8
Variational Principles, Constraints, and Rotating Systems

This chapter deals with two related topics: constrained Lagrangian (and Hamiltonian) systems and rotating systems. Constrained systems are illustrated by a particle constrained to move on a sphere. Such constraints that involve conditions on the *configuration* variables are called "holonomic."[1] For rotating systems, one needs to distinguish systems that are viewed from rotating coordinate systems (passively rotating systems) and systems that themselves are rotated (actively rotating systems—such as a Foucault pendulum and weather systems rotating with the Earth). We begin with a more detailed look at variational principles, and then we turn to a version of the Lagrange multiplier theorem that will be useful for our analysis of constraints.

8.1 A Return to Variational Principles

In this section we take a closer look at variational principles. Technicalities involving infinite-dimensional manifolds prevent us from presenting the full story from that point of view. For these, we refer to, for example, Smale [1964], Palais [1968], and Klingenberg [1978]. For the classical geometric theory without the infinite-dimensional framework, the reader may consult,

[1] In this volume we shall not discuss "nonholonomic" constraints such as rolling constraints. We refer to Bloch, Krishnaprasad, Marsden, and Murray [1996], Koon and Marsden [1997b], and Zenkov, Bloch, and Marsden [1998] for a discussion of nonholonomic systems and further references.

for example, Bolza [1973], Whittaker [1927], Gelfand and Fomin [1963], or Hermann [1968].

Hamilton's Principle. We begin by setting up the space of paths joining two points.

Definition 8.1.1. *Let Q be a manifold and let $L : TQ \to \mathbb{R}$ be a regular Lagrangian. Fix two points q_1 and q_2 in Q and an interval $[a, b]$, and define the **path space** from q_1 to q_2 by*

$$\begin{aligned}&\Omega(q_1, q_2, [a, b]) \\ &\qquad = \{\, c : [a, b] \to Q \mid c \text{ is a } C^2 \text{ curve}, \ c(a) = q_1, \ c(b) = q_2 \,\} \qquad (8.1.1)\end{aligned}$$

and the map $\mathfrak{S} : \Omega(q_1, q_2, [a, b]) \to \mathbb{R}$ by

$$\mathfrak{S}(c) = \int_a^b L(c(t), \dot{c}(t))\, dt.$$

What we shall *not* prove is that $\Omega(q_1, q_2, [a, b])$ is a smooth infinite-dimensional manifold. This is a special case of a general result in the topic of manifolds of mappings, wherein spaces of maps from one manifold to another are shown to be smooth infinite-dimensional manifolds. Accepting this, we can prove the following.

Proposition 8.1.2. *The tangent space $T_c\Omega(q_1, q_2, [a, b])$ to the manifold $\Omega(q_1, q_2, [a, b])$ at a point, that is, a curve $c \in \Omega(q_1, q_2, [a, b])$, is the set of C^2 maps $v : [a, b] \to TQ$ such that $\tau_Q \circ v = c$ and $v(a) = 0$, $v(b) = 0$, where $\tau_Q : TQ \to Q$ denotes the canonical projection.*

Proof. The tangent space to a manifold consists of tangents to smooth curves in the manifold. The tangent vector to a curve $c_\lambda \in \Omega(q_1, q_2, [a, b])$ with $c_0 = c$ is

$$v = \left.\frac{d}{d\lambda} c_\lambda \right|_{\lambda=0}. \qquad (8.1.2)$$

However, $c_\lambda(t)$, for each fixed t, is a curve through $c_0(t) = c(t)$. Hence

$$\left.\frac{d}{d\lambda} c_\lambda(t) \right|_{\lambda=0}$$

is a tangent vector to Q based at $c(t)$. Hence $v(t) \in T_{c(t)}Q$; that is, $\tau_Q \circ v = c$. The restrictions $c_\lambda(a) = q_1$ and $c_\lambda(b) = q_2$ lead to $v(a) = 0$ and $v(b) = 0$, but otherwise v is an arbitrary C^2 function. ■

One refers to v as an ***infinitesimal variation*** of the curve c subject to fixed endpoints, and we use the notation $v = \delta c$. See Figure 8.1.1.

Now we can state and sketch the proof of a main result in the calculus of variations in a form due to Hamilton [1834].

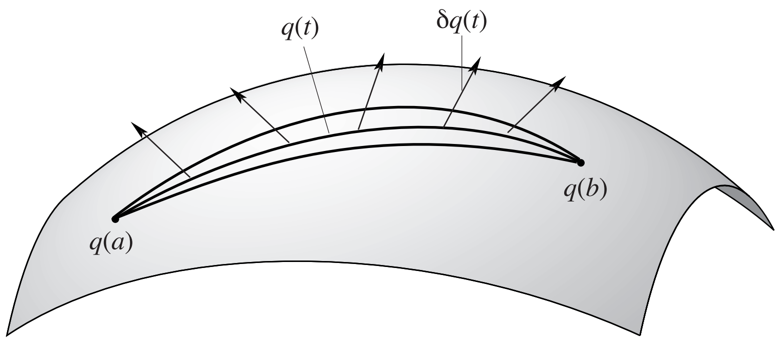

FIGURE 8.1.1. The variation $\delta q(t)$ of a curve $q(t)$ is a field of vectors tangent to the configuration manifold along that curve.

Theorem 8.1.3 (Variational Principle of Hamilton). *Let L be a Lagrangian on TQ. A curve $c_0 : [a,b] \to Q$ joining $q_1 = c_0(a)$ to $q_2 = c_0(b)$ satisfies the Euler–Lagrange equations*

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) = \frac{\partial L}{\partial q^i} \tag{8.1.3}$$

if and only if c_0 is a critical point of the function $\mathfrak{S} : \Omega(q_1, q_2, [a,b]) \to \mathbb{R}$, that is, $\mathbf{d}\mathfrak{S}(c_0) = 0$. If L is regular, either condition is equivalent to c_0 being a base integral curve of X_E.

As in §7.1, the condition $\mathbf{d}\mathfrak{S}(c_0) = 0$ is denoted by

$$\delta \int_a^b L(c_0(t), \dot{c}_0(t))\, dt = 0; \tag{8.1.4}$$

that is, the integral is stationary when it is differentiated with c regarded as the independent variable.

Proof. We work out $\mathbf{d}\mathfrak{S}(c) \cdot v$ just as in §7.1. Write v as the tangent to the curve c_λ in $\Omega(q_1, q_2, [a,b])$ as in (8.1.2). By the chain rule,

$$\mathbf{d}\mathfrak{S}(c) \cdot v = \left.\frac{d}{d\lambda}\mathfrak{S}(c_\lambda)\right|_{\lambda=0} = \left.\frac{d}{d\lambda}\int_a^b L(c_\lambda(t), \dot{c}_\lambda(t))\, dt\right|_{\lambda=0}. \tag{8.1.5}$$

Differentiating (8.1.5) under the integral sign, and using local coordinates,[2] we get

$$\mathbf{d}\mathfrak{S}(c) \cdot v = \int_a^b \left(\frac{\partial L}{\partial q^i} v^i + \frac{\partial L}{\partial \dot{q}^i}\dot{v}^i\right) dt. \tag{8.1.6}$$

[2]If the curve $c_0(t)$ does not lie in a single coordinate chart, divide the curve $c(t)$ into a finite partition each of whose elements lies in a chart and apply the argument below.

Since v vanishes at both ends, the second term in (8.1.6) can be integrated by parts to give

$$\mathbf{d}\mathfrak{S}(c) \cdot v = \int_a^b \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} \right) v^i \, dt. \tag{8.1.7}$$

Now, $\mathbf{d}\mathfrak{S}(c) = 0$ means that $\mathbf{d}\mathfrak{S}(c) \cdot v = 0$ for all $v \in T_c\Omega(q_1, q_2, [a, b])$. This holds if and only if

$$\frac{\partial L}{\partial q^i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) = 0, \tag{8.1.8}$$

since the integrand is continuous and v is arbitrary, except for $v = 0$ at the ends. (This last assertion was proved in Theorem 7.3.3.) ■

The reader can check that Hamilton's principle proceeds virtually unchanged for time-dependent Lagrangians. We shall use this remark below.

The Principle of Critical Action. Next we discuss variational principles with the constraint of constant energy imposed. To compensate for this constraint, we let the interval $[a, b]$ be variable.

Definition 8.1.4. *Let L be a regular Lagrangian and let Σ_e be a regular energy surface for the energy E of L, that is, e is a regular value of E and $\Sigma_e = E^{-1}(e)$. Let $q_1, q_2 \in Q$ and let $[a, b]$ be a given interval. Define $\Omega(q_1, q_2, [a, b], e)$ to be the set of pairs (τ, c), where $\tau : [a, b] \to \mathbb{R}$ is C^2, satisfies $\dot{\tau} > 0$, and where $c : [\tau(a), \tau(b)] \to Q$ is a C^2 curve with*

$$c(\tau(a)) = q_1, \quad c(\tau(b)) = q_2,$$

and

$$E\left(c(\tau(t)), \dot{c}(\tau(t))\right) = e, \qquad \text{for all } t \in [a, b].$$

Arguing as in Proposition 8.1.2, computation of the derivatives of curves $(\tau_\lambda, c_\lambda)$ in $\Omega(q_1, q_2, [a, b], e)$ shows that the tangent space to $\Omega(q_1, q_2, [a, b], e)$ at (τ, c) consists of the space of pairs of C^2 maps

$$\alpha : [a, b] \to \mathbb{R} \quad \text{and} \quad v : [\tau(a), \tau(b)] \to TQ$$

such that $v(t) \in T_{c(t)}Q$,

$$\begin{aligned} \dot{c}(\tau(a))\alpha(a) + v(\tau(a)) = 0, \\ \dot{c}(\tau(b))\alpha(b) + v(\tau(b)) = 0, \end{aligned} \tag{8.1.9}$$

and

$$\mathbf{d}E[c(\tau(t)), \dot{c}(\tau(t))] \cdot [\dot{c}(\tau(t))\alpha(t) + v(\tau(t)), \ddot{c}(\tau(t))\dot{\alpha}(t) + \dot{v}(\tau(t))] = 0. \tag{8.1.10}$$

Theorem 8.1.5 (Principle of Critical Action)**.** *Let $c_0(t)$ be a solution of the Euler–Lagrange equations and let $q_1 = c_0(a)$ and $q_2 = c_0(b)$. Let e be the energy of $c_0(t)$ and assume that it is a regular value of E. Define the map $\mathcal{A} : \Omega(q_1, q_2, [a, b], e) \to \mathbb{R}$ by*

$$\mathcal{A}(\tau, c) = \int_{\tau(a)}^{\tau(b)} A(c(t), \dot{c}(t))\, dt, \tag{8.1.11}$$

where A is the action of L. Then

$$\mathbf{d}\mathcal{A}(\mathrm{Id}, c_0) = 0, \tag{8.1.12}$$

where Id *is the identity map. Conversely, if* (Id, c_0) *is a critical point of $\mathcal{A}$ and c_0 has energy e, a regular value of E, then c_0 is a solution of the Euler–Lagrange equations.*

In coordinates, (8.1.11) reads

$$\mathcal{A}(\tau, c) = \int_{\tau(a)}^{\tau(b)} \frac{\partial L}{\partial \dot{q}^i} \dot{q}^i\, dt = \int_{\tau(a)}^{\tau(b)} p_i\, dq^i, \tag{8.1.13}$$

the integral of the canonical one-form along the curve $\gamma = (c, \dot{c})$. Being the line integral of a one-form, $\mathcal{A}(\tau, c)$ is independent of the parametrization τ. Thus, one may think of $\mathcal{A}$ as defined on the space of (unparametrized) curves joining q_1 and q_2.

Proof. If the curve c has energy e, then

$$\mathcal{A}(\tau, c) = \int_{\tau(a)}^{\tau(b)} [L(q^i, \dot{q}^i) + e]\, dt.$$

Differentiating $\mathcal{A}$ with respect to τ and c by the method of Theorem 8.1.3 gives

$$\begin{aligned} &\mathbf{d}\mathcal{A}(\mathrm{Id}, c_0) \cdot (\alpha, v) \\ &\quad = \alpha(b)\,[L(c_0(b), \dot{c}_0(b)) + e] - \alpha(a)\,[L(c_0(a), \dot{c}_0(a)) + e] \\ &\qquad + \int_a^b \left(\frac{\partial L}{\partial q^i}(c_0(t), \dot{c}_0(t)) v^i(t) + \frac{\partial L}{\partial \dot{q}^i}(c_0(t), \dot{c}_0(t)) \dot{v}^i(t) \right) dt. \end{aligned} \tag{8.1.14}$$

Integrating by parts gives

$$\begin{aligned} &\mathbf{d}\mathcal{A}(\mathrm{Id}, c_0) \cdot (\alpha, v) \\ &\quad = \left[\alpha(t)\,[L(c_0(t), \dot{c}_0(t)) + e] + \frac{\partial L}{\partial \dot{q}^i}(c_0(t), \dot{c}_0(t)) v^i(t) \right]_a^b \\ &\qquad + \int_a^b \left(\frac{\partial L}{\partial q^i}(c_0(t), \dot{c}_0(t)) - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i}(c_0(t), \dot{c}_0(t)) \right) v^i(t)\, dt. \end{aligned} \tag{8.1.15}$$

Using the boundary conditions $v = -\dot{c}\alpha$, noted in the description of the tangent space $T_{(\mathrm{Id},c_0)}\Omega(q_1, q_2, [a,b], e)$ and the energy constraint $(\partial L/\partial \dot{q}^i)\dot{c}^i - L = e$, the boundary terms cancel, leaving

$$\mathbf{d}\mathcal{A}(\mathrm{Id}, c_0) \cdot (\alpha, v) = \int_a^b \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} \right) v^i \, dt. \tag{8.1.16}$$

However, we can choose v arbitrarily; notice that the presence of α in the linearized energy constraint means that no restrictions are placed on the variations v^i on the open set where $\dot{c} \neq 0$. The result therefore follows. ■

If $L = K - V$, where K is the kinetic energy of a Riemannian metric, then Theorem 8.1.5 states that a curve c_0 is a solution of the Euler–Lagrange equations if and only if

$$\delta_e \int_a^b 2K(c_0, \dot{c}_0)\, dt = 0, \tag{8.1.17}$$

where δ_e indicates a variation holding the energy and endpoints but not the parametrization fixed; this is symbolic notation for the precise statement in Theorem 8.1.5. Using the fact that $K \geq 0$, a calculation of the Euler–Lagrange equations (Exercise 8.1-3) shows that (8.1.17) is the same as

$$\delta_e \int_a^b \sqrt{2K(c_0, \dot{c}_0)}\, dt = 0, \tag{8.1.18}$$

that is, arc length is extremized (subject to constant energy). This is ***Jacobi's form of the principle of "least action"*** and represents a key to linking mechanics and geometric optics, which was one of Hamilton's original motivations. In particular, geodesics are characterized as extremals of arc length. Using the Jacobi metric (see §7.7) one gets yet another variational principle.[3]

Phase Space Form of the Variational Principle. The above variational principles for Lagrangian systems carry over to some extent to Hamiltonian systems.

Theorem 8.1.6 (Hamilton's Principle in Phase Space)**.** *Consider a Hamiltonian H on a given cotangent bundle T^*Q. A curve $(q^i(t), p_i(t))$ in T^*Q satisfies Hamilton's equations iff*

$$\delta \int_a^b [p_i \dot{q}^i - H(q^i, p_i)]\, dt = 0 \tag{8.1.19}$$

for variations over curves $(q^i(t), p_i(t))$ in phase space, where $\dot{q}^i = dq^i/dt$ and where q^i are fixed at the endpoints.

---

[3]Other interesting variational principles are those of Gauss, Hertz, Gibbs, and Appell. A modern account, along with references, is Lewis [1996].

