P_Z &= \omega(Y_h, X_{P_Z}) \\ &= -\omega(X_{P_Z}, \dot{\gamma}) = -X_{P_Z} \mathbin{\lrcorner} \mathrm{d}h = -L_{P_Z} h. \end{aligned}$$

Consequently, the Hamiltonian h satisfies the momentum equation for Z.

To prove the converse we start by observing that because γ is a smooth curve in D, it follows that $v(t) \in D_{q(t)}$ and $T_{\gamma(t)}\dot{\gamma}(t) \in D_{q(t)}$ for every $t \in I$. Since (92) holds, we obtain $T_{\gamma(t)}\dot{\gamma}(t) = v(t)$. From $T_{\gamma(t)}\tau_Q(Y_h(\gamma(t))) = v(t)$, it follows that $T_{\gamma(t)}\tau_Q\lambda(t) = 0$ for every $t \in I$, where $\lambda(t) = \dot{\gamma}(t) - Y_h(\gamma(t))$. In addition, we have $\lambda(t) \in T_{\gamma(t)}D \cap T_{\gamma(t)}\tau_Q$, since $\dot{\gamma}(t)$ and $Y_h(\gamma(t))$ both lie in $T_{\gamma(t)}D$. Therefore for every $t \in I$ we see that $\lambda(t) \in D_{\gamma(t)}$, if we identify $\ker T_{\gamma(t)}\tau_Q$ with $T_{\gamma(t)}Q$. On the other hand, the assumption that the momentum equation holds for γ implies $Y_h \mathbin{\lrcorner} \mathrm{d}P_Z = -L_{X_{P_Z}} h = \langle \mathrm{d}P_Z \mid \dot{\gamma} \rangle$. Consequently, $\langle \mathrm{d}P_Z \mid \lambda(t) \rangle = 0$ for every $Z \in \mathcal{Z}$. Because $\lambda(t)$ belongs to the tangent space of each fiber of the tangent bundle projection map τ_Q and the fiber derivative of P_Z is $\mathrm{k}^\flat(Z)$, see (95), it follows that $\mathrm{k}(\gamma(t))(Z(\gamma(t)), \lambda(t)) = 0$ for every $Z \in \mathcal{Z}$. Because $\mathrm{span}_{Z \in \mathcal{Z}}\{Z(\gamma(t))\} = D_{\gamma(t)}$, we obtain $\lambda(t) = 0$ for every $t \in I$. In other words, $\dot{\gamma}(t) = Y_h(\gamma(t))$ for every $t \in I$, that is, $t \mapsto \gamma(t)$ is an integral curve of Y_h. □

1.8.3 *Homogeneous functions*

Because we have assumed that the nonholonomic constraints are linear, the constraint manifold D is a smooth vector subbundle of the tangent bundle TQ of configuration space Q. For every nonnegative integer k let $C^\infty_k(D)$ be the set of all smooth functions f on D such that for each $q \in Q$ the restriction $f|D_q$ of f to the *vector space* D_q is a homogeneous function of degree k, that is, $f(r\, u_q) = r^k\, f(u_q)$ for every $r \in \mathbb{R}\backslash\{0\}$ and every $u_q \in D_q$. For fixed $q \in Q$ the function $f|D_q$ is smooth at the origin of D_q. Therefore $f|D_q$ is a homogeneous *polynomial* on D_q of degree k.

Let $C^\infty_0(D)$ be the space of smooth functions on D which are constant on the fibers of the vector bundle $\tau = \tau_Q|D : D \subseteq TQ \rightarrow Q$. Then $C^\infty_0(D) = \{\tau^* f \in C^\infty(D) \big| \, f \in C^\infty(Q)\}$. This proves

Lemma 1.8.3.37. *The map* $\tau^* : C^\infty(Q) \rightarrow C^\infty_0(D)$ *is an isomorphism of algebras.*

Define the space
$C^\infty_1(D) = \{f \in C^\infty(D) \big| \, f_q|D_q : D_q \subseteq T_qQ \rightarrow \mathbb{R}$ is linear for every $q \in Q\}$. Because $\mathrm{k}^\flat_q|D_q : D_q \subseteq T_qQ \rightarrow D^*_q \subseteq T^*_qQ$ is a linear isomorphism, there is a unique $Z(q) \in D_q$ such that $\mathrm{k}(q)(Z(q), u_q) = (f|D_q)(u_q)$ for every $u_q \in D_q$. Moreover, the map $q \mapsto Z(q)$ is smooth. Therefore $f = P_Z|D$, where P_Z is the momentum function corresponding to the vector field Z on Q.

