7
Lagrangian Mechanics

Our approach so far has emphasized the Hamiltonian point of view. However, there is an independent point of view, that of Lagrangian mechanics, based on variational principles. This alternative viewpoint, computational convenience, and the fact that the Lagrangian is very useful in covariant relativistic theories can be used as arguments for the importance of the Lagrangian formulation. Ironically, it was Hamilton [1834] who discovered the variational basis of Lagrangian mechanics.

7.1 Hamilton's Principle of Critical Action

Much of mechanics can be based on variational principles. Indeed, it is the variational formulation that is the most covariant, being useful for relativistic systems as well. In the next chapter we shall see the utility of the Lagrangian approach in the study of rotating frames and moving systems, and we will also use it as an important way to approach Hamilton–Jacobi theory.

Consider a ***configuration manifold*** Q and the velocity phase space TQ. We consider a function $L : TQ \to \mathbb{R}$ called the ***Lagrangian***. Speaking informally, Hamilton's ***principle of critical action*** states that

$$\delta \int L\left(q^i, \frac{dq^i}{dt}\right) dt = 0, \tag{7.1.1}$$

where we take variations among paths $q^i(t)$ in Q with fixed endpoints. (We will study this process a little more carefully in §8.1.) Taking the variation in (7.1.1), the chain rule gives

$$\int \left[\frac{\partial L}{\partial q^i} \delta q^i + \frac{\partial L}{\partial \dot{q}^i} \frac{d}{dt} \delta q^i \right] dt \tag{7.1.2}$$

for the left-hand side. Integrating the second term by parts and using the boundary conditions $\delta q^i = 0$ at the endpoints of the time interval in question, we get

$$\int \left[\frac{\partial L}{\partial q^i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) \right] \delta q^i \, dt = 0. \tag{7.1.3}$$

If this is to hold for all such variations $\delta q^i(t)$, then

$$\frac{\partial L}{\partial q^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} = 0, \tag{7.1.4}$$

which are the ***Euler–Lagrange equations***.

We set $p_i = \partial L / \partial \dot{q}^i$, assume that the transformation $(q^i, \dot{q}^j) \mapsto (q^i, p_j)$ is invertible, and define the ***Hamiltonian*** by

$$H(q^i, p_j) = p_i \dot{q}^i - L(q^i, \dot{q}^i). \tag{7.1.5}$$

Note that

$$\dot{q}^i = \frac{\partial H}{\partial p_i},$$

since

$$\frac{\partial H}{\partial p_i} = \dot{q}^i + p_j \frac{\partial \dot{q}^j}{\partial p_i} - \frac{\partial L}{\partial \dot{q}^j} \frac{\partial \dot{q}^j}{\partial p_i} = \dot{q}^i$$

from (7.1.5) and the chain rule. Likewise,

$$\dot{p}_i = -\frac{\partial H}{\partial q^i}$$

from (7.1.4) and

$$\frac{\partial H}{\partial q^j} = p_i \frac{\partial \dot{q}^i}{\partial q^j} - \frac{\partial L}{\partial q^j} - \frac{\partial L}{\partial \dot{q}^i} \frac{\partial \dot{q}^i}{\partial q^j} = -\frac{\partial L}{\partial q^j}.$$

In other words, the *Euler–Lagrange equations are equivalent to Hamilton's equations*.

Thus, it is reasonable to explore the geometry of the Euler–Lagrange equations using the canonical form on T^*Q pulled back to TQ using $p_i = \partial L / \partial \dot{q}^i$. We do this in the next sections.

This is one standard way to approach the geometry of the Euler–Lagrange equations. Another *is to use the variational principle itself*. The reader will notice that the canonical one-form $p_i dq^i$ appears as the *boundary terms* when we take the variations. This can, in fact, be used as a basis for the introduction of the canonical one-form in Lagrangian mechanics. We shall develop this approach in Chapter 8. See also Exercise 7.2-2.

Exercises

◇ **7.1-1.** Verify that the Euler–Lagrange and Hamilton equations are equivalent, even if L is time-dependent.