Proof. Computing as in (8.1.6), we find that

$$\delta \int_a^b [p_i \dot{q}^i - H(q^i, p_i)]\, dt = \int_a^b \left[(\delta p_i)\dot{q}^i + p_i(\delta \dot{q}^i) - \frac{\partial H}{\partial q^i}\delta q^i - \frac{\partial H}{\partial p_i}\delta p_i \right] dt. \tag{8.1.20}$$

Since $q^i(t)$ are fixed at the two ends, we have $p_i \delta q^i = 0$ at the two ends, and hence the second term of (8.1.20) can be integrated by parts to give

$$\int_a^b \left[\dot{q}^i(\delta p_i) - \dot{p}_i(\delta q^i) - \frac{\partial H}{\partial q^i}\delta q^i - \frac{\partial H}{\partial p_i}\delta p_i \right] dt, \tag{8.1.21}$$

which vanishes for all $\delta p_i, \delta q^i$ exactly when Hamilton's equations hold. ∎

Hamilton's principle in phase space (8.1.19) on an exact symplectic manifold $(P, \Omega = -\mathbf{d}\Theta)$ reads

$$\delta \int_a^b (\Theta - H dt) = 0, \tag{8.1.22}$$

again with suitable boundary conditions. Likewise, if we impose the constraint $H =$ constant, the principle of least action reads

$$\delta \int_{\tau(a)}^{\tau(b)} \Theta = 0. \tag{8.1.23}$$

In Cendra and Marsden [1987], Cendra, Ibort, and Marsden [1987], Marsden and Scheurle [1993a, 1993b], Holm, Marsden, and Ratiu [1998a] and Marsden, Ratiu, and Scheurle [2000] and Cendra, Marsden, and Ratiu [2001], it is shown how to form variational principles on certain symplectic manifolds for which the symplectic form Ω is not exact and even on some Poisson manifolds that arise by a reduction process. The variational principle for the Euler–Poincaré equations that was described in the introduction and that we shall encounter again in Chapter 13 is a special instance of this.

The one-form $\Theta_H := \Theta - H dt$ in (8.1.22), regarded as a one-form on $P \times \mathbb{R}$, is an example of a ***contact form*** and plays an important role in time-dependent and relativistic mechanics. Let

$$\Omega_H = -\mathbf{d}\Theta_H = \Omega + dH \wedge dt$$

and observe that the vector field X_H is characterized by the statement that its suspension $\tilde{X}_H = (X_H, 1)$, a vector field on $P \times \mathbb{R}$, lies in the kernel of Ω_H:

$$\mathbf{i}_{\tilde{X}_H} \Omega_H = 0.$$

Exercises

◇ **8.1-1.** In Hamilton's principle, show that the boundary conditions of fixed $q(a)$ and $q(b)$ can be changed to $p(b) \cdot \delta q(b) = p(a) \cdot \delta q(a)$. What is the corresponding statement for Hamilton's principle in phase space?

◇ **8.1-2.** Show that the equations for a particle in a magnetic field B and a potential V can be written as

$$\delta \int (K - V)\, dt = -\frac{e}{c} \int \delta q \cdot (v \times B)\, dt.$$

◇ **8.1-3.** Do the calculation showing that

$$\delta_e \int_a^b 2K(c_0, \dot{c}_0)\, dt = 0$$

and

$$\delta_e \int_a^b \sqrt{2K(c_0, \dot{c}_0)}\, dt = 0$$

are equivalent.

8.2 The Geometry of Variational Principles

In Chapter 7 we derived the "geometry" of Lagrangian systems on TQ by pulling back the geometry from the Hamiltonian side on T^*Q. Now we show how *all of this basic geometry of Lagrangian systems can be derived directly from Hamilton's principle*. The exposition below follows Marsden, Patrick, and Shkoller [1998].

A Brief Review. Recall that given a Lagrangian function $L : TQ \to \mathbb{R}$, we construct the corresponding ***action functional*** $\mathfrak{S}$ on C^2 curves $q(t)$, $a \le t \le b$, by (using coordinate notation)

$$\mathfrak{S}\big(q(\cdot)\big) \equiv \int_a^b L\left(q^i(t), \frac{dq^i}{dt}(t)\right) dt. \tag{8.2.1}$$

Hamilton's principle (Theorem 8.1.3) seeks the curves $q(t)$ for which the functional $\mathfrak{S}$ is stationary under variations of $q^i(t)$ with *fixed endpoints* at *fixed times*. Recall that this calculation gives

$$\mathbf{d}\mathfrak{S}\big(q(\cdot)\big) \cdot \delta q(\cdot) = \int_a^b \delta q^i \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right) dt + \frac{\partial L}{\partial \dot{q}^i}\delta q^i \bigg|_a^b . \tag{8.2.2}$$

The last term in (8.2.2) vanishes, since $\delta q(a) = \delta q(b) = 0$, so that the requirement that $q(t)$ be stationary for $\mathfrak{S}$ yields the Euler–Lagrange equations

$$\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} = 0. \tag{8.2.3}$$

Recall that L is called ***regular*** when the matrix $[\partial^2 L/\partial \dot{q}^i \partial \dot{q}^j]$ is everywhere nonsingular, and in this case the Euler–Lagrange equations are second-order ordinary differential equations for the required curves.

Since the action (8.2.1) is independent of the choice of coordinates, the Euler–Lagrange equations are coordinate-independent as well. Consequently, it is natural that the Euler–Lagrange equations may be intrinsically expressed using the language of differential geometry.

Recall that one defines the ***canonical 1-form*** Θ on the $2n$-dimensional cotangent bundle T^*Q of Q by

$$\Theta(\alpha_q) \cdot w_{\alpha_q} = \langle \alpha_q, T_{\alpha_q} \pi_Q(w_{\alpha_q}) \rangle,$$

where $\alpha_q \in T_q^*Q$, $w_{\alpha_q} \in T_{\alpha_q} T^*Q$, and $\pi_Q : T^*Q \to Q$ is the projection. The Lagrangian L defines a fiber-preserving bundle map $\mathbb{F}L : TQ \to T^*Q$, the Legendre transformation, by fiber differentiation:

$$\mathbb{F}L(v_q) \cdot w_q = \left. \frac{d}{d\epsilon} \right|_{\epsilon=0} L(v_q + \epsilon w_q).$$

One normally defines the ***Lagrange 1-form*** on TQ by pull-back,

$$\Theta_L = \mathbb{F}L^* \Theta,$$

and the ***Lagrange 2-form*** by $\Omega_L = -\mathbf{d}\Theta_L$. We then seek a vector field X_E (called the ***Lagrange vector field***) on TQ such that $X_E \,\lrcorner\, \Omega_L = \mathbf{d}E$, where the ***energy*** E is defined by

$$E(v_q) = \langle \mathbb{F}L(v_q), v_q \rangle - L(v_q) = \Theta_L(X_E)(v_q) - L(v_q).$$

If $\mathbb{F}L$ is a local diffeomorphism, which is equivalent to L being regular, then X_E exists and is unique, and its integral curves solve the Euler–Lagrange equations. The Euler–Lagrange equations are second-order equations in TQ. In addition, the flow F_t of X_E is symplectic, that is, preserves Ω_L: $F_t^* \Omega_L = \Omega_L$. These facts were proved using differential forms and Lie derivatives in the last three chapters.

The Variational Approach. Besides being more faithful to history, sometimes there are advantages to staying on the "Lagrangian side." Many examples can be given, but the theory of Lagrangian reduction (the Euler-Poincaré equations being an instance) is one example. Other examples are the direct variational approach to questions in black-hole dynamics given by Wald [1993] and the development of variational asymptotics (see Holm [1996], Holm, Marsden, and Ratiu [1998b], and references therein). In such studies, it is the variational principle that is the center of attention.

The development begins by removing the endpoint condition $\delta q(a) = \delta q(b) = 0$ from (8.2.2) but still keeping the time interval fixed. Equation (8.2.2) becomes

$$\mathbf{d}\mathfrak{S}\big(q(\cdot)\big) \cdot \delta q(\cdot) = \int_a^b \delta q^i \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} \right) dt + \left. \frac{\partial L}{\partial \dot{q}^i} \delta q^i \right|_a^b, \qquad (8.2.4)$$

but now the left side operates on more general δq, and correspondingly, the last term on the right side need not vanish. That last term of (8.2.4) is a linear pairing of the function $\partial L/\partial \dot{q}^i$, a function of q^i and $\dot{q}^i$, with the tangent vector δq^i. Thus, one may consider it a 1-form on TQ, namely, the Lagrange 1-form $(\partial L/\partial \dot{q}^i)dq^i$.

Theorem 8.2.1. *Given a C^k Lagrangian L, $k \geq 2$, there exists a unique C^{k-2} mapping $D_{EL}L : \ddot{Q} \to T^*Q$, defined on the* ***second-order submanifold***

$$\ddot{Q} := \left\{ \frac{d^2q}{dt^2}(0) \in T(TQ) \,\middle|\, q \text{ is a } C^2 \text{ curve in } Q \right\}$$

of $T(TQ)$, and a unique C^{k-1} 1-form Θ_L on TQ, such that for all C^2 variations $q_\epsilon(t)$ (on a fixed t-interval) of $q(t)$, where $q_0(t) = q(t)$, we have

$$\mathbf{d}\mathfrak{S}\big(q(\cdot)\big) \cdot \delta q(\cdot) = \int_a^b D_{EL}L\left(\frac{d^2q}{dt^2}\right) \cdot \delta q \, dt + \Theta_L\left(\frac{dq}{dt}\right) \cdot \hat{\delta q}\Bigg|_a^b, \qquad (8.2.5)$$

where

$$\delta q(t) = \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} q_\epsilon(t), \qquad \hat{\delta q}(t) = \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \frac{d}{dt} q_\epsilon(t).$$

The 1-form so defined is a called the ***Lagrange 1-form.***

Indeed, uniqueness and local existence follow from the calculation (8.2.2). The coordinate independence of the action implies the global existence of D_{EL} and the 1-form Θ_L.

Thus, using the variational principle, the Lagrange 1-form Θ_L is the "boundary part" of the functional derivative of the action when the boundary is varied. The analogue of the symplectic form is the negative exterior derivative of Θ_L; that is, $\Omega_L \equiv -\mathbf{d}\Theta_L$.

Lagrangian Flows Are Symplectic. One of Lagrange's basic discoveries was that the solutions of the Euler–Lagrange equations give rise to a symplectic map. It is a curious twist of history that he did this without the machinery of differential forms, the Hamiltonian formalism, or Hamilton's principle itself.

Assuming that L is regular, the variational principle gives coordinate-independent second-order ordinary differential equations. We temporarily denote the vector field on TQ so obtained by X, and its flow by F_t. Now consider the restriction of $\mathfrak{S}$ to the subspace $\mathcal{C}_L$ of solutions of the variational principle. The space $\mathcal{C}_L$ may be identified with the initial conditions for the flow; to $v_q \in TQ$ we associate the integral curve $s \mapsto F_s(v_q)$, $s \in [0, t]$. The value of $\mathfrak{S}$ on the base integral curve $q(s) = \pi_Q(F_s(v_q))$ is denoted by $\mathfrak{S}_t$, that is,

$$\mathfrak{S}_t = \int_0^t L(F_s(v_q))\,ds, \tag{8.2.6}$$

which is again called the ***action***. We regard $\mathfrak{S}_t$ as a real-valued function on TQ. Note that by (8.2.6), $d\mathfrak{S}_t/dt = L(F_t(v_q))$. The fundamental equation (8.2.5) becomes

$$\mathbf{d}\mathfrak{S}_t(v_q)\cdot w_{v_q} = \Theta_L\big(F_t(v_q)\big)\cdot \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} F_t(v_q+\epsilon w_{v_q}) - \Theta_L(v_q)\cdot w_{v_q},$$

where $\epsilon \mapsto v_q + \epsilon w_{v_q}$ symbolically represents a curve at v_q in TQ with derivative w_{v_q}. Note that the first term on the right-hand side of (8.2.5) vanishes, since we have restricted $\mathfrak{S}$ to solutions. The second term becomes the one stated, remembering that now $\mathfrak{S}_t$ is regarded as a function on TQ. We have thus derived the equation

$$\mathbf{d}\mathfrak{S}_t = F_t^*\Theta_L - \Theta_L. \tag{8.2.7}$$

Taking the exterior derivative of (8.2.7) yields the fundamental fact that the flow of X is symplectic:

$$0 = \mathbf{dd}\mathfrak{S}_t = \mathbf{d}(F_t^*\Theta_L - \Theta_L) = -F_t^*\Omega_L + \Omega_L,$$

which is equivalent to $F_t^*\Omega_L = \Omega_L$. *Thus, using the variational principle, the analogue that the evolution is symplectic is the equation* $\mathbf{d}^2 = 0$*, applied to the action restricted to the space of solutions of the variational principle.* Equation (8.2.7) also provides the differential–geometric equations for X. Indeed, taking one time-derivative of (8.2.7) gives $\mathbf{d}L = \pounds_X\Theta_{\mathcal{L}}$, so that

$$X \mathbin{\lrcorner} \Omega_L = -X \mathbin{\lrcorner} \mathbf{d}\Theta_L = -\pounds_X\Theta_L + \mathbf{d}(X \mathbin{\lrcorner} \Theta_L) = \mathbf{d}(X \mathbin{\lrcorner} \Theta_L - L) = \mathbf{d}E,$$

where we define $E = X \mathbin{\lrcorner} \Theta_L - L$. Thus, quite naturally, we find that $X = X_E$.

The Hamilton–Jacobi Equation. Next, we give a derivation of the Hamilton–Jacobi equation from variational principles. *Allowing* L *to be time-dependent*, Jacobi [1866] showed that the ***action integral*** defined by

$$S(q^i, \overline{q}^i, t) = \int_{t_0}^t L(q^i(s), \dot{q}^i(s), s)\,ds,$$

where $q^i(s)$ is the solution of the Euler–Lagrange equation subject to the conditions $q^i(t_0) = \overline{q}^i$ and $q^i(t) = q^i$, satisfies the Hamilton–Jacobi equation. There are several implicit assumptions in Jacobi's argument: L is regular and the time $|t - t_0|$ is assumed to be small, so that by the convex neighborhood theorem, S is a well-defined function of the endpoints. We can allow $|t - t_0|$ to be large as long as the solution $q(t)$ is near a nonconjugate solution.

Theorem 8.2.2 (Hamilton–Jacobi). *With the above assumptions, the function $S(q, \overline{q}, t)$ satisfies the Hamilton–Jacobi equation:*

$$\frac{\partial S}{\partial t} + H\left(q, \frac{\partial S}{\partial q}, t\right) = 0.$$

Proof. In this equation, $\overline{q}$ is held fixed. Define v, a tangent vector at $\overline{q}$, implicitly by

$$\pi_Q F_t(v) = q, \tag{8.2.8}$$

where $F_t : TQ \to TQ$ is the flow of the Euler–Lagrange equations, as in Theorem 7.4.5. As before, identifying the space of solutions $\mathcal{C}_L$ of the Euler–Lagrange equations with the set of initial conditions, which is TQ, we regard

$$\mathfrak{S}_t(v_q) := S(q, \overline{q}, t) := \int_0^t L(F_s(v_q), s)\, ds \tag{8.2.9}$$

as a real-valued function on TQ. Thus, by the chain rule and our previous calculations for $\mathfrak{S}_t$ (see (8.2.7)), equation (8.2.9) gives

$$\begin{aligned}\frac{\partial S}{\partial t} &= \frac{\partial \mathfrak{S}_t}{\partial t} + \mathbf{d}\mathfrak{S}_t \cdot \frac{\partial v}{\partial t} \\ &= L(F_t(v), t) + (F_t^* \Theta_L)\left(\frac{\partial v}{\partial t}\right) - \Theta_L\left(\frac{\partial v}{\partial t}\right),\end{aligned} \tag{8.2.10}$$

where $\partial v/\partial t$ is computed by keeping $\overline{q}$ and q fixed and only changing t. Notice that in (8.2.10), q and $\overline{q}$ are held fixed on both sides of the equation; $\partial S/\partial t$ is a *partial* and *not* a total time-derivative.