Let $\mathcal{X}^\infty(Q, D)$ be the space of smooth vector fields on Q with values in the distribution D. The discussion above proves

Lemma 1.8.3.38. *The mapping*

$$\mathcal{X}^\infty(Q, D) \rightarrow C^\infty_1(D) : Z \mapsto P_Z|D$$

is an isomorphism of vector spaces.

Let $C^\infty_{\mathrm{hom}}(D)$ be the vector space of smooth homogeneous functions on D.

Lemma 1.8.3.39. $C^\infty_{\mathrm{hom}}(D)$ *is a graded almost Poisson algebra. In particular, if* $f \in C^\infty_k(D)$ *and* $g \in C^\infty_\ell(D)$*, then* $\{f, g\} \in C^\infty_{k+\ell-1}(D)$*.*

Proof. For any $r \in \mathbb{R} \setminus \{0\}$ let $m_r : T_qQ \rightarrow T_qQ : v_q \mapsto r\, v_q$. Then $m^t_r : T^*_qQ \rightarrow T^*_qQ : \alpha_q \mapsto r\, \alpha_q$. Let θ_Q be the canonical 1-form on T^*Q. Then for every $\alpha_q \in T^*Q$ and $w_\alpha \in T_\alpha(T^*Q)$ we get

$$\begin{aligned} \langle (m^t_r)^*\theta_Q(\alpha_q) \mid w_\alpha \rangle &= \langle \theta_Q(r\, \alpha_q) \mid T_{\alpha_q} m_r\, w_\alpha \rangle \\ &= \langle r\, \alpha_q \mid T(\tau_Q \circ m_r) w_\alpha \rangle = \langle r\, \alpha_q \mid T\tau_Q w_\alpha \rangle \text{ since } \tau_Q \circ m_r = \tau_Q \\ &= \langle r\, \alpha_q \mid T\tau_Q w_\alpha \rangle = r\, \langle \theta_Q(\alpha_q) \mid w_\alpha \rangle. \end{aligned}$$

So $(m_r^t)^*\theta_Q = r\,\theta_Q$. Therefore

$$(m_r^t)^*\omega_Q = (m_r^t)^*(-\,\mathrm{d}\theta_Q) = -\,\mathrm{d}\big((m_r^t)^*\theta_Q\big) = -\,\mathrm{d}(r\,\theta_Q) = -r\,\mathrm{d}\theta_Q = r\,\omega_Q$$

Note that the map m_r leaves the distribution D invariant and Tm_r leaves the distribution H invariant. Since $f \in C_k^\infty(D)$ if and only if $m_r^* f = r^k\, f$, we obtain

$$(m_r^t)^*\,\mathrm{d}f = \mathrm{d}(m_r^* f) = \mathrm{d}(r^k\, f) = r^k\ \mathrm{d}f.$$

Therefore for each $u, u' \in D$ we have

$$\begin{aligned} r^k\,\omega(Y_f(u), u') &= \langle \mathrm{d}f(m_r(u)) \mid T_{u'}m_r(u')\rangle \\ &= \omega(m_r(u))(Y_f(m_r(u))) \mid T_{u'}m_r(u')\rangle \\ &= \big((m_r^t)^*\omega\big)(u)(m_r^* Y_f(u), u'), \end{aligned}$$

that is,

$$m_r^* Y_f = r^{k-1}\,Y_f. \tag{97}$$

Consequently,

$$\begin{aligned} m_r^*(\{f,g\})(u) &= \{f,g\}(m_r(u)) = \langle \mathrm{d}g(m_r(u)) \mid Y_f(m_r(u))\rangle \\ &= r^{k-1}\,\langle \mathrm{d}g(m_r(u)) \mid T_u m_r Y_f(u)\rangle, \ \text{using (97) and } f \in C_k^\infty(D) \\ &= r^{k-1}\,\langle (m_r^t)^*(\mathrm{d}g)(u) \mid Y_f(u)\rangle \\ &= r^{k+\ell-1}\,\langle \mathrm{d}g(u) \mid Y_f(u)\rangle, \ \text{because } g \in C_\ell^\infty(D) \\ &= r^{k+\ell-1}\,\{f,g\}(u). \end{aligned}$$

□

1.8.4 *Momenta as coordinates*

Let U be an open subset of Q. With $\varphi^i \in C^\infty(D)$ let $\{\varphi^i\}_{i=1}^n$ be a system of coordinates on U. Let $\{Z_j\}_{j=1}^d$ be smooth vector fields on U such that $Z_j(U) \subseteq D$.