Implicitly differentiating the defining condition (8.2.8) with respect to t gives

$$T\pi_Q \cdot X_E(F_t(v)) + T\pi_Q \cdot TF_t \cdot \frac{\partial v}{\partial t} = 0.$$

Thus, since $T\pi_Q \cdot X_E(u) = u$ by the second-order equation property, we get

$$T\pi_Q \cdot TF_t \cdot \frac{\partial v}{\partial t} = -\dot{q},$$

where $(q, \dot{q}) = F_t(v) \in T_qQ$. Thus,

$$(F_t^* \Theta_L)\left(\frac{\partial v}{\partial t}\right) = \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i.$$

Also, since the base point of v does not change with t, $T\pi_Q \cdot (\partial v/\partial t) = 0$, so $\Theta_L(\partial v/\partial t) = 0$. Thus, (8.2.10) becomes

$$\frac{\partial S}{\partial t} = L(q, \dot{q}, t) - \frac{\partial L}{\partial \dot{q}}\dot{q} = -H(q, p, t),$$

where $p = \partial L/\partial \dot{q}$ as usual.

It remains only to show that $\partial S/\partial q = p$. To do this, we differentiate (8.2.8) implicitly with respect to q to give

$$T\pi_Q \cdot TF_t(v) \cdot (T_q v \cdot u) = u. \tag{8.2.11}$$

Then, from (8.2.9) and (8.2.7),

$$\begin{aligned} T_q S(q, \overline{q}, t) \cdot u &= \mathbf{d}\mathfrak{S}_t(v) \cdot (T_q v \cdot u) \\ &= (F_t^* \Theta_L)\,(T_q v \cdot u) - \Theta_L(T_q v \cdot u). \end{aligned}$$

As in (8.2.10), the last term vanishes, since the base point $\overline{q}$ of v is fixed. Then, letting $p = \mathbb{F}L(F_t(v))$, we get, from the definition of Θ_L and pull-back,

$$(F_t^* \Theta_L)\,(T_q v \cdot u) = \langle p, T\pi_Q \cdot TF_t(v) \cdot (T_q v \cdot u)\rangle = \langle p, u\rangle$$

in view of (8.2.11). ■

The fact that $\partial S/\partial q = p$ also follows from the definition of S and the fundamental formula (8.2.4). Just as we derived $p = \partial S/\partial q$, we can derive $\partial S/\partial \overline{q} = -\overline{p}$; in other words, *$S$ is the generating function for the canonical transformation* $(q, p) \mapsto (\overline{q}, \overline{p})$.

Some History of the Euler–Lagrange Equations. In the following paragraphs we make a few historical remarks concerning the Euler–Lagrange equations.[4] Naturally, much of the story focuses on Lagrange. Section V of Lagrange's *Mécanique Analytique* [1788] contains the equations of motion in Euler–Lagrange form (8.1.3). Lagrange writes $Z = T - V$ for what we would call the Lagrangian today. In the previous section Lagrange came to these equations by asking for a coordinate-invariant expression for mass times acceleration. His conclusion is that it is given (in abbreviated notation) by $(d/dt)(\partial T/\partial v) - \partial T/\partial q$, which transforms under arbitrary substitutions of position variables as a one-form. Lagrange does *not* recognize the equations of motion as being equivalent to the variational principle

$$\delta \int L\, dt = 0.$$

This was observed only a few decades later by Hamilton [1834]. The peculiar fact about this is that Lagrange *did* know the general form of the differential equations for variational problems, and he actually had commented on

[4]Many of these interesting historical points were conveyed to us by Hans Duistermaat, to whom we are very grateful. The reader can also profitably consult some of the standard texts such as those of Whittaker [1927], Wintner [1941], and Lanczos [1949] for additional interesting historical information.

Euler's proof of this—his early work on this in 1759 was admired very much by Euler. He immediately applied it to give a proof of the Maupertuis principle of least action, as a consequence of Newton's equations of motion. This principle, apparently having its roots in the early work of Leibniz, is a less natural principle in the sense that the curves are varied only over those that have a constant energy. It is also Hamilton's principle that applies in the *time-dependent* case, when H is *not* conserved and that also generalizes to allow for certain external forces as well.

This discussion in the *Mécanique Analytique* precedes the equations of motion in general coordinates, and so is written in the case that the kinetic energy is of the form $\sum_i m_i v_i^2$, where the m_i are positive constants. Wintner [1941] is also amazed by the fact that the more complicated Maupertuis principle precedes Hamilton's principle. One possible explanation is that Lagrange did not consider L as an interesting physical quantity—for him it was only a convenient function for writing down the equations of motion in a coordinate-invariant fashion. The time span between his work on variational calculus and the *Mécanique Analytique* (1788, 1808) could also be part of the explanation—he may not have been thinking of the variational calculus when he addressed the question of a coordinate-invariant formulation of the equations of motion.

Section V starts by discussing the evident fact that the position and velocity at time t depend on the initial position and velocity, which can be chosen freely. We might write this as (suppressing the coordinate indices for simplicity) $q = q(t, q_0, v_0)$, $v = v(t, q_0, v_0)$, and in modern terminology we would talk about the flow in $x = (q, v)$-space. One problem in reading Lagrange is that he does not explicitly write the variables on which his quantities depend. In any case, he then makes an infinitesimal variation in the initial condition and looks at the corresponding variations of position and velocity at time t. In our notation, $\delta x = (\partial x/\partial x_0)(t, x_0)\delta x_0$. We would say that he considers the tangent mapping of the flow on the tangent bundle of $X = TQ$. Now comes the first interesting result. He makes two such variations, one denoted by δx and the other by Δx, and he writes down a bilinear form $\omega(\delta x, \Delta x)$, in which we recognize ω as the pull-back of the canonical symplectic form on the cotangent bundle of Q, by means of the fiber derivative $\mathbb{F}L$. What he then shows is that this symplectic product is constant as a function of t. This is nothing other than the *invariance of the symplectic form ω under the flow in TQ.*

It is striking that Lagrange obtains the invariance of the symplectic form in TQ and not in T^*Q just as we do in the text where this is derived from Hamilton's principle. In fact, Lagrange does *not* look at the equations of motion in the cotangent bundle via the transformation $\mathbb{F}L$; again it is Hamilton who observes that these take the canonical Hamiltonian form. This is retrospectively puzzling, since later on in Section V, Lagrange states very explicitly that it is useful to pass to the (q, p)-coordinates by means of the coordinate transformation $\mathbb{F}L$, and one even sees written down a system of ordinary differential equations *in Hamiltonian form*, but with the total energy function H replaced by some other mysterious function $-\Omega$. Lagrange does use the letter H for the constant value of energy, apparently in honor of Huygens. He also knew about the conservation of momentum as a result of translational symmetry.

The part where he does this deals with the case in which he perturbs the system by perturbing the potential from $V(q)$ to $V(q) - \Omega(q)$, leaving the kinetic energy unchanged. To this perturbation problem he applies his famous method of variation of constants, which is presented here in a truly nonlinear framework! In our notation, he keeps $t \mapsto x(t, x_0)$ as the solution of the unperturbed system, and then looks at the differential equations for $x_0(t)$ that make $t \mapsto x(t, x_0(t))$ a solution of the perturbed system. The result is that if V is the vector field of the unperturbed system and $V + W$ is the vector field of the perturbed system, then

$$\frac{dx_0}{dt} = ((e^{tV})^* W)(x_0).$$

In words, $x_0(t)$ is the solution of the time-dependent system, the vector field of which is obtained by pulling back W by means of the flow of V after time t. In the case that Lagrange considers, the dq/dt-component of the perturbation is equal to zero, and the dp/dt-component is equal to $\partial\Omega/\partial q$. Thus, it is obviously in a Hamiltonian form; here one does not use anything about Legendre transformations (which Lagrange does not seem to know). But Lagrange knows already that the flow of the unperturbed system preserves the symplectic form, and he shows that the pull-back of his W under such a transformation is a vector field in Hamiltonian form. Actually, this is a time-dependent vector field, defined by the function

$$G(t, q_0, p_0) = -\Omega(q(t, q_0, p_0)).$$

A potential point of confusion is that Lagrange denotes this by $-\Omega$ and writes down expressions like $d\Omega/dp$, and one might first think that these are zero because Ω was assumed to depend only on q. Lagrange presumably means that

$$\frac{dq_0}{dt} = \frac{\partial G}{\partial p_0}, \qquad \frac{dp_0}{dt} = -\frac{\partial G}{\partial q_0}.$$

Most classical textbooks on mechanics, for example Routh [1877, 1884], correctly point out that Lagrange has the invariance of the symplectic form in (q, v) coordinates (rather than in the canonical (q, p) coordinates). Less attention is usually paid to the variation of constants equation in Hamiltonian form, but it must have been generally known that Lagrange derived these—see, for example, Weinstein [1981]. In fact, we should point out that the whole question of linearizing the Euler–Lagrange and Hamilton equations and retaining the mechanical structure is remarkably subtle (see Marsden, Ratiu, and Raugel [1991], for example).

Lagrange continues by introducing the *Poisson brackets* for arbitrary functions, arguing that these are useful in writing the time-derivative of arbitrary functions of arbitrary variables, along solutions of systems in Hamiltonian form. He also continues by saying that if Ω is small, then $x_0(t)$ in zero-order approximation is a constant, and he obtains the next-order approximation by an integration over t; here Lagrange introduces the first steps of the so-called *method of averaging*. When Lagrange discovered (in 1808) the invariance of the symplectic form, the variations-of-constants equations in Hamiltonian form, and the Poisson brackets, he was already 73 years old. It is quite probable that Lagrange generously gave some of these bracket ideas to Poisson at this time. In any case, it is clear that Lagrange had a surprisingly large part of the symplectic picture of classical mechanics.

Exercises

◇ **8.2-1.** Derive the Hamilton–Jacobi equation starting with the phase space version of Hamilton's principle.

8.3 Constrained Systems

We begin this section with the Lagrange multiplier theorem for purposes of studying constrained dynamics.

The Lagrange Multiplier Theorem. We state the theorem with a sketch of the proof, referring to Abraham, Marsden, and Ratiu [1988] for details. We shall not be absolutely precise about the technicalities (such as how to interpret dual spaces).

First, consider the case of functions defined on linear spaces. Let V and Λ be Banach spaces and let $\varphi : V \to \Lambda$ be a smooth map. Suppose 0 is a regular value of φ, so that $C := \varphi^{-1}(0)$ is a submanifold. Let $h : V \to \mathbb{R}$ be a smooth function and define $\overline{h} : V \times \Lambda^* \to \mathbb{R}$ by

$$\overline{h}(x, \lambda) = h(x) - \langle \lambda, \varphi(x) \rangle \,. \tag{8.3.1}$$

Theorem 8.3.1 (Lagrange Multiplier Theorem for Linear Spaces)**.** *The following are equivalent conditions on $x_0 \in C$:*

(i) *x_0 is a critical point of $h|C$; and*

(ii) *there is a $\lambda_0 \in \Lambda^*$ such that (x_0, λ_0) is a critical point of $\overline{h}$.*

Sketch of Proof. Since

$$\mathbf{D}\overline{h}(x_0, \lambda_0) \cdot (x, \lambda) = \mathbf{D}h(x_0) \cdot x - \langle \lambda_0, \mathbf{D}\varphi(x_0) \cdot x \rangle - \langle \lambda, \varphi(x_0) \rangle$$

and $\varphi(x_0) = 0$, the condition $\mathbf{D}\overline{h}(x_0, \lambda_0) \cdot (x, \lambda) = 0$ is equivalent to

$$\mathbf{D}h(x_0) \cdot x = \langle \lambda_0, \mathbf{D}\varphi(x_0) \cdot x \rangle \tag{8.3.2}$$

for all $x \in V$ and $\lambda \in \Lambda^*$. The tangent space to C at x_0 is $\ker \mathbf{D}\varphi(x_0)$, so (8.3.2) implies that $h|C$ has a critical point at x_0.

Conversely, if $h|C$ has a critical point at x_0, then $\mathbf{D}h(x_0) \cdot x = 0$ for all x satisfying $\mathbf{D}\varphi(x_0) \cdot x = 0$. By the implicit function theorem, there is a smooth coordinate change that straightens out C; that is, it allows us to assume that $V = W \oplus \Lambda$, $x_0 = 0$, C is (in a neighborhood of 0) equal to W, and φ (in a neighborhood of the origin) is the projection to Λ. With these simplifications, condition (i) means that the first partial derivative of h vanishes. We choose λ_0 to be $\mathbf{D}_2 h(x_0)$ regarded as an element of Λ^*; then (8.3.2) clearly holds. ∎

The Lagrange multiplier theorem is a convenient test for constrained critical points, as we know from calculus. It also leads to a convenient test for constrained maxima and minima. For instance, to test for a minimum, let $\alpha > 0$ be a constant, let (x_0, λ_0) be a critical point of $\overline{h}$, and consider

$$h_\alpha(x, \lambda) = h(x) - \langle \lambda, \varphi(x) \rangle + \alpha \|\lambda - \lambda_0\|^2, \tag{8.3.3}$$

which also has a critical point at (x_0, λ_0). Clearly, if h_α has a minimum at (x_0, λ_0), then $h|C$ has a minimum at x_0. This observation is convenient, since one can use the unconstrained second derivative test on h_α, which leads to the theory of ***bordered Hessians***. (For an elementary discussion, see Marsden and Tromba [1996, p. 220ff].)

A second remark concerns the generalization of the Lagrange multiplier theorem to the case where V is a manifold but h is still real-valued. Such a context is as follows. Let M be a manifold and let $N \subset M$ be a submanifold. Suppose $\pi : E \to M$ is a vector bundle over M and φ is a section of E that is transverse to fibers. Assume $N = \varphi^{-1}(0)$.

Theorem 8.3.2 (Lagrange Multiplier Theorem for Manifolds)**.** *The following are equivalent for $x_0 \in N$ and $h : M \to \mathbb{R}$ smooth:*

(i) *x_0 is a critical point of $h|N$; and*

(ii) *there is a section λ_0 of the dual bundle E^* such that $\lambda_0(x_0)$ is a critical point of $\overline{h} : E^* \to \mathbb{R}$ defined by*

$$\overline{h}(\lambda_x) = h(x) - \langle \lambda_x, \varphi(x) \rangle . \tag{8.3.4}$$

In (8.3.4), λ_x denotes an arbitrary element of E_x^*. We leave it to the reader to adapt the proof of the previous theorem to this situation.

Holonomic Constraints. Many mechanical systems are obtained from higher-dimensional ones by adding constraints. Rigidity in rigid-body mechanics and incompressibility in fluid mechanics are two such examples, while constraining a free particle to move on a sphere is another.

Typically, constraints are of two types. Holonomic constraints are those imposed on the configuration space of a system, such as those mentioned in the preceding paragraph. Others, such as *rolling constraints*, involve the conditions on the velocities and are termed *nonholonomic*.

A ***holonomic constraint*** can be defined for our purposes as the specification of a submanifold $N \subset Q$ of a given configuration manifold Q. (More generally, a holonomic constraint is an integrable subbundle of TQ.) Since we have the natural inclusion $TN \subset TQ$, a given Lagrangian $L : TQ \to \mathbb{R}$ can be restricted to TN to give a Lagrangian L_N. We now have two Lagrangian systems, namely those associated to L and to L_N, assuming that both are regular. We now relate the associated variational principles and the Hamiltonian vector fields.

Suppose that $N = \varphi^{-1}(0)$ for a section $\varphi : Q \to E^*$, the dual of a vector bundle E over Q. The variational principle for L_N can be phrased as

$$\delta \int L_N(q, \dot{q})\, dt = 0, \tag{8.3.5}$$

where the variation is over curves with fixed endpoints and subject to the constraint $\varphi(q(t)) = 0$. By the Lagrange multiplier theorem, (8.3.5) is equivalent to

$$\delta \int [L(q(t), \dot{q}(t)) - \langle \lambda(q(t), t), \varphi(q(t)) \rangle]\, dt = 0 \tag{8.3.6}$$

for some function $\lambda(q, t)$ taking values in the bundle E and where the variation is over curves q in Q and curves λ in E.[5] In coordinates, (8.3.6) reads

$$\delta \int [L(q^i, \dot{q}^i) - \lambda^a(q^i, t)\varphi_a(q^i)]\, dt = 0. \tag{8.3.7}$$

The corresponding Euler–Lagrange equations in the variables q^i, λ^a are

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} = \frac{\partial L}{\partial q^i} - \lambda^a \frac{\partial \varphi_a}{\partial q^i} \tag{8.3.8}$$

and

$$\varphi_a = 0. \tag{8.3.9}$$

[5] This conclusion assumes some regularity in t on the Lagrange multiplier λ. One can check (after the fact) that this assumption is justified by relating λ to the forces of constraint, as in the next theorem.

They are viewed as equations in the unknowns $q^i(t)$ and $\lambda^a(q^i, t)$; if E is a trivial bundle, we can take λ to be a function only of t.[6]

We summarize these findings as follows.

Theorem 8.3.3. *The Euler–Lagrange equations for L_N on the manifold $N \subset Q$ are equivalent to the equations* (8.3.8) *together with the constraints* $\varphi = 0$.

We interpret the term $-\lambda^a \partial\varphi_a/\partial q^i$ as the *force of constraint*, since it is the force that is added to the Euler–Lagrange operator (see §7.8) in the *unconstrained space* in order to maintain the constraints. In the next section we will develop the geometric interpretation of these forces of constraint.

Notice that $\mathcal{L} = L - \lambda^a\varphi_a$ as a Lagrangian in q and λ is degenerate in λ; that is, the time-derivative of λ does not appear, so its conjugate momentum π_a is constrained to be zero. Regarding $\mathcal{L}$ as defined on TE, the corresponding Hamiltonian on T^*E is formally

$$\mathcal{H}(q, p, \lambda, \pi) = H(q, p) + \lambda^a\varphi_a, \tag{8.3.10}$$

where H is the Hamiltonian corresponding to L.

One has to be a little careful in interpreting Hamilton's equations, because $\mathcal{L}$ is degenerate; the general theory appropriate for this situation is the *Dirac theory of constraints*, which we discuss in §8.5. However, in the present context this theory is quite simple and proceeds as follows. One calls $C \subset T^*E$ defined by $\pi_a = 0$ the ***primary constraint set***; it is the image of the Legendre transform, provided that the original L was regular. The canonical form Ω is pulled back to C to give a presymplectic form (a closed but possibly degenerate two-form) Ω_C, and one seeks $X_\mathcal{H}$ such that

$$\mathbf{i}_{X_\mathcal{H}}\Omega_C = \mathbf{d}\mathcal{H}. \tag{8.3.11}$$

In this case, the degeneracy of Ω_C gives no equation for λ; that is, the evolution of λ is indeterminate. The other Hamiltonian equations are equivalent to (8.3.8) and (8.3.9), so in this sense the Lagrangian and Hamiltonian pictures are still equivalent.

Exercises

◇ **8.3-1.** Write out the second derivative of h_α at (x_0, λ_0) and relate your answer to the bordered Hessian.

◇ **8.3-2.** Derive the equations for a simple pendulum using the Lagrange multiplier method and compare them with those obtained using generalized coordinates.

[6] The combination $\mathcal{L} = L - \lambda^a\varphi_a$ is related to the Routhian construction for a Lagrangian with cyclic variables; see §8.9.

◇ **8.3-3** (Neumann [1859]).

(a) Derive the equations of motion of a particle of unit mass on the sphere S^{n-1} under the influence of a quadratic potential $A\mathbf{q} \cdot \mathbf{q}$, $\mathbf{q} \in \mathbb{R}^n$, where A is a fixed real diagonal matrix.

(b) Form the matrices $X = (q^i q^j)$, $P = (\dot{q}^i q^j - q^j \dot{q}^j)$. Show that the system in (a) is equivalent to $\dot{X} = [P, X]$, $\dot{P} = [X, A]$. (This was observed first by K. Uhlenbeck.) Equivalently, show that

$$(-X + P\lambda + A\lambda^2)^{\cdot} = [-X + P\lambda + A\lambda^2, -P - A\lambda].$$

(c) Verify that

$$E(X, P) = -\frac{1}{4}\operatorname{trace}(P^2) + \frac{1}{2}\operatorname{trace}(AX)$$

is the total energy of this system.

(d) Verify that for $k = 1, \ldots, n-1$,

$$f_k(X, P) = \frac{1}{2(k+1)}\operatorname{trace}\left(-\sum_{i=0}^{k} A^i X A^{k-i} + \sum_{\substack{i+j+l=k-1 \\ i,j,l \geq 0}} A^i P A^j P A^l \right),$$

are conserved on the flow of the C. Neumann problem (Ratiu [1981b]).

8.4 Constrained Motion in a Potential Field

We saw in the preceding section how to write the equations for a constrained system in terms of variables on the containing space. We continue this line of investigation here by specializing to the case of motion in a potential field. In fact, we shall determine by geometric methods the extra terms that need to be added to the Euler–Lagrange equations, that is, the forces of constraint, to ensure that the constraints are maintained.

Let Q be a (weak) Riemannian manifold and let $N \subset Q$ be a submanifold. Let

$$\mathbb{P} : (TQ)|N \to TN \tag{8.4.1}$$

be the orthogonal projection of TQ to TN defined pointwise on N.

Consider a Lagrangian $L : TQ \to \mathbb{R}$ of the form $L = K - V \circ \tau_Q$, that is, kinetic minus potential energy. The Riemannian metric associated to the kinetic energy is denoted by $\langle\!\langle\, , \rangle\!\rangle$. The restriction $L_N = L|TN$ is also of the form kinetic minus potential, using the metric induced on N and the potential $V_N = V|N$. We know from §7.7 that if E_N is the energy of L_N, then

$$X_{E_N} = S_N - \mathrm{ver}(\nabla V_N), \tag{8.4.2}$$

where S_N is the spray of the metric on N and ver( ) denotes vertical lift. Recall that integral curves of (8.4.2) are solutions of the Euler–Lagrange equations. Let S be the geodesic spray on Q.

First notice that ∇V_N and ∇V are related in a very simple way: For $q \in N$,

$$\nabla V_N(q) = \mathbb{P} \cdot [\nabla V(q)].$$

Thus, the main complication is in the geodesic spray.

Proposition 8.4.1. *$S_N = T\mathbb{P} \circ S$ at points of TN.*

Proof. For the purpose of this proof we can ignore the potential and let $L = K$. Let $R = TQ|N$, so that $\mathbb{P} : R \to TN$ and therefore

$$T\mathbb{P} : TR \to T(TN), \quad S : R \to T(TQ), \quad \text{and} \quad T\tau_Q \circ S = \text{identity},$$

since S is second-order. But

$$TR = \{\, w \in T(TQ) \mid T\tau_Q(w) \in TN \,\},$$

so $S(TN) \subset TR$, and hence $T\mathbb{P} \circ S$ makes sense at points of TN.

If $v \in TQ$ and $w \in T_v(TQ)$, then $\Theta_L(v) \cdot w = \langle\!\langle v, T_v\tau_Q(w)\rangle\!\rangle$. Letting $i : R \to TQ$ be the inclusion, we claim that

$$\mathbb{P}^*\Theta_{L|TN} = i^*\Theta_L. \tag{8.4.3}$$

Indeed, for $v \in R$ and $w \in T_vR$, the definition of pull-back gives

$$\mathbb{P}^*\Theta_{L|TN}(v) \cdot w = \langle\!\langle \mathbb{P}v, (T\tau_Q \circ T\mathbb{P})(w)\rangle\!\rangle = \langle\!\langle \mathbb{P}v, T(\tau_Q \circ \mathbb{P})(w)\rangle\!\rangle. \tag{8.4.4}$$

Since on R, $\tau_Q \circ \mathbb{P} = \tau_Q$, $\mathbb{P}^* = \mathbb{P}$, and $w \in T_vR$, (8.4.4) becomes

$$\begin{aligned}\mathbb{P}^*\Theta_{L|TN}(v) \cdot w &= \langle\!\langle \mathbb{P}v, T\tau_Q(w)\rangle\!\rangle = \langle\!\langle v, \mathbb{P}T\tau_Q(w)\rangle\!\rangle = \langle\!\langle v, T\tau_Q(w)\rangle\!\rangle \\ &= \Theta_L(v) \cdot w = (i^*\Theta_L)(v) \cdot w.\end{aligned}$$

Taking the exterior derivative of (8.4.3) gives

$$\mathbb{P}^*\Omega_{L|TN} = i^*\Omega_L. \tag{8.4.5}$$

In particular, for $v \in TN$, $w \in T_vR$, and $z \in T_v(TN)$, the definition of pull-back and (8.4.5) give

$$\begin{aligned}\Omega_L(v)(w, z) &= (i^*\Omega_L)(v)(w, z) = (\mathbb{P}^*\Omega_{L|TN})(v)(w, z) \\ &= \Omega_{L|TN}(\mathbb{P}v)(T\mathbb{P}(w), T\mathbb{P}(z)) \\ &= \Omega_{L|TN}(v)(T\mathbb{P}(w), z).\end{aligned} \tag{8.4.6}$$

But

$$\mathbf{d}E(v) \cdot z = \Omega_L(v)(S(v), z) = \Omega_{L|TN}(v)(S_N(v), z),$$

since S and S_N are Hamiltonian vector fields for E and $E|TN$, respectively. From (8.4.6),

$$\Omega_{L|TN}(v)(T\mathbb{P}(S(v)), z) = \Omega_L(v)(S(v), z) = \Omega_{L|TN}(v)(S_N(v), z),$$

so by weak nondegeneracy of $\Omega_{L|TN}$ we get the desired relation

$$S_N = T\mathbb{P} \circ S. \qquad \blacksquare$$

Corollary 8.4.2. *For $v \in T_qN$:*

(i) $(S - S_N)(v)$ *is the vertical lift of a vector $Z(v) \in T_qQ$ relative to v;*

(ii) $Z(v) \perp T_qN$*; and*

(iii) $Z(v) = -\nabla_v v + \mathbb{P}(\nabla_v v)$ *is minus the normal component of $\nabla_v v$, where in $\nabla_v v$, v is extended to a vector field on Q tangent to N.*

Proof. (i) Since $T\tau_Q(S(v)) = v = T\tau_Q(S_N(v))$, we have

$$T\tau_Q(S - S_N)(v) = 0,$$

that is, $(S - S_N)(v)$ is vertical. The statement now follows from the comments following Definition 7.7.1.

(ii) For $u \in T_qQ$, we have $T\mathbb{P} \cdot \operatorname{ver}(u, v) = \operatorname{ver}(\mathbb{P}u, v)$, since

$$\begin{aligned} \operatorname{ver}(\mathbb{P}u, v) &= \frac{d}{dt}(v + t\mathbb{P}u)\bigg|_{t=0} = \frac{d}{dt}\mathbb{P}(v + tu)\bigg|_{t=0} \\ &= T\mathbb{P} \cdot \operatorname{ver}(u, v). \end{aligned} \qquad (8.4.7)$$

By Part (i), $S(v) - S_N(v) = \operatorname{ver}(Z(v), v)$ for some $Z(v) \in T_qQ$, so that using the previous theorem, (8.4.7), and $\mathbb{P} \circ \mathbb{P} = \mathbb{P}$, we get

$$\begin{aligned} \operatorname{ver}(\mathbb{P}Z(v), v) &= T\mathbb{P} \cdot \operatorname{ver}(Z(v), v) \\ &= T\mathbb{P}(S(v) - S_N(v)) \\ &= T\mathbb{P}(S(v) - T\mathbb{P} \circ S(v)) = 0. \end{aligned}$$

Therefore, $\mathbb{P}Z(v) = 0$, that is, $Z(v) \perp T_qN$.

(iii) Let $v(t)$ be a curve of tangents to N; $v(t) = \dot{c}(t)$, where $c(t) \in N$. Then in a chart,

$$S(c(t), v(t)) = \big(c(t), v(t), v(t), \gamma_{c(t)}(v(t), v(t))\big)$$

by (7.5.5). Extending $v(t)$ to a vector field v on Q tangent to N we get, in a standard chart,

$$\nabla_v v = -\gamma_c(v, v) + \mathbf{D}v(c) \cdot v = -\gamma_c(v, v) + \frac{dv}{dt}$$

by (7.5.19), so on TN,

$$S(v) = \frac{dv}{dt} - \operatorname{ver}(\nabla_v v, v).$$

Since $dv/dt \in TN$, (8.4.7) and the previous proposition give

$$S_N(v) = T\mathbb{P}\frac{dv}{dt} - \operatorname{ver}(\mathbb{P}(\nabla_v v), v) = \frac{dv}{dt} - \operatorname{ver}(\mathbb{P}(\nabla_v v), v).$$

Thus, by part (i),

$$\operatorname{ver}(Z(v), v) = S(v) - S_N(v) = \operatorname{ver}(-\nabla_v v + \mathbb{P}\nabla_v v, v). \qquad \blacksquare$$

The map $Z : TN \to TQ$ is called the ***force of constraint***. We shall prove below that if the codimension of N in Q is one, then

$$Z(v) = -\nabla_v v + \mathbb{P}(\nabla_v v) = -\langle \nabla_v v, n \rangle n,$$

where n is the unit normal vector field to N in Q, equals the negative of the quadratic form associated to the second fundamental form of N in Q, a result due to Gauss. (We shall define the second fundamental form, which measures how "curved" N is within Q, shortly.) It is not obvious at first that the expression $\mathbb{P}(\nabla_v v) - \nabla_v v$ depends only on the pointwise values of v, but this follows from its identification with $Z(v)$.

To prove the above statement, we recall that the Levi-Civita covariant derivative has the property that for vector fields $u, v, w \in \mathfrak{X}(Q)$ the following identity is satisfied:

$$w[\langle u, v \rangle] = \langle \nabla_w u, v \rangle + \langle u, \nabla_w v \rangle, \tag{8.4.8}$$

as may be easily checked. Assume now that u and v are vector fields tangent to N and n is the unit normal vector field to N in Q. The identity (8.4.8) yields

$$\langle \nabla_v u, n \rangle + \langle u, \nabla_v n \rangle = 0. \tag{8.4.9}$$

The ***second fundamental form*** in Riemannian geometry is defined to be the map

$$(u, v) \mapsto -\langle \nabla_u n, v \rangle \tag{8.4.10}$$

with u, v, n as above. It is a classical result that this bilinear form is symmetric and hence is uniquely determined by polarization from its quadratic form $-\langle \nabla_v n, v \rangle$. In view of equation (8.4.9), this quadratic from has the alternative expression $\langle \nabla_v v, n \rangle$, which, after multiplication by n, equals $-Z(v)$, thereby proving the claim above.

As indicated, this discussion of the second fundamental form is under the assumption that the codimension of N in Q is one—keep in mind that our discussion of forces of constraint requires no such restriction.

As before, interpret $Z(v)$ as the constraining force needed to keep particles in N. Notice that N is totally geodesic (that is, geodesics in N are geodesics in Q) iff $Z = 0$.

Some interesting studies in the problem of showing convergence of solutions in the limit of strong constraining forces are Rubin and Ungar [1957], Ebin [1982], and van Kampen and Lodder [1984].

Exercises

◇ **8.4-1.** Compute the force of constraint Z and the second fundamental form for the sphere of radius R in $\mathbb{R}^3$.

◇ **8.4-2.** Assume that L is a regular Lagrangian on TQ and $N \subset Q$. Let $i : TN \to TQ$ be the embedding obtained from $N \subset Q$ and let Ω_L be the Lagrange two-form on TQ. Show that $i^*\Omega_L$ is the Lagrange two-form $\Omega_{L|TN}$ on TN. Assuming that L is hyperregular, show that the Legendre transform defines a symplectic embedding $T^*N \subset T^*Q$.