The functions $\{\tau^*\varphi^i\}_{i=1}^n$, where $\tau = \tau_Q|D$, together with the momenta $\{P_j = P_{Z_j}\}_{j=1}^d$ form a *system of coordinates* on $V = \tau^{-1}(U)$ in D if and only if for each $q \in U$ the vectors $\{Z_j(q)\}_{j=1}^d$ form a basis of D_q. The map

$$\lambda : U \times \mathbb{R}^d \to V \subseteq D : (q,c) \mapsto \sum_{j=1}^d c_j\,Z_j(q)$$

is the inverse of a trivialization of the vector bundle $\tau : V \subseteq D \to U \subseteq Q$. Substituting $v = \sum_{j=1}^{d} c_j \, Z_j(q)$ into the definition of the momentum $P_j = P_{Z_j}$ gives the relation

$$P_j(q,v) = P_{Z_j}(q,v) = \mathrm{k}(q)(Z_j(q), v) = \sum_{\ell=1}^{d} \mathrm{k}(q)(Z_j(q), Z_\ell(q))\, c_\ell \tag{98}$$

between the coordinates P_j and c_ℓ on U. Note that $c_j = P_j$ for $1 \le j \le d$ if and only if $\{Z_j(q)\}_{j=1}^{d}$ is a k-orthonormal basis of D_q for every $q \in U$.

In terms of the coordinates $\{\tau^*\varphi^1, \ldots, \tau^*\varphi^n, P_1, \ldots P_n\}$ on $V \subseteq D$ the equations of motion are

$$\dot{\varphi}^i = \sum_{j=1}^{d} (Z_j)^i \, c_j, \text{ for } 1 \le i \le n \tag{99}$$

$$\dot{P}_j = -(L_{P_j} h)|V, \text{ for } 1 \le j \le d. \tag{100}$$

Here equation (99) is the second order equation condition for φ^i, where $(Z_j)^i = \langle \mathrm{d}\varphi^i \mid Z_j \rangle$ are the components of the vector field Z_j with respect to the coordinates $\{\varphi^i\}_{i=1}^{n}$ on $U \subseteq Q$. Solving (98), the c_j's can be expressed in terms of the P_j's. Equation (100) is the momentum equation for Z_j, where the right hand side has to be expressed in terms of φ^i and P_j.

Using the fact that $h = \frac{1}{2}\mathrm{k}(q)(v,v) + V(q)$, where $T = \frac{1}{2}\mathrm{k}(q)(v,v) \in C_2^\infty(D)$ and $V \in C_0^\infty(D)$ and lemma 1.8.3.39, we obtain

$$\dot{\varphi}^i = \{h, \varphi^i\} = \{T, \varphi^i\} \in C_1^\infty(D),$$

which is a $\{\varphi^i\}_{i=1}^{n}$-dependent linear form on D and

$$\dot{P}_j = \{h, P_j\} = \{T, P_j\} + \{V, P_j\},$$

where $\{T, P_j\} \in C_2^\infty(D)$ is a $\{\varphi^i\}_{i=1}^{n}$-dependent quadratic form in $\{P_j\}_{j=1}^{d}$ and $\{V, P_j\} \in C_0^\infty(D)$ is a function of $\{\varphi^i\}_{i=1}^{n}$.

1.9 A projection principle

In this section we describe a *projection method* to obtain the distributional Hamiltonian equations of motion (63).

Equation (55) enables us to decompose $X_f|D$, the restriction to D of the Hamiltonian vector field X_f of f, into its components $X_{f,H}$ and X_{f,H^ω} in H and H^ω, respectively. In other words,

$$X_f \mid_D = X_{f,H} + X_{f,H^\omega}. \tag{101}$$

Taking equations (56), (57) and (59) into account, we see that the distributional Hamiltonian vector field Y_f of f is equal to the H-component of the Hamiltonian vector field X_f of f with respect to the decomposition (101). In other words, $Y_f = X_{f,H}$.

On the other hand, the first equality in (33) allows us to decompose $X_f\mid_D$ into its components X_{f,F^ω} and $X_{f,TD}$ in F^ω and TD, respectively.

Lemma 1.9.40. $X_{f,TD} = Y_f$.

Proof. Since $H \subseteq TD$ and $F^\omega \subseteq H^\omega$, we can intersect the decompositions $T_D(TQ) = F^\omega \oplus TD$ and $T_D(TQ) = H^\omega \oplus H$ to obtain

$$T_D(TQ) = F^\omega \oplus (H^\omega \cap TD) \oplus H. \tag{102}$$

We know that H is a symplectic distribution on D, and $F^\omega \subseteq \ker T\tau_Q$ is isotropic. Since $F^\omega \oplus (H^\omega \cap TD) = H^\omega$ and $\dim F^\omega = \dim(H^\omega \cap TD)$, it follows that F^ω and $H^\omega \cap TD$ are Lagrangian in H^ω. Using the decomposition (102) write $X_f = X_f^1 + X_f^2 + X_f^3$ where X_f^1 has values in F^ω, X_f^2 has values in $H^\omega \cap TD$, and X_f^3 has values in H. In a similar fashion, using (102) we can write

$$\mathrm{d}f = \partial_{F^\omega} f \oplus \partial_{H^\omega \cap TD} f \oplus \partial_H f.$$

Taking into account the decomposition (56) we get

$$\begin{aligned} X_f \lrcorner\, \omega &= (X_f^1 + X_f^2 + X_f^3) \lrcorner\, (\varpi_{H^\omega} \oplus \varpi_H) \\ &= (X_f^1 \lrcorner\, \varpi_{H^\omega}) \oplus (X_f^2 \lrcorner\, \varpi_{H^\omega}) \oplus (X_f^3 \lrcorner\, \varpi_H) \\ &= \partial_{F^\omega} f \oplus \partial_{H^\omega \cap TD} f \oplus \partial_H f. \end{aligned}$$

Since $X_f^1 \lrcorner\, \varpi_{H^\omega}$ annihilates F^ω, and $X_f^2 \lrcorner\, \varpi_{H^\omega}$ annihilates $H^\omega \cap TD$, it follows that

$$X_f^1 \lrcorner\, \varpi_{H^\omega} = \partial_{H^\omega \cap TD} f, \quad X_f^2 \lrcorner\, \varpi_{H^\omega} = \partial_{F^\omega} f, \quad \text{and} \quad X_f^3 \lrcorner\, \varpi_H = \partial_H f.$$

Lemma 1.6.3.18 implies that $\partial_{F^\omega} f = 0$. Hence, $X_f^2 = 0$. Therefore, $X_f^1 = X_{f,H^\omega} = X_{f,F^\omega}$, and $X_f^3 = X_{f,TD} = X_{f,H} = Y_f$. □

Let $\mathcal{P} : T_D(TQ) \to TD$ be the projection along the fibers of F^ω corresponding to the direct sum decomposition $T_D(TQ) = F^\omega \oplus TQ$. The statement of theorem 1.6.3.20 can be reformulated as the following *projection principle.*

Proposition 1.9.41. *A dynamically admissible motion of a distributional Hamiltonian system (D, H, ϖ, h) is the image under the tangent bundle projection τ_Q to Q of an integral curve of $\mathcal{P} \circ X_h\mid_D$.*

1.10 Accessible sets

An abstract structure of a Hamiltonian system with linear nonholonomic constraints is given by a quadruple (D, H, ϖ, h), where D is a manifold, (H, ϖ) is a symplectic distribution on D, and h is a smooth function on D. For each $f \in C^\infty(D)$, the distributional Hamiltonian vector field of f is defined as the unique vector field Y_f on D with values in H satisfying the equation $Y_f \lrcorner\, \varpi = \partial_H f$. By theorem 1.6.3.20 the evolution of our nonholonomically constrained system is given by the distributional Hamiltonian vector Y_h of the Hamiltonian h.

This abstract structure was obtained in §6.3 as the result of an analysis of dynamics of a Hamiltonian system with configuration space Q, Hamiltonian $h \in C^\infty(TQ)$, and linear nonholonomic constraints given by a distribution D on Q. However, once we have obtained the symplectic distribution (H, ϖ) on D we can forget about the linear structure of D and its embedding into TQ, and retain only the manifold structure of D.

In the following we shall need the notion of a *generalized distribution* on D, that is a linear subset H of the tangent bundle TD of D locally spanned by smooth vector fields. For $u \in D$, the number of linearly independent vector fields spanning $H_u \subseteq T_u D$ is called the *rank* of H at u. A distribution is a generalized distribution of constant rank. An *accessible* set of a generalized distribution H on D is the set of points of D that can be connected by piecewise integral curves of vector fields with values in H. Accessible sets of H are also called *reachable* sets of H or *orbits* of the family of local vector fields with values in H.