◇ **8.4-3.** In $\mathbb{R}^3$, let

$$H(\mathbf{q}, \mathbf{p}) = \frac{1}{2m} \left[\|\mathbf{p}\|^2 - (\mathbf{p} \cdot \mathbf{q})^2 \right] + mgq^3,$$

where $\mathbf{q} = (q^1, q^2, q^3)$. Show that Hamilton's equations in $\mathbb{R}^3$ *automatically* preserve T^*S^2 and give the equations for the spherical pendulum when restricted to this invariant (symplectic) submanifold. (Hint: Use the formulation of Lagrange's equations with constraints in §8.3.)

◇ **8.4-4.** Redo the C. Neumann problem in Exercise 8.3-3 using Corollary 8.4.2 and the interpretation of the constraining force in terms of the second fundamental form.

8.5 Dirac Constraints

If (P, Ω) is a symplectic manifold, a submanifold $S \subset P$ is called a ***symplectic submanifold*** when $\omega := i^*\Omega$ is a symplectic form on S, $i : S \to P$ being the inclusion. Thus, S inherits a Poisson bracket structure; its relationship to the bracket structure on P is given by a formula of Dirac [1950] that will be derived in this section. Dirac's work was motivated by the study of constrained systems, especially relativistic ones, where one thinks of S as a constraint subspace of phase space (see Gotay, Isenberg, and Marsden [1997] and references therein for more information). Let us work in the finite-dimensional case; the reader is invited to study the intrinsic infinite-dimensional version using Remark 1 below.

Dirac's Formula. Let $\dim P = 2n$ and $\dim S = 2k$. In a neighborhood of a point z_0 of S, choose coordinates $z^1, \ldots, z^{2n}$ on P such that S is given by

$$z^{2k+1} = 0, \ldots, z^{2n} = 0,$$

and so $z^1, \ldots, z^{2k}$ provide local coordinates for S.

Consider the matrix whose entries are

$$C^{ij}(z) = \{z^i, z^j\}, \quad i, j = 2k+1, \ldots, 2n.$$

Assume that the coordinates are chosen such that C^{ij} is an invertible matrix at z_0 and hence in a neighborhood of z_0. (Such coordinates always exist, as is easy to see.) Let the inverse of C^{ij} be denoted by $[C_{ij}(z)]$. Let F be a smooth function on P and $F|S$ its restriction to S. We are interested in relating $X_{F|S}$ and X_F as well as the brackets $\{F, G\}|S$ and $\{F|S, G|S\}$.

Proposition 8.5.1 (Dirac's Bracket Formula)**.** *In a coordinate neighborhood as described above, and for $z \in S$, we have*

$$X_{F|S}(z) = X_F(z) - \sum_{i,j=2k+1}^{2n} \{F, z^i\} C_{ij}(z) X_{z^j}(z) \tag{8.5.1}$$

and

$$\{F|S, G|S\}(z) = \{F, G\}(z) - \sum_{i,j=2k+1}^{2n} \{F, z^i\} C_{ij}(z) \{z^j, G\}. \tag{8.5.2}$$

Proof. To verify (8.5.1), we show that the right-hand side satisfies the condition required for $X_{F|S}(z)$, namely that it be a vector field on S and that

$$\omega_z(X_{F|S}(z), v) = \mathbf{d}(F|S)_z \cdot v \tag{8.5.3}$$

for $v \in T_zS$. Since S is symplectic,

$$T_zS \cap (T_zS)^\Omega = \{0\},$$

where $(T_zS)^\Omega$ denotes the Ω-orthogonal complement. Since

$$\dim(T_zS) + \dim(T_zS)^\Omega = 2n,$$

we get

$$T_zP = T_zS \oplus (T_zS)^\Omega. \tag{8.5.4}$$

If $\pi_z : T_zP \to T_zS$ is the associated projection operator, one can verify that

$$X_{F|S}(z) = \pi_z \cdot X_F(z), \tag{8.5.5}$$

so in fact, (8.5.1) is a formula for π_z in coordinates; equivalently,

$$(\mathrm{Id} - \pi_z)X_F(z) = \sum_{i,j=2k+1}^{2n} \{F, z^i\}C_{ij}(z)X_{z^j}(z) \tag{8.5.6}$$

gives the projection to $(T_zS)^\Omega$. To verify (8.5.6), we need to check that the right-hand side

(i) is an element of $(T_zS)^\Omega$;

(ii) equals $X_F(z)$ if $X_F(z) \in (T_zS)^\Omega$; and

(iii) equals 0 if $X_F(z) \in T_zS$.

To prove (i), observe that $X_K(z) \in (T_zS)^\Omega$ means

$$\Omega(X_K(z), v) = 0 \qquad \text{for all } v \in T_zS;$$

that is,

$$\mathbf{d}K(z) \cdot v = 0 \qquad \text{for all } v \in T_zS.$$

But for $K = z^j$, $j = 2k+1, \ldots, 2n$, $K \equiv 0$ on S, and hence $\mathbf{d}K(z) \cdot v = 0$. Thus, $X_{z^j}(z) \in (T_zS)^\Omega$, so (i) holds.

For (ii), if $X_F(z) \in (T_zS)^\Omega$, then

$$\mathbf{d}F(z) \cdot v = 0 \qquad \text{for all } v \in T_zS$$

and, in particular, for $v = \partial/\partial z^i$, $i = 1, \ldots, 2k$. Therefore, for $z \in S$, we can write

$$\mathbf{d}F(z) = \sum_{j=2k+1}^{2n} a_j \, dz^j \tag{8.5.7}$$

and hence

$$X_F(z) = \sum_{j=2k+1}^{2n} a_j X_{z^j}(z). \tag{8.5.8}$$

The a_j are determined by pairing (8.5.8) with dz^i, $i = 2k+1, \ldots, 2n$, to give

$$-\left\langle dz^i, X_F(z)\right\rangle = \{F, z^i\} = \sum_{j=2k+1}^{2n} a_j \{z^j, z^i\} = \sum_{j=2k+1}^{2n} a_j C^{ji},$$

or

$$a_j = \sum_{i=2k+1}^{2n} \{F, z^i\} C_{ij}, \tag{8.5.9}$$

which proves (ii). Finally, for (iii), $X_F(z) \in T_zS = ((T_zS)^\Omega)^\Omega$ means that $X_F(z)$ is Ω orthogonal to each X_{z^j}, $j = 2k+1, \ldots, 2n$. Thus, $\{F, z^j\} = 0$, so the right-hand side of (8.5.6) vanishes.

Formula (8.5.6) is therefore proved, and so, equivalently, (8.5.1) holds. Formula (8.5.2) follows by writing $\{F|S, G|S\} = \omega(X_{F|S}, X_{G|S})$ and substituting (8.5.1). In doing this, the last two terms cancel. ■

In (8.5.2) notice that $\{F|S, G|S\}(z)$ is intrinsic to $F|S$, $G|S$, and S. The bracket does not depend on how $F|S$ and $G|S$ are extended off S to functions F, G on P. This is not true for just $\{F, G\}(z)$, which *does* depend on the extensions, but the extra term in (8.5.2) cancels this dependence.

Remarks.

1. A coordinate-free way to write (8.5.2) is as follows. Write $S = \psi^{-1}(m_0)$, where $\psi : P \to M$ is a submersion on S. For $z \in S$ and $m = \psi(z)$, let

$$C_m : T_m^*M \times T_m^*M \to \mathbb{R} \tag{8.5.10}$$

be given by

$$C_m(\mathbf{d}F_m, \mathbf{d}G_m) = \{F \circ \psi, G \circ \psi\}(z) \tag{8.5.11}$$

for $F, G \in \mathcal{F}(M)$. Assume that C_m is invertible, with "inverse"

$$C_m^{-1} : T_mM \times T_mM \to \mathbb{R}.$$

Then

$$\{F|S, G|S\}(z) = \{F, G\}(z) - C_m^{-1}(T_z\psi \cdot X_F(z), T_z\psi \cdot X_G(z)). \tag{8.5.12}$$

2. There is another way to derive and write Dirac's formula, using complex structures. Suppose $\langle\!\langle\, , \rangle\!\rangle_z$ is an inner product on T_zP and

$$\mathbb{J}_z : T_zP \to T_zP$$

is an orthogonal transformation satisfying $\mathbb{J}_z^2 = -\,\mathrm{Identity}$ and, as in §5.3,

$$\Omega_z(u, v) = \langle\!\langle \mathbb{J}_z u, v \rangle\!\rangle \tag{8.5.13}$$

for all $u, v \in T_zP$. With the inclusion $i : S \to P$ as before, we get corresponding structures induced on S; let

$$\omega = i^*\Omega. \tag{8.5.14}$$

If ω is nondegenerate, then (8.5.14) and the induced metric define an associated complex structure $\mathbb{K}$ on S. At a point $z \in S$, suppose one has arranged to choose $\mathbb{J}_z$ to map T_zS to itself, and that $\mathbb{K}_z$ is the restriction of $\mathbb{J}_z$ to T_zS. At z, we then get

$$(T_zS)^\perp = (T_zS)^\Omega,$$

and thus symplectic projection coincides with orthogonal projection. From (8.5.5), and using coordinates as described earlier, but for which the $X_{z^j}(z)$ are also orthogonal, we get

$$\begin{aligned} X_{F|S}(z) &= X_F(z) - \sum_{j=2k+1}^{2n} \langle X_F(z), X_{z^j}(z) \rangle \, X_{z^j}(z) \\ &= X_F(z) + \sum_{j=2k+1}^{2n} \Omega(X_F(z), \mathbb{J}^{-1} X_{z^j}(z)) X_{z^j}. \end{aligned} \tag{8.5.15}$$

This is equivalent to (8.5.1) and so also gives (8.5.2); to see this, one shows that

$$\mathbb{J}^{-1} X_{z^j}(z) = - \sum_{i=2k+1}^{2n} X_{z^i}(z) C_{ij}(z). \tag{8.5.16}$$

Indeed, the symplectic pairing of each side with X_{z^p} gives δ_j^p.

3. For a relationship between Poisson reduction and Dirac's formula, see Marsden and Ratiu [1986].

Examples

(a) Holonomic Constraints. To treat ***holonomic constraints*** by the Dirac formula, proceed as follows. Let $N \subset Q$ be as in §8.4, so that

$TN \subset TQ$; with $i : N \to Q$ the inclusion, one obtains $(Ti)^*\Theta_L = \Theta_{L_N}$ by considering the following commutative diagram:

$$\begin{array}{ccc} TN & \xrightarrow{\;Ti\;} & TQ|N \\ {\scriptstyle \mathbb{F}L_N}\big\downarrow & & \big\downarrow{\scriptstyle \mathbb{F}L} \\ T^*N & \xleftarrow[\text{projection}]{} & T^*Q|N \end{array}$$

This realizes TN as a symplectic submanifold of TQ, and so Dirac's formula can be applied, reproducing (8.4.2). See Exercise 8.4-2. ♦

(b) KdV Equation. Suppose[7] one starts with a Lagrangian of the form

$$L(v_q) = \langle \alpha(q), v \rangle - h(q), \tag{8.5.17}$$

where α is a one-form on Q, and h is a function on Q. In coordinates, (8.5.17) reads

$$L(q^i, \dot{q}^i) = \alpha_i(q)\dot{q}^i - h(q^i). \tag{8.5.18}$$

The corresponding momenta are

$$p_i = \frac{\partial L}{\partial \dot{q}^i} = \alpha_i, \quad \text{i.e.,} \quad p = \alpha(q), \tag{8.5.19}$$

while the Euler–Lagrange equations are

$$\frac{d}{dt}(\alpha_i(q^j)) = \frac{\partial L}{\partial q^i} = \frac{\partial \alpha_j}{\partial q^i}\dot{q}^j - \frac{\partial h}{\partial q^i},$$

that is,

$$\frac{\partial \alpha_i}{\partial q^j}\dot{q}^j - \frac{\partial \alpha_j}{\partial q^i}\dot{q}^j = -\frac{\partial h}{\partial q^i}. \tag{8.5.20}$$

In other words, with $v^i = \dot{q}^i$,

$$\mathbf{i}_v \mathbf{d}\alpha = -\mathbf{d}h. \tag{8.5.21}$$

If $\mathbf{d}\alpha$ is nondegenerate on Q, then (8.5.21) defines Hamilton's equations for a vector field v on Q with Hamiltonian h and symplectic form $\Omega_\alpha = -\mathbf{d}\alpha$.

This collapse, or reduction, from TQ to Q is another instance of the Dirac theory and how it deals with degenerate Lagrangians in attempting

[7] We thank P. Morrison and M. Gotay for the following comment on how to view the KdV equation using constraints; see Gotay [1988].

to form the corresponding Hamiltonian system. Here the primary constraint manifold is the graph of α. Note that if we form the Hamiltonian on the primaries, then

$$H = p_i \dot{q}^i - L = \alpha_i \dot{q}^i - \alpha_i \dot{q}^i + h(q) = h(q), \tag{8.5.22}$$

that is, $H = h$, as expected from (8.5.21).

To put the KdV equation $u_t + 6uu_x + u_{xxx} = 0$ in this context, let $u = \psi_x$; that is, ψ is an indefinite integral for u. Observe that the KdV equation is the Euler–Lagrange equation for

$$L(\psi, \psi_t) = \int \left[\tfrac{1}{2}\psi_t \psi_x + \psi_x^3 - \tfrac{1}{2}(\psi_{xx})^2 \right] dx, \tag{8.5.23}$$

that is, $\delta \int L \, dt = 0$ gives $\psi_{xt} + 6\psi_x \psi_{xx} + \psi_{xxxx} = 0$, which is the KdV equation for u. Here α is given by

$$\langle \alpha(\psi), \varphi \rangle = \tfrac{1}{2} \int \psi_x \varphi \, dx, \tag{8.5.24}$$

and so by formula 6 in the table in §4.4,

$$-\mathbf{d}\alpha(\psi)(\psi_1, \psi_2) = \tfrac{1}{2} \int (\psi_1 \psi_{2x} - \psi_2 \psi_{1x}) \, dx, \tag{8.5.25}$$

which equals the KdV symplectic structure (3.2.9). Moreover, (8.5.22) gives the Hamiltonian

$$H = \int \left[\tfrac{1}{2}(\psi_{xx})^2 - \psi_x^3 \right] dx = \int \left[\tfrac{1}{2}(u_x)^2 - u^3 \right] dx, \tag{8.5.26}$$

also coinciding with Example (c) of §3.2. ♦

Exercises

◇ **8.5-1.** Derive formula (8.4.2) from (8.5.1).

◇ **8.5-2.** Work out Dirac's formula for

(a) $T^*S^1 \subset T^*\mathbb{R}^2$; and

(b) $T^*S^2 \subset T^*\mathbb{R}^3$.

In each case, note that the embedding makes use of the metric. Reconcile your analysis with what you found in Exercise 8.4-2.

8.6 Centrifugal and Coriolis Forces

In this section we discuss, in an elementary way, the basic ideas of centrifugal and Coriolis forces. This section takes the view of rotating *observers*, while the next sections take the view of rotating *systems*.

Rotating Frames. Let V be a three-dimensional oriented inner product space that we regard as "inertial space." Let ψ_t be a curve in $\mathrm{SO}(V)$, the group of orientation-preserving orthogonal linear transformations of V to V, and let X_t be the (possibly time-dependent) vector field generating ψ_t; that is,

$$X_t(\psi_t(\mathbf{v})) = \frac{d}{dt}\psi_t(\mathbf{v}), \tag{8.6.1}$$

or, equivalently,

$$X_t(\mathbf{v}) = (\dot{\psi}_t \circ \psi_t^{-1})(\mathbf{v}). \tag{8.6.2}$$

Differentiation of the orthogonality condition $\psi_t \cdot \psi_t^T = \mathrm{Id}$ shows that X_t is skew-symmetric.