Theorem 1.10.42. (Sussmann's theorem) *Every accessible set of a generalized distribution on D is an immersed submanifold of D.*

Proof. See [112]. □

Let $L \subseteq D$ be an accessible set of H. According to theorem 10.1 it is an immersed submanifold of D. The restriction H_L of H to points in L is contained in TL. Hence, H_L is a distribution on L. Let ϖ_L be the restriction of ϖ to H_L. By construction ϖ_L is a nondegenerate 2-form on the distribution H_L. For each $f \in C^\infty(D)$, let f_L be the restriction of f to L. Clearly f_L is a smooth function on L. Let Y_{f_L} be the distributional Hamiltonian vector field of f_L on L relative to the symplectic distribution (H_L, ϖ_L). In other words, Y_{f_L} is the unique vector field on L with values in H_L such that

$$Y_{f_L} \lrcorner \varpi_L = \partial_{H_L} f_L. \tag{103}$$

Since Y_h has values in H, it follows that $Y_h(u) = Y_{h_L}(u)$ for all $u \in L$. Hence, L is preserved by the evolution of our nonholonomic Hamiltonian system. This proves

Lemma 1.10.43. *For every accessible set L of H, the nonholonomic Hamiltonian system (D, H, ϖ, h) induces on L the structure of a nonholonomic Hamiltonian system (L, H_L, ϖ_L, h_L).*

Since accessible sets of H form a partition of D, we can write

$$(D, H, \varpi, h) = \bigcup_{L \text{ a.s.} H} (L, H_L, \varpi_L, h_L). \tag{104}$$

Here the union is taken over accessible sets L of H. We say that a nonholonomic Hamiltonian system (D, H, ϖ, h) is *simple* if D is the unique accessible set of H. Since each accessible set L of H is a unique accessible set of H_L, it follows that (104) is the decomposition of (D, H, ϖ, h) into its simple components.

Accessible sets of our symplectic distribution (H, ϖ) on D are important because they provide restrictions on the evolution of the system given by integral curves of the vector field Y_h.

Theorem 1.10.44. (Stefan's theorem) *Accessible sets of a generalized distribution on D defines a structure of a smooth foliation with singularities on D.*

Proof. See [111] □

Taking into account the fact that D is a distribution on Q, the following proposition relates accessible sets of H to accessible sets of D.

Proposition 1.10.45. *Let $M \subseteq Q$ be the unique accessible set of D through $q_0 \in Q$, and let $u_0 \in D_{q_0}$. Then the restriction $D_M = \tau_Q^{-1}(M) \cap D$ of D to points of M is the accessible set of H through u_0.*

Proof. Let X be a vector field on Q with values in D. Consider the function $f \in C^\infty(D)$ defined by $f(u) = \mathrm{k}(X(\tau_Q(u)), u)$ for all $u \in D$. Let $t \mapsto u(t)$ be an integral curve of Y_f with $t \mapsto q(t) = \tau_Q(u(t))$ its projection to Q. Evaluating both sides of the equation $\dot{u}(t) \lrcorner \varpi = \mathrm{d}f(u(t))$ on vectors in

$T_{u(t)}D$ of the form $\iota_{u(t)}v$, where $v \in D_{q(t)}$, and taking into account lemma 1.5.9, we get

$$\begin{aligned} \mathrm{k}(v, T\tau_Q(\dot{u}(t))) &= \langle \dot{u}(t) \,\lrcorner\, \varpi \mid \iota_{u(t)}v\rangle = \langle \mathrm{d}f(u(t)) \mid \iota_{u(t)}v\rangle \\ &= \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} f(u(t)+sv) = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} \mathrm{k}(X(q(t)), u+sv) \\ &= \mathrm{k}(X(q(t)), v), \quad \text{for all } v \in D. \end{aligned}$$

Since $T\tau_Q(\dot{u}(t))$ and $X(q(t))$ lie in D, it follows that $T\tau_Q(\dot{u}(t)) = X(q(t))$. But, $T\tau_Q(\dot{u}(t)) = \dot{q}(t)$. Therefore, $t \mapsto q(t)$ is an integral curve of X.

In theorem 1.6.3.20 we have shown that an integral curve $t \mapsto q(t)$ of a vector field on Q with values in D lifts to a curve $t \mapsto u(t)$, which is an integral curve of a vector field on D with values in H. This shows that if L is the accessible set of H through $u_0 \in D$ and $u \in L$, then the whole fiber $D_{\tau_Q}(u)$ is contained in L. Consequently, if M is the accessible set of D containing $q_0 = \tau_Q(u_0)$ then $D_M = \tau_Q^{-1}(M) \cap D \subseteq L$. On the other hand, the restriction H_{D_M} of H to points in D_M is contained in TD_M. Hence H_{D_M} is a distribution on D_M. Thus the accessible set L of H through $u_0 \in D_M$ is contained in D_M. Therefore, $L = D_M$. □

Corollary 1.10.46. *A nonholonomic system is simple if and only if Q is the unique accessible set of D.*

The upshot of the above discussion is that accessible sets provide restrictions on the evolution of a nonholonomically constrained system which are independent of external forces and depend *only* on the constraint distribution.