A vector $\boldsymbol{\omega}$ in three-space defines a skew-symmetric 3×3 linear transformation $\hat{\boldsymbol{\omega}}$ using the cross product; specifically, it is defined by the equation

$$\hat{\boldsymbol{\omega}}(\mathbf{v}) = \boldsymbol{\omega} \times \mathbf{v}.$$

Conversely, any skew matrix can be so represented in a unique way. As we shall see later (see §9.2, especially equation (9.2.4)), this is a fundamental link between the Lie algebra of the rotation group and the cross product. This relation also will play a crucial role in the dynamics of a rigid body.

In particular, we can represent the skew matrix X_t this way:

$$X_t(\mathbf{v}) = \boldsymbol{\omega}(t) \times \mathbf{v}, \tag{8.6.3}$$

which defines $\boldsymbol{\omega}(t)$, the ***instantaneous rotation vector***.

Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be a fixed (inertial) orthonormal frame in V and let $\{\, \boldsymbol{\xi}_i = \psi_t(\mathbf{e}_i) \mid i = 1, 2, 3 \,\}$ be the corresponding ***rotating frame***. Given a point $\mathbf{v} \in V$, let $\mathbf{q} = (q^1, q^2, q^3)$ denote the vector in $\mathbb{R}^3$ defined by $\mathbf{v} = q^i \mathbf{e}_i$ and let $\mathbf{q}_R \in \mathbb{R}^3$ be the corresponding coordinate vector representing the components of the same vector $\mathbf{v}$ in the rotating frame, so $\mathbf{v} = q_R^i \boldsymbol{\xi}_i$. Let $A_t = A(t)$ be the *matrix* of ψ_t relative to the basis $\mathbf{e}_i$, that is, $\boldsymbol{\xi}_i = A_i^j \mathbf{e}_j$; then

$$\mathbf{q} = A_t \mathbf{q}_R, \quad \text{i.e.,} \quad q^j = A_i^j q_R^i, \tag{8.6.4}$$

and (8.6.2) in matrix notation becomes

$$\hat{\boldsymbol{\omega}} = \dot{A}_t A_t^{-1}. \tag{8.6.5}$$

Newton's Law in a Rotating Frame. Assume that the point $\mathbf{v}(t)$ moves in V according to Newton's second law with a potential energy $U(\mathbf{v})$. Using $U(\mathbf{q})$ for the corresponding function induced on $\mathbb{R}^3$, Newton's law reads

$$m\ddot{\mathbf{q}} = -\nabla U(\mathbf{q}), \tag{8.6.6}$$

which are the Euler–Lagrange equations for

$$L(\mathbf{q}, \dot{\mathbf{q}}) = \frac{m}{2} \langle \dot{\mathbf{q}}, \dot{\mathbf{q}} \rangle - U(\mathbf{q}) \tag{8.6.7}$$

or Hamilton's equations for

$$H(\mathbf{q}, \mathbf{p}) = \frac{1}{2m} \langle \mathbf{p}, \mathbf{p} \rangle + U(\mathbf{q}). \tag{8.6.8}$$

To find the equation satisfied by $\mathbf{q}_R$, differentiate (8.6.4) with respect to time,

$$\dot{\mathbf{q}} = \dot{A}_t \mathbf{q}_R + A_t \dot{\mathbf{q}}_R = \dot{A}_t A_t^{-1} \mathbf{q} + A_t \dot{\mathbf{q}}_R, \tag{8.6.9}$$

that is,

$$\dot{\mathbf{q}} = \boldsymbol{\omega}(t) \times \mathbf{q} + A_t \dot{\mathbf{q}}_R, \tag{8.6.10}$$

where, by abuse of notation, $\boldsymbol{\omega}$ is also used for the representation of $\boldsymbol{\omega}$ in the inertial frame $\mathbf{e}_i$. Differentiating (8.6.10),

$$\begin{aligned}\ddot{\mathbf{q}} &= \dot{\boldsymbol{\omega}} \times \mathbf{q} + \boldsymbol{\omega} \times \dot{\mathbf{q}} + \dot{A}_t \dot{\mathbf{q}}_R + A_t \ddot{\mathbf{q}}_R \\ &= \dot{\boldsymbol{\omega}} \times \mathbf{q} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{q} + A_t \dot{\mathbf{q}}_R) + \dot{A}_t A_t^{-1} A_t \dot{\mathbf{q}}_R + A_t \ddot{\mathbf{q}}_R,\end{aligned}$$

that is,

$$\ddot{\mathbf{q}} = \dot{\boldsymbol{\omega}} \times \mathbf{q} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{q}) + 2(\boldsymbol{\omega} \times A_t \dot{\mathbf{q}}_R) + A_t \ddot{\mathbf{q}}_R. \tag{8.6.11}$$

The ***angular velocity*** in the rotating frame is (see (8.6.4))

$$\boldsymbol{\omega}_R = A_t^{-1} \boldsymbol{\omega}, \quad \text{i.e.,} \quad \boldsymbol{\omega} = A_t \boldsymbol{\omega}_R. \tag{8.6.12}$$

Differentiating (8.6.12) with respect to time gives

$$\dot{\boldsymbol{\omega}} = \dot{A}_t \boldsymbol{\omega}_R + A_t \dot{\boldsymbol{\omega}}_R = \dot{A}_t A_t^{-1} \boldsymbol{\omega} + A_t \dot{\boldsymbol{\omega}}_R = A_t \dot{\boldsymbol{\omega}}_R, \tag{8.6.13}$$

since $\dot{A}_t A_t^{-1} \boldsymbol{\omega} = \boldsymbol{\omega} \times \boldsymbol{\omega} = 0$. Multiplying (8.6.11) by A_t^{-1} gives

$$A_t^{-1} \ddot{\mathbf{q}} = \dot{\boldsymbol{\omega}}_R \times \mathbf{q}_R + \boldsymbol{\omega}_R \times (\boldsymbol{\omega}_R \times \mathbf{q}_R) + 2(\boldsymbol{\omega}_R \times \dot{\mathbf{q}}_R) + \ddot{\mathbf{q}}_R. \tag{8.6.14}$$

Since $m\ddot{\mathbf{q}} = -\nabla U(\mathbf{q})$, we have

$$m A_t^{-1} \ddot{\mathbf{q}} = -\nabla U_R(\mathbf{q}_R), \tag{8.6.15}$$

where the ***rotated potential*** U_R is the *time-dependent* potential defined by

$$U_R(\mathbf{q}_R, t) = U(A_t \mathbf{q}_R) = U(\mathbf{q}), \tag{8.6.16}$$

so that $\nabla U(\mathbf{q}) = A_t \nabla U_R(\mathbf{q}_R)$. Therefore, by (8.6.15), Newton's equations (8.6.6) become

$$\begin{aligned} m\ddot{\mathbf{q}}_R + 2(\boldsymbol{\omega}_R \times m\dot{\mathbf{q}}_R) + m\boldsymbol{\omega}_R \times (\boldsymbol{\omega}_R \times \mathbf{q}_R) + m\dot{\boldsymbol{\omega}}_R \times \mathbf{q}_R \\ = -\nabla U_R(\mathbf{q}_R, t), \end{aligned}$$

that is,

$$\begin{aligned} m\ddot{\mathbf{q}}_R = &- \nabla U_R(\mathbf{q}_R, t) - m\boldsymbol{\omega}_R \times (\boldsymbol{\omega}_R \times \mathbf{q}_R) \\ &- 2m(\boldsymbol{\omega}_R \times \dot{\mathbf{q}}_R) - m\dot{\boldsymbol{\omega}}_R \times \mathbf{q}_R, \end{aligned} \tag{8.6.17}$$

which expresses the equations of motion entirely in terms of rotated quantities.

Ficticious Forces. There are three types of "fictitious forces" that suggest themselves if we try to identify (8.6.17) with $m\mathbf{a} = \mathbf{F}$:

(i) ***Centrifugal force*** $m\boldsymbol{\omega}_R \times (\mathbf{q}_R \times \boldsymbol{\omega}_R)$;

(ii) ***Coriolis force*** $2m\dot{\mathbf{q}}_R \times \boldsymbol{\omega}_R$; and

(iii) ***Euler force*** $m\mathbf{q}_R \times \dot{\boldsymbol{\omega}}_R$.

Note that the Coriolis force $2m\boldsymbol{\omega}_R \times \dot{\mathbf{q}}_R$ is orthogonal to $\boldsymbol{\omega}_R$ and $m\dot{\mathbf{q}}_R$, while the centrifugal force

$$m\boldsymbol{\omega}_R \times (\boldsymbol{\omega}_R \times \mathbf{q}_R) = m[(\boldsymbol{\omega}_R \cdot \mathbf{q}_R)\boldsymbol{\omega}_R - \|\boldsymbol{\omega}_R\|^2 \mathbf{q}_R]$$

is in the plane of $\boldsymbol{\omega}_R$ and $\mathbf{q}_R$. Also note that the Euler force is due to the *nonuniformity* of the rotation rate.

Lagrangian Form. It is of interest to ask the sense in which (8.6.17) is Lagrangian or Hamiltonian. To answer this, it is useful to begin with the Lagrangian approach, which, we will see, is simpler. Substitute (8.6.10) into (8.6.7) to express the Lagrangian in terms of rotated quantities:

$$\begin{aligned} L &= \frac{m}{2} \langle \boldsymbol{\omega} \times \mathbf{q} + A_t \dot{\mathbf{q}}_R, \boldsymbol{\omega} \times \mathbf{q} + A_t \dot{\mathbf{q}}_R \rangle - U(\mathbf{q}) \\ &= \frac{m}{2} \langle \boldsymbol{\omega}_R \times \mathbf{q}_R + \dot{\mathbf{q}}_R, \boldsymbol{\omega}_R \times \mathbf{q}_R + \dot{\mathbf{q}}_R \rangle - U_R(\mathbf{q}_R, t), \end{aligned} \tag{8.6.18}$$

which defines a new (time-dependent!) Lagrangian $L_R(\mathbf{q}_R, \dot{\mathbf{q}}_R, t)$. Remarkably, (8.6.17) are precisely the Euler–Lagrange equations for L_R; that is, (8.6.17) are equivalent to

$$\frac{d}{dt} \frac{\partial L_R}{\partial \dot{\mathbf{q}}_R^i} = \frac{\partial L_R}{\partial \mathbf{q}_R^i},$$

as is readily verified. If one thinks about performing a time-dependent transformation in the variational principle, then in fact, one sees that this is reasonable.

Hamiltonian Form. To find the sense in which (8.6.17) is Hamiltonian, perform a Legendre transformation on L_R. The conjugate momentum is

$$\mathbf{p}_R = \frac{\partial L_R}{\partial \dot{\mathbf{q}}_R} = m(\boldsymbol{\omega}_R \times \mathbf{q}_R + \dot{\mathbf{q}}_R), \tag{8.6.19}$$

and so the Hamiltonian has the expression

$$\begin{aligned} H_R(\mathbf{q}_R, \mathbf{p}_R) &= \langle \mathbf{p}_R, \dot{\mathbf{q}}_R \rangle - L_R \\ &= \frac{1}{m} \langle \mathbf{p}_R, \mathbf{p}_R - m\boldsymbol{\omega}_R \times \mathbf{q}_R \rangle - \frac{1}{2m} \langle \mathbf{p}_R, \mathbf{p}_R \rangle + U_R(\mathbf{q}_R, t) \\ &= \frac{1}{2m} \langle \mathbf{p}_R, \mathbf{p}_R \rangle + U_R(\mathbf{q}_R, t) - \langle \mathbf{p}_R, \boldsymbol{\omega}_R \times \mathbf{q}_R \rangle . \end{aligned} \tag{8.6.20}$$

Thus, (8.6.17) are equivalent to Hamilton's canonical equations with Hamiltonian (8.6.20) and with the canonical symplectic form. In general, H_R is time-dependent. Alternatively, if we perform the momentum shift

$$\mathfrak{p}_R = \mathbf{p}_R - m\boldsymbol{\omega}_R \times \mathbf{q}_R = m\dot{\mathbf{q}}_R, \tag{8.6.21}$$

then we get

$$\begin{aligned} \tilde{H}_R(\mathbf{q}_R, \mathfrak{p}_R) &:= H_R(\mathbf{q}_R, \mathbf{p}_R) \\ &= \frac{1}{2m} \langle \mathfrak{p}_R, \mathfrak{p}_R \rangle + U_R(\mathbf{q}_R) - \frac{m}{2} \|\boldsymbol{\omega}_R \times \mathbf{q}_R\|^2, \end{aligned} \tag{8.6.22}$$

which is in the usual form of kinetic plus potential energy, but now the potential is *amended* by the centrifugal potential $m\|\boldsymbol{\omega}_R \times \mathbf{q}_R\|^2/2$, and the canonical symplectic structure

$$\Omega_{\text{can}} = d\mathbf{q}_R^i \wedge d(\mathbf{p}_R)_i$$

gets transformed, by the momentum shifting lemma, or directly, to

$$d\mathbf{q}_R^i \wedge d(\mathbf{p}_R)_i = d\mathbf{q}_R^i \wedge d(\mathfrak{p}_R)_i + \epsilon_{ijk}\omega_R^i d\mathbf{q}_R^i \wedge d\mathbf{q}_R^j,$$

where ϵ_{ijk} is the alternating tensor. Note that

$$\tilde{\Omega}_R = \tilde{\Omega}_{\text{can}} + *\boldsymbol{\omega}_R, \tag{8.6.23}$$

where $*\boldsymbol{\omega}_R$ means the two-form associated to the vector $\boldsymbol{\omega}_R$, and that (8.6.23) has the same form as the corresponding expression for a particle in a magnetic field (§6.7).

In general, the momentum shift (8.6.21) is time-dependent, so care is needed in interpreting the sense in which the equations for $\mathfrak{p}_R$ and $\mathbf{q}_R$ are Hamiltonian. In fact, the equations should be computed as follows. Let X_H be a Hamiltonian vector field on P and let $\zeta_t : P \to P$ be a *time-dependent* map with generator Y_t:

$$\frac{d}{dt}\zeta_t(z) = Y_t(\zeta_t(z)). \tag{8.6.24}$$

Assume that ζ_t is symplectic for each t. If $\dot{z}(t) = X_H(z(t))$ and we let $w(t) = \zeta_t(z(t))$, then w satisfies

$$\dot{w} = T\zeta_t \cdot X_H(z(t)) + Y_t(\zeta_t(z(t))), \tag{8.6.25}$$

that is,

$$\dot{w} = X_K(w) + Y_t(w) \tag{8.6.26}$$

where $K = H \circ \zeta_t^{-1}$. The extra term Y_t in (8.6.26) is, in the example under consideration, the Euler force.

So far we have been considering a fixed system as seen from different rotating observers. Analogously, one can consider systems that themselves are subjected to a superimposed rotation, an example being the Foucault pendulum. It is clear that the physical behavior in the two cases can be different—in fact, the Foucault pendulum and the example in the next section show that one can get a real physical effect from rotating a system—obviously, rotating observers can cause nontrivial changes in the *description* of a system but cannot make any *physical* difference. Nevertheless, the strategy for the analysis of rotating systems is analogous to the above. The easiest approach, as we have seen, is to transform the Lagrangian. The reader may wish to reread §2.10 for an easy and specific instance of this.

Exercises

◇ **8.6-1.** Generalize the discussion of Newton's law seen in a rotating frame to that of a particle moving in a magnetic field as seen from a rotating observer. Do so first directly and then by Lagrangian methods.

8.7 The Geometric Phase for a Particle in a Hoop

This discussion follows Berry [1985] with some small modifications (due to Marsden, Montgomery, and Ratiu [1990]) necessary for a geometric interpretation of the results. Figure 8.7.1, shows a planar hoop (not necessarily circular) in which a bead slides without friction.

As the bead is sliding, the hoop is rotated in its plane through an angle $\theta(t)$ with angular velocity $\boldsymbol{\omega}(t) = \dot{\theta}(t)\mathbf{k}$. Let s denote the arc length along the hoop, measured from a reference point on the hoop, and let $\mathbf{q}(s)$ be the vector from the origin to the corresponding point on the hoop; thus the shape of the hoop is determined by this function $\mathbf{q}(s)$. The unit tangent vector is $\mathbf{q}'(s)$, and the position of the reference point $\mathbf{q}(s(t))$ relative to an inertial frame in space is $R_{\theta(t)}\mathbf{q}(s(t))$, where R_θ is the rotation in the plane of the hoop through an angle θ. Note that

$$\dot{R}_\theta R_\theta^{-1}\mathbf{q} = \boldsymbol{\omega} \times \mathbf{q} \quad \text{and} \quad R_\theta \boldsymbol{\omega} = \boldsymbol{\omega}.$$

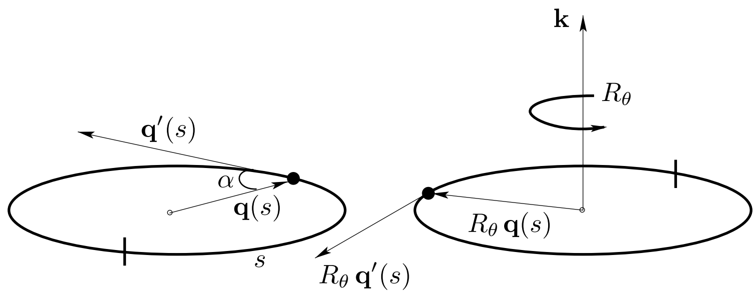

FIGURE 8.7.1. A particle sliding in a rotating hoop.

The Equations of Motion. The configuration space is a fixed closed curve (the hoop) in the plane with length ℓ. The Lagrangian $L(s, \dot{s}, t)$ is simply the kinetic energy of the particle. Since

$$\frac{d}{dt} R_{\theta(t)} \mathbf{q}(s(t)) = R_{\theta(t)} \mathbf{q}'(s(t)) \dot{s}(t) + R_{\theta(t)} [\boldsymbol{\omega}(t) \times \mathbf{q}(s(t))],$$

the Lagrangian is

$$L(s, \dot{s}, t) = \frac{1}{2} m \|\mathbf{q}'(s)\dot{s} + \boldsymbol{\omega} \times \mathbf{q}\|^2. \tag{8.7.1}$$

Note that the momentum conjugate to s is $p = \partial L / \partial \dot{s}$; that is,

$$p = m\mathbf{q}' \cdot [\mathbf{q}'\dot{s} + \boldsymbol{\omega} \times \mathbf{q}] = mv, \tag{8.7.2}$$

where v is the component of the velocity *with respect to the inertial frame* tangent to the curve. The Euler–Lagrange equations

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{s}} = \frac{\partial L}{\partial s}$$

become

$$\frac{d}{dt} [\mathbf{q}' \cdot (\mathbf{q}'\dot{s} + \boldsymbol{\omega} \times \mathbf{q})] = (\mathbf{q}'\dot{s} + \boldsymbol{\omega} \times \mathbf{q}) \cdot (\mathbf{q}''\dot{s} + \boldsymbol{\omega} \times \mathbf{q}').$$

Using $\|\mathbf{q}'\|^2 = 1$, its consequence $\mathbf{q}' \cdot \mathbf{q}'' = 0$, and simplifying, we get

$$\ddot{s} + \mathbf{q}' \cdot (\dot{\boldsymbol{\omega}} \times \mathbf{q}) - (\boldsymbol{\omega} \times \mathbf{q}) \cdot (\boldsymbol{\omega} \times \mathbf{q}') = 0. \tag{8.7.3}$$

The second and third terms in (8.7.3) are the Euler and centrifugal forces, respectively. Since $\boldsymbol{\omega} = \dot{\theta}\mathbf{k}$, we can rewrite (8.7.3) as

$$\ddot{s} = \dot{\theta}^2 \mathbf{q} \cdot \mathbf{q}' - \ddot{\theta} q \sin\alpha, \tag{8.7.4}$$

where α is as in Figure 8.7.1 and $q = \|\mathbf{q}\|$.

Averaging. From (8.7.4) and Taylor's formula with remainder, we get

$$s(t) = s_0 + \dot{s}_0 t + \int_0^t (t-\tau)\{\dot{\theta}(\tau)^2 \mathbf{q}(s(\tau)) \cdot \mathbf{q}'(s(\tau)) \\ - \ddot{\theta}(\tau) q(s(\tau)) \sin\alpha(s(\tau))\}\, d\tau. \qquad (8.7.5)$$

The angular velocity $\dot{\theta}$ and acceleration $\ddot{\theta}$ are assumed small with respect to the particle's velocity, so by the averaging theorem (see, for example, Hale [1963]), the s-dependent quantities in (8.7.5) can be replaced by their averages around the hoop:

$$s(t) \approx s_0 + \dot{s}_0 t + \int_0^t (t-\tau)\left\{\dot{\theta}(\tau)^2 \frac{1}{\ell}\int_0^\ell \mathbf{q}\cdot\mathbf{q}'\, ds \right. \\ \left. - \ddot{\theta}(\tau)\frac{1}{\ell}\int_0^\ell q(s)\sin\alpha(s)\, ds \right\} d\tau. \qquad (8.7.6)$$

Technical Aside. The essence of averaging in this case can be seen as follows. Suppose $g(t)$ is a rapidly varying function whose oscillations are bounded in magnitude by a constant C and $f(t)$ is slowly varying on an interval $[a,b]$. Over one period of g, say $[\alpha,\beta]$, we have

$$\int_\alpha^\beta f(t)g(t)\, dt \approx \overline{g}\int_\alpha^\beta f(t)\, dt, \qquad (8.7.7)$$

where

$$\overline{g} = \frac{1}{\beta-\alpha}\int_\alpha^\beta g(t)\, dt$$

is the average of g. The assumption that the oscillations of g are bounded by C means that

$$|g(t) - \overline{g}| \le C \quad \text{for all } t \in [\alpha,\beta].$$

The error in (8.7.7) is $\int_\alpha^\beta f(t)(g(t)-\overline{g})\, dt$, whose absolute value is bounded as follows. Let M be the maximum value of f on $[\alpha,\beta]$ and m be the minimum. Then

$$\left|\int_\alpha^\beta f(t)[g(t)-\overline{g}]\, dt\right| = \left|\int_\alpha^\beta (f(t)-m)[g(t)-\overline{g}]\, dt\right| \\ \le (\beta-\alpha)(M-m)C \\ \le (\beta-\alpha)^2 DC,$$

where D is the maximum of $|f'(t)|$ for $\alpha \le t \le \beta$. Now these errors over each period are added up over $[a,b]$. Since the error estimate has the *square* of $\beta-\alpha$ as a factor, one still gets something small as the period of g tends to 0.

In (8.7.5) we change variables from t to s, do the averaging, and then change back.

The Phase Formula. The first inner integral in (8.7.6) over s vanishes (since the integrand is $(d/ds)\|\mathbf{q}(s)\|^2$), and the second is $2A$, where A is the area enclosed by the hoop. Integrating by parts,

$$\int_0^T (T-\tau)\ddot{\theta}(\tau)\,d\tau = -T\dot{\theta}(0) + \int_0^T \dot{\theta}(\tau)\,d\tau = -T\dot{\theta}(0) + 2\pi, \qquad (8.7.8)$$

assuming that the hoop makes one complete revolution in time T. Substituting (8.7.8) in (8.7.6) gives

$$s(T) \approx s_0 + \dot{s}_0 T + \frac{2A}{\ell}\dot{\theta}_0 T - \frac{4\pi A}{\ell}, \qquad (8.7.9)$$

where $\dot{\theta}_0 = \dot{\theta}(0)$. The initial velocity of the bead *relative to the hoop* is $\dot{s}_0$, while its component along the curve *relative to the inertial frame* is (see (8.7.2))

$$v_0 = \mathbf{q}'(0) \cdot [\mathbf{q}'(0)\dot{s}_0 + \boldsymbol{\omega}_0 \times \mathbf{q}(0)] = \dot{s}_0 + \boldsymbol{\omega}_0 q(s_0)\sin\alpha(s_0). \qquad (8.7.10)$$

Now we replace $\dot{s}_0$ in (8.7.9) by its expression in terms of v_0 from (8.7.10) and average over all initial conditions to get

$$\langle s(T) - s_0 - v_0 T\rangle = -\frac{4\pi A}{\ell}, \qquad (8.7.11)$$

which means that *on average*, the shift in position is by $4\pi A/\ell$ between the rotated and nonrotated hoop. Note that if $\dot{\theta}_0 = 0$ (the situation assumed by Berry [1985]), then averaging over initial conditions is not necessary.

This extra length $4\pi A/\ell$ is sometimes called the geometric phase or the ***Berry–Hannay phase***. This example is related to a number of interesting effects, both classically and quantum-mechanically, such as the Foucault pendulum and the Aharonov–Bohm effect. The effect is known as *holonomy* and can be viewed as an instance of *reconstruction* in the context of symmetry and reduction. For further information and additional references, see Aharonov and Anandan [1987], Montgomery [1988], Montgomery [1990], and Marsden, Montgomery, and Ratiu [1989, 1990]. For related ideas in soliton dynamics, see Alber and Marsden [1992].

Exercises

◇ **8.7-1.** Consider the dynamics of a ball in a slowly rotating planar hoop, as in the text. However, this time, consider rotating the hoop about an axis that is not perpendicular to the plane of the hoop, but makes an angle θ with the normal. Compute the geometric phase for this problem.

◇ **8.7-2.** Study the geometric phase for a particle in a general spatial hoop that is moved through a closed curve in SO(3).

◇ **8.7-3.** Consider the dynamics of a ball in a slowly rotating planar hoop, as in the text. However, this time, consider a charged particle with charge e and a fixed magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ in the vicinity of the hoop. Compute the geometric phase for this problem.

8.8 Moving Systems

The particle in the rotating hoop is an example of a rotated or, more generally, a *moving system*. Other examples are a pendulum on a merry-go-round (Exercise 8.8-4) and a fluid on a rotating sphere (like the Earth's ocean and atmosphere). As we have emphasized, systems of this type are not to be confused with rotating observers! Actually rotating a system causes real physical effects, such as the trade winds and hurricanes.

This section develops a general context for such systems. Our purpose is to show how to systematically derive Lagrangians and the resulting equations of motion for moving systems, like the bead in the hoop of the last section. This will also prepare the reader who wants to pursue the question of how moving systems fit in the context of phases (Marsden, Montgomery, and Ratiu [1990]).

The Lagrangian. Consider a Riemannian manifold $\mathcal{S}$, a submanifold Q, and a space M of embeddings of Q into $\mathcal{S}$. Let $m_t \in M$ be a given curve. If a particle in Q is following a curve $q(t)$, and if Q moves by superposing the motion m_t, then the path of the particle in $\mathcal{S}$ is given by $m_t(q(t))$. Thus, its velocity in $\mathcal{S}$ is given by

$$T_{q(t)}m_t \cdot \dot{q}(t) + \mathcal{Z}_t(m_t(q(t))), \tag{8.8.1}$$

where $\mathcal{Z}_t(m_t(q)) = (d/dt)m_t(q)$. Consider a Lagrangian on TQ of the usual form of kinetic minus potential energy:

$$L_{m_t}(q, v) = \frac{1}{2}\|T_{q(t)}m_t \cdot v + \mathcal{Z}_t(m_t(q))\|^2 - V(q) - U(m_t(q)), \tag{8.8.2}$$

where V is a given potential on Q, and U is a given potential on $\mathcal{S}$.

The Hamiltonian. We now compute the Hamiltonian associated to this Lagrangian by taking the associated Legendre transform. If we take the derivative of (8.8.2) with respect to v in the direction of w, we obtain

$$\frac{\partial L_{m_t}}{\partial v} \cdot w = p \cdot w = \left\langle T_{q(t)}m_t \cdot v + \mathcal{Z}_t\left(m_t(q(t))\right)^T, T_{q(t)}m_t \cdot w \right\rangle_{m_t(q(t))}, \tag{8.8.3}$$

where $p \cdot w$ means the natural pairing between the covector $p \in T^*_{q(t)}Q$ and the vector $w \in T_{q(t)}Q$, while $\langle \, , \rangle_{m_t(q(t))}$ denotes the metric inner product on $\mathcal{S}$ at the point $m_t(q(t))$ and ^{T} denotes the orthogonal projection to the tangent space $Tm_t(Q)$ using the metric of $\mathcal{S}$ at $m_t(q(t))$. We endow Q with the (possibly time-dependent) metric induced by the mapping m_t. In other words, we choose the metric on Q that makes m_t into an isometry for each t. Using this definition, (8.8.3) gives

$$p \cdot w = \left\langle v + \left(T_{q(t)} m_t\right)^{-1} \cdot \mathcal{Z}_t \left(m_t(q(t))\right)^T, w \right\rangle_{q(t)};$$

that is,

$$p = \left(v + \left(T_{q(t)} m_t\right)^{-1} \cdot \left[\mathcal{Z}_t \left(m_t(q(t))^T\right]\right)^\flat, \tag{8.8.4}$$

where $\flat$ is the index-lowering operation at $q(t)$ using the metric on Q.

Physically, if $\mathcal{S}$ is $\mathbb{R}^3$, then p is the inertial momentum (see the hoop example in the preceding section). This extra term $\mathcal{Z}_t(m_t(q))^T$ is associated with a connection called the ***Cartan connection*** on the bundle $Q \times M \to M$, with horizontal lift defined to be $\mathcal{Z}(m) \mapsto (Tm^{-1} \cdot \mathcal{Z}(m)^T, \mathcal{Z}(m))$. (See, for example, Marsden and Hughes [1983] for an account of some aspects of Cartan's contributions.)

The corresponding Hamiltonian (given by the standard prescription $H = pv - L$) picks up a cross term and takes the form

$$H_{m_t}(q,p) = \frac{1}{2}\|p\|^2 - \mathcal{P}(Z_t) - \frac{1}{2}\|\mathcal{Z}_t^\perp\|^2 + V(q) + U(m_t(q)), \tag{8.8.5}$$

where the time-dependent vector field Z_t on Q is defined by

$$Z_t(q) = \left(T_{q(t)} m_t\right)^{-1} \cdot [\mathcal{Z}_t(m_t(q)]^T$$

and where $\mathcal{P}(Z_t(q))(q,p) = \langle p, Z_t(q)\rangle$ and $\mathcal{Z}_t^\perp$ denotes the component perpendicular to $m_t(Q)$. The Hamiltonian vector field of this cross term, namely $X_{\mathcal{P}(\mathcal{Z}_t)}$, represents the noninertial forces and also has the natural interpretation as a horizontal lift of the vector field $\mathcal{Z}_t$ relative to a certain connection on the bundle $T^*Q \times M \to M$, naturally derived from the Cartan connection.

Remarks on Averaging. Let G be a Lie group that acts on T^*Q in a Hamiltonian fashion and leaves H_0 (defined by setting $\mathcal{Z} = 0$ and $U = 0$ in (8.8.5)) invariant. (Lie groups are discussed in the next chapter, so these remarks can be omitted on a first reading.) In our examples, G is either $\mathbb{R}$ acting on T^*Q by the flow of H_0 (the hoop), or a subgroup of the isometry group of Q that leaves V and U invariant, and acts on T^*Q by cotangent lift (this is appropriate for the Foucault pendulum). In any case, we assume that G has an invariant measure relative to which we can average.