1.11 Constants of motion

Additional restrictions on the evolution of a nonholonomically constrained system are provided by *constants of motion.*

Theorem 1.11.47. **(Nonholonomic Noether theorem)** *A function $f \in C^\infty(D)$ is a constant of motion of the distributional Hamiltonian system (D, H, ϖ, h) if and only if its distributional Hamiltonian vector field Y_f preserves the Hamiltonian h.*

Proof. The dynamics on D is given by the distributional Hamiltonian vector field Y_h of the Hamiltonian h. Hence,

$$\dot{f} = \langle \mathrm{d}f \mid Y_h\rangle = \varpi(Y_f, Y_h) = -\varpi(Y_h, Y_f) = -\langle \mathrm{d}h \mid Y_f\rangle.$$

Therefore, $\dot{f} = 0$ if and only if Y_f preserves h. □

Theorem 1.11.47 is a nonholonomic counterpart of the Noether's theorem relating symmetries and conservation laws in unconstrained Hamiltonian systems. In the unconstrained case, a function $f \in C^\infty(TQ)$ is a constant of motion if and only if its Hamiltonian vector field X_f preserves the Hamiltonian. However, X_f also preserves the symplectic form ω. So it is an infinitesimal symmetry of the Hamiltonian system (T^*Q, ω, h). In the presence of constraints, the relationship between symmetries and constants of motion is more involved. The condition that the distributional Hamiltonian vector field Y_f preserves the Hamiltonian h does *not* imply that it is an infinitesimal symmetry of the nonholonomic Hamiltonian system, because it need not preserve either H or ϖ. Conservation laws corresponding to symmetries of a nonholonomically constrained system will be discussed in chapter 3.

Lemma 1.11.48. *Suppose that the nonholonomic system (D, H, ϖ, h) is simple. Then the only Casimir functions of its almost Poisson algebra $(C^\infty(D), \{\ ,\ \})$ are the constant functions.*

We now consider a special case of the conservation law given in theorem 1.11.47. Let Z be a vector field on Q and let $\Phi_t : Q \to Q$ be the local 1-parameter group of local diffeomorphism of Q generated by Z. The tangent map $T\Phi_t : TQ \to TQ$ forms a local group of local diffeomorphisms of TQ generated by the tangent lift Z_{TQ} of Z to TQ. Let P_Z be a function on TQ defined by

$$P_Z(u) = \mathrm{k}(u, Z(\tau_Q(u))), \quad \text{for each } u \in TQ. \tag{105}$$

Proposition 1.11.49. *If Z is a vector field on Q with values in D such that its tangent lift Z_{TQ} preserves the Lagrangian $\ell(u) = \frac{1}{2}\,\mathrm{k}(u,u) - V(\tau_Q(u))$, then P_Z is a constant of motion of the nonholonomically constrained system with Lagrangian ℓ and constraint distribution D.*

Proof. Since Z_{TQ} preserves the Lagrangian ℓ, it follows that it preserves the kinetic energy $k(u) = \frac{1}{2}\,\mathrm{k}(u,u)$ and the potential $V(\tau_Q(u))$, separately. Hence, Z_{TQ} preserves the Hamiltonian $h(u) = \frac{1}{2}\,\mathrm{k}(u,u) + V(\tau_Q(u))$. Moreover, Z preserves the kinetic energy metric k, that is $L_Z\mathrm{k} = 0$. By lemma 1.8.1.33, Z_{TQ} is the Hamiltonian vector field of P_Z defined by equation (105).