Assuming the "averaging principle" (see Arnold [1989], for example) we replace H_{m_t} by its G-average,

$$\langle H_{m_t}\rangle (q,p) = \frac{1}{2}\|p\|^2 - \langle \mathcal{P}(Z_t)\rangle - \frac{1}{2}\left\langle \|\mathcal{Z}_t^\perp\|^2\right\rangle + V(q) + \langle U(m_t(q))\rangle . \tag{8.8.6}$$

In (8.8.6) we shall assume that the term $\frac{1}{2}\left\langle \|\mathcal{Z}_t^\perp\|^2\right\rangle$ is small and discard it. Thus, define

$$\begin{aligned}\mathcal{H}(q,p,t) &= \frac{1}{2}\|p\|^2 - \langle \mathcal{P}(Z_t)\rangle + V(q) + \langle U(m_t(q))\rangle \\ &= \mathcal{H}_0(q,p) - \langle \mathcal{P}(Z_t)\rangle + \langle U(m_t(q))\rangle .\end{aligned} \tag{8.8.7}$$

Consider the dynamics on $T^*Q \times M$ given by the vector field

$$(X_{\mathcal{H}}, Z_t) = (X_{\mathcal{H}_0} - X_{\langle \mathcal{P}(Z_t)\rangle} + X_{\langle U \circ m_t\rangle}, Z_t). \tag{8.8.8}$$

The vector field, consisting of the extra terms in this representation due to the superposed motion of the system, namely

$$\operatorname{hor}(Z_t) = (-X_{\langle \mathcal{P}(Z_t)\rangle}, Z_t), \tag{8.8.9}$$

has a natural interpretation as the horizontal lift of Z_t relative to a connection on $T^*Q \times M$, which is obtained by averaging the Cartan connection and is called the ***Cartan–Hannay–Berry connection***. The holonomy of this connection is the ***Hannay–Berry phase*** of a slowly moving constrained system. For details of this approach, see Marsden, Montgomery, and Ratiu [1990].

Exercises

◇ **8.8-1.** Consider the particle in a hoop of §8.7. For this problem, identify all the elements of formula (8.8.2) and use that identification to obtain the Lagrangian (8.7.1).

◇ **8.8-2.** Consider the particle in a rotating hoop discussed in §2.8.

(a) Use the tools of this section to obtain the Lagrangian given in §2.8.

(b) Suppose that the hoop rotates freely. Can you still use the tools of part (a)? If so, compute the new Lagrangian and point out the differences between the two cases.

(c) Analyze, in the same fashion as in §2.8, the equilibria of the free system. Does this system also bifurcate?

◇ **8.8-3.** Set up the equations for the Foucault pendulum using the ideas in this section.

◇ **8.8-4.** Consider again the mechanical system in Exercise 2.8-6, but this time hang a *spherical* pendulum from the rotating arm. Investigate the geometric phase when the arm is swung once around. (Consider doing the experiment!) Is the term $\|Z_t^\perp\|^2$ really small in this example?

8.9 Routh Reduction

An abelian version of Lagrangian reduction was known to Routh by around 1860. A modern account was given in Arnold [1988], and motivated by that, Marsden and Scheurle [1993a] gave a geometrization and a generalization of the Routh procedure to the nonabelian case.

In this section we give an elementary classical description in preparation for more sophisticated reduction procedures, such as Euler–Poincaré reduction in Chapter 13.

We assume that Q is a product of a manifold S and a number, say k, of copies of the circle S^1, namely $Q = S \times (S^1 \times \cdots \times S^1)$. The factor S, called ***shape space***, has coordinates denoted by $x^1, \ldots, x^m$, and coordinates on the other factors are written $\theta^1, \ldots, \theta^k$. Some or all of the factors of S^1 can be replaced by $\mathbb{R}$ if desired, with little change. We assume that the variables θ^a, $a = 1, \ldots, k$, are ***cyclic***, that is, they do not appear explicitly in the Lagrangian, although their velocities do.

As we shall see after Chapter 9 is studied, invariance of L under the action of the abelian group $G = S^1 \times \cdots \times S^1$ is another way to express that fact that θ^a are cyclic variables. That point of view indeed leads ultimately to deeper insight, but here we focus on some basic calculations done "by hand" in coordinates.

A basic class of examples (for which Exercises 8.9-1 and 8.9-2 provide specific instances) are those for which the Lagrangian L has the form kinetic minus potential energy:

$$L(x, \dot{x}, \dot{\theta}) = \frac{1}{2} g_{\alpha\beta}(x) \dot{x}^\alpha \dot{x}^\beta + g_{a\alpha}(x) \dot{x}^\alpha \dot{\theta}^a + \frac{1}{2} g_{ab}(x) \dot{\theta}^a \dot{\theta}^b - V(x), \quad (8.9.1)$$

where there is a sum over α, β from 1 to m and over a, b from 1 to k. Even in simple examples, such as the double spherical pendulum or the simple pendulum on a cart (Exercise 8.9-2), the matrices $g_{\alpha\beta}$, $g_{a\alpha}$, g_{ab} can depend on x.

Because θ^a are cyclic, the corresponding conjugate momenta

$$p_a = \frac{\partial L}{\partial \dot{\theta}^a} \quad (8.9.2)$$

are conserved quantities. In the case of the Lagrangian (8.9.1), these momenta are given by

$$p_a = g_{a\alpha} \dot{x}^\alpha + g_{ab} \dot{\theta}^b.$$

Definition 8.9.1. *The **classical Routhian** is defined by setting* $p_a = \mu_a = \text{constant}$ *and performing a partial Legendre transformation in the variables* θ^a :

$$R^\mu(x, \dot{x}) = \left[L(x, \dot{x}, \dot{\theta}) - \mu_a \dot{\theta}^a \right] \Big|_{p_a=\mu_a}, \tag{8.9.3}$$

where it is understood that the variable $\dot{\theta}^a$ *is eliminated using the equation* $p_a = \mu_a$ *and* μ_a *is regarded as a constant.*

Now consider the Euler–Lagrange equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}^a} - \frac{\partial L}{\partial x^a} = 0; \tag{8.9.4}$$

we attempt to write these as Euler–Lagrange equations for a function from which $\dot{\theta}^a$ has been eliminated. We claim that the Routhian R^μ does the job. To see this, we compute the Euler–Lagrange expression for R^μ using the chain rule:

$$\begin{aligned} \frac{d}{dt}\left(\frac{\partial R^\mu}{\partial \dot{x}^\alpha}\right) - \frac{\partial R^\mu}{\partial x^\alpha} &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^\alpha} + \frac{\partial L}{\partial \dot{\theta}^a}\frac{\partial \dot{\theta}^a}{\partial \dot{x}^\alpha}\right) \\ &\quad - \left(\frac{\partial L}{\partial x^\alpha} + \frac{\partial L}{\partial \dot{\theta}^a}\frac{\partial \dot{\theta}^a}{\partial x^\alpha}\right) - \frac{d}{dt}\left(\mu_a \frac{\partial \dot{\theta}^a}{\partial \dot{x}^\alpha}\right) + \mu_a \frac{\partial \dot{\theta}^a}{\partial x^\alpha}. \end{aligned}$$

The first and third terms vanish by (8.9.4), and the remaining terms vanish using $\mu_a = p_a$. Thus, we have proved the following result.

Proposition 8.9.2. *The Euler–Lagrange equations* (8.9.4) *for* $L(x, \dot{x}, \dot{\theta})$ *together with the conservation laws* $p_a = \mu_a$ *are equivalent to the Euler–Lagrange equations for the Routhian* $R^\mu(x, \dot{x})$ *together with* $p_a = \mu_a$.

The Euler–Lagrange equations for R^μ are called the ***reduced Euler–Lagrange equations***, since the configuration space Q with variables (x^a, θ^a) has been *reduced* to the configuration space S with variables x^α.

In what follows we shall make the following notational conventions: g^{ab} denote the entries of the inverse matrix of the $m \times m$ matrix $[g_{ab}]$, and similarly, $g^{\alpha\beta}$ denote the entries of the inverse of the $k \times k$ matrix $[g_{\alpha\beta}]$. We will not use the entries of the inverse of the whole matrix tensor on Q, so there is no danger of confusion.

Proposition 8.9.3. *For* L *given by* (8.9.1) *we have*

$$R^\mu(x, \dot{x}) = g_{a\alpha} g^{ac} \mu_c \dot{x}^\alpha + \frac{1}{2}\left(g_{\alpha\beta} - g_{a\alpha} g^{ac} g_{c\beta}\right) \dot{x}^\alpha \dot{x}^\beta - V_\mu(x), \tag{8.9.5}$$

where

$$V_\mu(x) = V(x) + \frac{1}{2} g^{ab} \mu_a \mu_b$$

*is the **amended potential**.*

Proof. We have $\mu_a = g_{a\alpha}\dot{x}^\alpha + g_{ab}\dot{\theta}^b$, so

$$\dot{\theta}^a = g^{ab}\mu_b - g^{ab}g_{b\alpha}\dot{x}^\alpha. \tag{8.9.6}$$

Substituting this in the definition of R^μ gives

$$\begin{aligned} R^\mu(x,\dot{x}) = {} & \frac{1}{2}g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta + (g_{a\alpha}\dot{x}^\alpha)\left(g^{ac}\mu_c - g^{ac}g_{c\beta}\dot{x}^\beta\right) \\ & + \frac{1}{2}g_{ab}\left(g^{ac}\mu_c - g^{ac}g_{c\beta}\dot{x}^\beta\right)\left(g^{bd}\mu_d - g^{bd}g_{d\gamma}\dot{x}^\gamma\right) \\ & - \mu_a\left(g^{ac}\mu_c - g^{ac}g_{c\beta}\dot{x}^\beta\right) - V(x). \end{aligned}$$

The terms linear in $\dot{x}$ are

$$g_{a\alpha}g^{ac}\mu_c\dot{x}^\alpha - g_{ab}g^{ac}\mu_c g^{bd}g_{d\gamma}\dot{x}^\gamma + \mu_a g^{ac}g_{c\beta}\dot{x}^\beta = g_{a\alpha}g^{ac}\mu_c\dot{x}^\alpha,$$

while the terms quadratic in $\dot{x}$ are

$$\frac{1}{2}(g_{\alpha\beta} - g_{a\alpha}g^{ac}g_{c\beta})\dot{x}^\alpha\dot{x}^\beta,$$

and the terms dependent only on x are $-V_\mu(x)$, as required. ■

Note that R^μ has picked up a term linear in the velocity, and the potential as well as the kinetic energy matrix (the ***mass matrix***) have both been modified.

The term linear in the velocities has the form $A^a_\alpha\mu_a\dot{x}^\alpha$, where $A^a_\alpha = g^{ab}g_{b\alpha}$. The Euler–Lagrange expression for this term can be written

$$\frac{d}{dt}A^a_\alpha\mu_a - \frac{\partial}{\partial x^\alpha}A^a_\beta\mu_a\dot{x}^\beta = \left(\frac{\partial A^a_\alpha}{\partial x^\beta} - \frac{\partial A^a_\beta}{\partial x^\alpha}\right)\mu_a\dot{x}^\beta,$$

which is denoted by $B^a_{\alpha\beta}\mu_a\dot{x}^\beta$. If we think of the one-form $A^a_\alpha dx^\alpha$, then $B^a_{\alpha\beta}$ is its exterior derivative. The quantities A^a_α are called ***connection coefficients***, and $B^a_{\alpha\beta}$ are called the ***curvature coefficients***.

Introducing the modified (simpler) Routhian, obtained by deleting the terms linear in $\dot{x}$,

$$\tilde{R}^\mu = \frac{1}{2}\left(g_{\alpha\beta} - g_{a\alpha}g^{ab}g_{b\beta}\right)\dot{x}^\alpha\dot{x}^\beta - V_\mu(x),$$

the equations take the form

$$\frac{d}{dt}\frac{\partial\tilde{R}^\mu}{\partial\dot{x}^\alpha} - \frac{\partial\tilde{R}^\mu}{\partial x^\alpha} = -B^a_{\alpha\beta}\mu_a\dot{x}^\beta, \tag{8.9.7}$$

which is the form that makes intrinsic sense and generalizes to the case of nonabelian groups. The extra terms have the structure of magnetic, or Coriolis, terms that we have seen in a variety of earlier contexts.

The above gives a hint of the large amount of geometry hidden behind the apparently simple process of Routh reduction. In particular, *connections* A^a_α and their *curvatures* $B^a_{\alpha\beta}$ play an important role in more general theories, such as those involving nonablelian symmetry groups (like the rotation group).

Another suggestive hint of more general theories is that the kinetic term in (8.9.5) can be written in the following way:

$$\frac{1}{2}(\dot{x}^\alpha, -A^a_\delta \dot{x}^\delta) \begin{pmatrix} g_{\alpha\beta} & g_{\alpha b} \\ g_{a\beta} & g_{ab} \end{pmatrix} \begin{pmatrix} \dot{x}^\beta \\ -A^b_\gamma \dot{x}^\gamma \end{pmatrix},$$

which also exhibits its positive definite nature.

Routh himself (in the mid 1800s) was very interested in rotating mechanical systems, such as those possessing an angular momentum conservation law. In this context, Routh used the term "steady motion" for dynamic motions that were uniform rotations about a fixed axis. *We may identify these with equilibria of the reduced Euler–Lagrange equations*.

Since the Coriolis term does not affect conservation of energy (we have seen this earlier with the dynamics of a particle in a magnetic field), we can apply the Lagrange–Dirichlet test to reach the following conclusion:

Proposition 8.9.4 (Routh's Stability Criterion)**.** *Steady motions correspond to critical points x_e of the amended potential V_μ. If $d^2V_\mu(x_e)$ is positive definite, then the steady motion x_e is stable.*

When more general symmetry groups are involved, one speaks of ***relative equilibria*** rather than steady motions, a change of terminology due to Poincaré around 1890. This is the beginning of a more sophisticated theory of stability, leading up to the ***energy–momentum method*** outlined in §1.7.

Exercises

◇ **8.9-1.** Carry out Routh reduction for the spherical pendulum.

◇ **8.9-2.** Carry out Routh reduction for the planar pendulum on a cart, as in Figure 8.9.1.

◇ **8.9-3** (Two-body problem)**.** Compute the amended potential for the planar motion of a particle moving in a central potential $V(r)$. Compare the result with the "effective potential" found in, for example, Goldstein [1980].

◇ **8.9-4.** Let L be a Lagrangian on TQ and let

$$\hat{R}^\mu(q, \dot{q}) = L(q, \dot{q}) + A^a_\alpha \mu_a \dot{q}^a,$$

where A^a is an $\mathbb{R}^k$-valued one-form on TQ and $\mu \in \mathbb{R}^{k*}$.

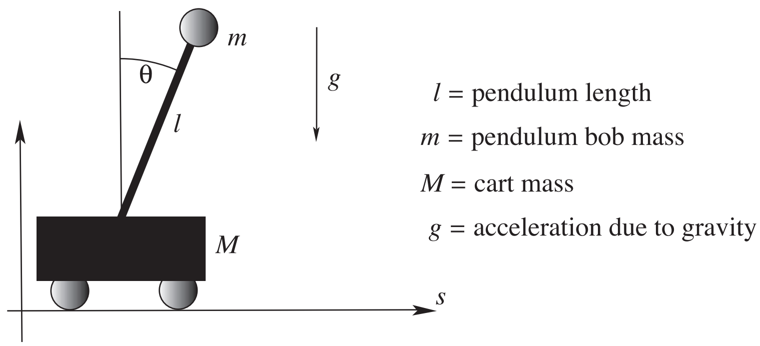

FIGURE 8.9.1. A pendulum on a cart.

(a) Write Hamilton's principle for L as a Lagrange–d'Alembert principle for $\hat{R}^\mu$.

(b) Letting $\hat{H}^\mu$ be the Hamiltonian associated with $\hat{R}^\mu$, show that the original Euler–Lagrange equations for L can be written as

$$\dot{q}^\alpha = \frac{\partial \hat{H}^\mu}{\partial p_\alpha},$$

$$\dot{p}_\alpha = \frac{\partial \hat{H}^\mu}{\partial q^\alpha} + \beta^a_{\alpha\beta}\mu_b \frac{\partial \hat{H}^\mu}{\partial p_\beta}.$$