Using decomposition (55) we can express the restriction of Z_{TQ} to D as the sum of its components in H and H^ω. In other words, $Z_{TQ}\mid_D=$

$\widetilde{Z}_H + \widetilde{Z}_{H^\omega}$. Moreover, $\widetilde{Z}_H = Y_{P_Z}$ is the distributional Hamiltonian vector field of P_Z. From the fact that Z has values in D, it follows that Z_{TQ} has values in F, and $\widetilde{Z}_{H^\omega}$ has values in $F \cap H^\omega$. Therefore by lemma 1.6.3.18 $\langle \mathrm{d}h \mid \widetilde{Z}_{H^\omega} \rangle = 0$ which implies that $\widetilde{Z}_{H^\omega}$ preserves h. Since Z_{TQ} preserves h, it follows that $Y_{P_Z} = \widetilde{Z}_H = Z_{TQ} - \widetilde{Z}_{H^\omega}$ preserves h. Then theorem 1.11.47 ensures that P_Z is a constant of motion. □

Suppose that we have k conserved functions $f_1, ..., f_k$. For each accessible set L of H and each $c = (c_a) \in \mathbb{R}^k$, consider the level set

$$L_c = \{u \in D \mid f_a(u) = c_a \quad \mathrm{a} = 1, \ldots, k\}.$$

Suppose that locally L_c is a submanifold of L. Since the functions f_a are constants of motion, the restriction of Y_h to L_c is a vector field Y_{h,L_c} on L_c with values in $H_{L_c} = H \cap TL_c$. Let ϖ_{L_c} be the restriction of ϖ to H_{L_c}, and h_{L_c} the restriction of h to L_c. We want to describe the vector field Y_{h,L_c} in terms of the data $(L_c, H_{L_c}, \varpi_{L_c}, h_{L_c})$. Equation (59) restricted to H_{L_c} implies that

$$Y_{h,L_c} \mathbin{\lrcorner} \varpi_{L_c} = \mathrm{d}h_{L_c}. \tag{106}$$

If ϖ_{L_c} is nondegenerate, then Y_{h,L_c} is the distributional Hamiltonian vector field of h_{L_c} relative to a symplectic distribution (H_{L_c}, ϖ_{L_c}) on L_c, which we will denote by $Y_{h_{L_c}}$. Hence, we have proved

Theorem 1.11.50. *Suppose that locally L_c is a submanifold of L and that (H_{L_c}, ϖ_{L_c}) is a symplectic distribution on L_c. Then the evolution of Y_h in L_c is given by the distributional Hamiltonian vector field $Y_{h_{L_c}}$ of h_{L_c} relative to (H_{L_c}, ϖ_{L_c}).*

If the hypotheses of proposition 1.11.49 are satisfied for every accessible set L of H and for every value $c \in \mathbb{R}^d$ of the constants of motions $f_1, ..., f_d$, then the decomposition (104) of the nonholonomic Hamiltonian system (D, H, ϖ, h) admits a refinement, namely,

$$(D, H, \varpi, h) = \bigcup_{L \text{ a.s. } H} \; \bigcup_{c \in \mathbb{R}^d} (L_c, H_{L_c}, \varpi_{L_c}, h_{L_c}). \tag{107}$$

We can refine the above decomposition further by considering accessible sets of H_{L_c}. Let M be an accessible set H_{L_c}. By Sussmann's theorem 1.10.42, M is an immersed submanifold of L_c. The restriction H_M of H to points in M coincides with $H_{L_c} \cap TM$. Let ϖ_M be the restriction of ϖ_{L_c}

to H_M and h_M the restriction of h_{L_c} to M. Applying the decomposition (104) to $(L_c, H_{L_c}, \varpi_{L_c}, h_{L_c})$ we obtain

$$(D, H, \varpi, h) = \bigcup_{L \text{ a.s. } H} \bigcup_{c \in \mathbb{R}^d} \bigcup_{M \text{ a.s. } H_{L_c}} (M, H_M, \varpi_M, h_M). \tag{108}$$

In some examples, every distribution H_{L_c} is involutive and on every integral manifold M of H_{L_c} the 2-form ϖ_M on $H_M = TM$ gives rise to a symplectic form ω_M on M. When this happens, the distributional Hamiltonian systems (M, H_M, ϖ_M, h_M) are Hamiltonian systems (M, ω_M, h_M) and the original distributional Hamiltonian system (D, H, ϖ, h) defines a foliation of D by Hamiltonian systems, namely,

$$(D, H, \varpi, h) = \bigcup_{L \text{ a.s. } H} \bigcup_{c \in \mathbb{R}^d} \bigcup_{M \text{ a.s. } H_{L_c}} (M, \omega_M, h_M). \tag{109}$$

1.12 Notes

We review the standard approach to dynamics of systems with linear nonholonomic constraints. Our only addition is the systematic use of the Levi Civita connection of the kinetic energy metric. We note that hypothesis 2.1, which leads to the Lagrange-d'Alembert principle, has been verified experimentally by Lewis and Murray [67].

We use the name the Lagrange-d'Alembert principle on the basis of an analogy with the formalisms of d'Alembert and Lagrange. But they themselves never applied their formalisms to nonholonomically constrained systems. These applications are definitely of a later date, namely the second half of the 19$^{\text{th}}$ century with a lot of confusion, which was only cleared up at the end of that century.

Here is a brief account of the history. Whittaker [118] refers to Ferrers [43] as the first one who wrote down the equations of motion for a nonholonomically constrained system. But we do not understand what Ferrers meant. In §240 Routh [96] uses the Lagrange-d'Alembert principle in our form. Then Lindelöf [70] and Appell in the first edition of [4] and others used, in the case of a translation invariant nonholonomically constrained system, the *wrong* principle, namely, they take the Lagrangian on the tangent bundle of the orbit space of the translation symmetry; and then apply the Euler-Lagrange equation. Both Chaplygin [23], who refers to Lindelöf [70], and Korteweg [62], who refers to the first edition of Appell [4], correct this error. Appell corrected his mistake in the second edition of [4]. For a more detailed discussion of this point see §4 of chapter 3. From this time on the Lagrange-d'Alembert principle has been accepted as giving the "right" equations of motion.

More recently Arnol'd et al. [7] has introduced vakonomic mechanics which have nonholonomic constraints but the equations of motion are based on a variational principle. Vakonomic systems are of interest, but rolling without slipping is *not* vakonomic.

For an up to date treatment of *nonlinearly* nonholonomically constrained systems see Marle [74].

Our presentation of the distributional Hamiltonian formulation of nonholonomically constrained systems follows the theory developed in [13] and [32]. It is a special case of the partially symplectic formulation proposed by Bocharov and Vinogradov [17].

Dalsmo and van der Schaft [36] placed nonholonomically constrained systems in the context of Dirac structures. We recall the definition of a Dirac structure. The Pontryagin bundle of a manifold M is the direct sum of its tangent and cotangent bundles, that is, $TM \oplus_M T^*M$. We endow the total space of the Pontryagin bundle with a nondegenerate symmetric bilinear form of signature (dim$M$,dim$M$) given by

$$\langle (u,p),(v,q)\rangle = \langle q \mid u\rangle + \langle p \mid v\rangle$$

for every $(u,p),(v,q) \in T_mM \oplus T^*_mM$. A Dirac structure on a manifold M is a subbundle B of the $TM \oplus T^*M$ which is maximal isotropic with respect of the above bilinear form, see Courant [29]. We now show that the distributional Hamiltonian system (D, ϖ, H, h) with nonholonomic constraint distribution $D \subset TQ$, Hamiltonian h, and symplectic distribution (H, ϖ) on D is a Dirac structure. Here the symplectic form ϖ is the restriction to H of the symplectic form ω on TQ, which is obtained by pulling back the canonical symplectic form on T^*Q by the Legendre transformation $\mathrm{k}^\flat : TQ \to T^*Q$ given by the kinetic energy metric k of the system. Following Yoshimura and Marsden [119] the Dirac structure corresponding to (D, ϖ, H, h) is given by

$$\{(u,p) \in TD \oplus_D T^*D \,\big|\, u \in H \text{ and } p - u \lrcorner\, \omega \in H^0 \}.$$

Here H^0 is the annihilator of H in T^*D. The pair (Y_h, dh), where $Y_h \lrcorner\, \varpi = \partial_H h$, is a section of this Dirac structure.

The projection principle in the text is due to Marle [73]. However, projection principles were first used by Gibbs [47] and by Appell in the second edition [4] to derive what Pars [89] calls the Gibbs-Appell equations of motion for a nonholonomically constrained system.

Van der Schaft and Maschke [113] introduced an almost Poisson bracket to study nonholonomically constrained Hamiltonian systems. Equation (87) is the nonholonomic almost Poisson bracket analogue of the formula for the Dirac bracket given in [39]. The intrinsic relation between the almost Poisson bracket and the symplectic distribution (H, ϖ) was given by Koon and Marsden [61]. The term *almost Poisson bracket* was coined by Cantrijn, de León, and de Diego [19].

Results on accessible sets of distributions used here are taken from papers of Stefan [111] and Sussmann [112].

In [103] there is a proof of a nonholonomic version of Noether's theorem of §11. A foliation of the constraint manifold by Hamiltonian systems given by constants of motion was discovered by Kemppainen in the dynamics of a rolling disk [57].

Bloch, Krishnaprasad, Marsden and Murray establish the momentum equation in [16], but only for G-invariant vector fields which were tangent to G-orbits on configuration space Q under the hypothesis that the action of G on Q is free and proper. The version in the text is based on [105].