◇ **7.1-2.** Show that the conservation of energy equation results if in Hamilton's principle, variations corresponding to reparametrizations of the given curve $q(t)$ are chosen.

7.2 The Legendre Transform

Fiber Derivatives. Given a Lagrangian $L : TQ \to \mathbb{R}$, define a map $\mathbb{F}L : TQ \to T^*Q$, called the ***fiber derivative***, by

$$\mathbb{F}L(v) \cdot w = \left.\frac{d}{ds}\right|_{s=0} L(v + sw), \tag{7.2.1}$$

where $v, w \in T_qQ$. Thus, $\mathbb{F}L(v) \cdot w$ is the derivative of L at v along the fiber T_qQ in the direction w. Note that $\mathbb{F}L$ is fiber-preserving; that is, it maps the fiber T_qQ to the fiber T_q^*Q. In a local chart $U \times E$ for TQ, where U is open in the model space E for Q, the fiber derivative is given by

$$\mathbb{F}L(u, e) = (u, \mathbf{D}_2 L(u, e)), \tag{7.2.2}$$

where $\mathbf{D}_2 L$ denotes the partial derivative of L with respect to its second argument. For finite-dimensional manifolds, with (q^i) denoting coordinates on Q and $(q^i, \dot{q}^i)$ the induced coordinates on TQ, the fiber derivative has the expression

$$\mathbb{F}L(q^i, \dot{q}^i) = \left(q^i, \frac{\partial L}{\partial \dot{q}^i} \right), \tag{7.2.3}$$

that is, $\mathbb{F}L$ is given by

$$p_i = \frac{\partial L}{\partial \dot{q}^i}. \tag{7.2.4}$$

The ***associated energy function*** is defined by $E(v) = \mathbb{F}L(v) \cdot v - L(v)$.

In many examples it is the relationship (7.2.4) that gives physical meaning to the momentum variables. We call $\mathbb{F}L$ the ***Legendre transform***.

Lagrangian Forms. Let Ω denote the canonical symplectic form on T^*Q. Using $\mathbb{F}L$, we obtain a one-form Θ_L and a closed two-form Ω_L on TQ by setting

$$\Theta_L = (\mathbb{F}L)^*\Theta \quad \text{and} \quad \Omega_L = (\mathbb{F}L)^*\Omega. \tag{7.2.5}$$

We call Θ_L the ***Lagrangian one-form*** and Ω_L the ***Lagrangian two-form***. Since $\mathbf{d}$ commutes with pull-back, we get $\Omega_L = -\mathbf{d}\Theta_L$. Using the local expressions for Θ and Ω, a straightforward pull-back computation yields the following local formula for Θ_L and Ω_L: If E is the model space for Q, U is the range in E of a chart on Q, and $U \times E$ is the corresponding range of the induced chart on TQ, then for $(u, e) \in U \times E$ and tangent vectors $(e_1, e_2), (f_1, f_2)$ in $E \times E$, we have

$$\begin{aligned} &T_{(u,e)}\mathbb{F}L \cdot (e_1, e_2) \\ &\qquad = (u, \mathbf{D}_2 L(u,e), e_1, \mathbf{D}_1(\mathbf{D}_2 L(u,e)) \cdot e_1 + \mathbf{D}_2(\mathbf{D}_2 L(u,e)) \cdot e_2), \end{aligned} \tag{7.2.6}$$

so that using the local expression for Θ and the definition of pull-back,

$$\Theta_L(u,e) \cdot (e_1, e_2) = \mathbf{D}_2 L(u,e) \cdot e_1. \tag{7.2.7}$$

Similarly, one finds that

$$\begin{aligned} &\Omega_L(u,e) \cdot ((e_1, e_2), (f_1, f_2)) \\ &\qquad = \mathbf{D}_1(\mathbf{D}_2 L(u,e) \cdot e_1) \cdot f_1 - \mathbf{D}_1(\mathbf{D}_2 L(u,e) \cdot f_1) \cdot e_1 \\ &\qquad\quad + \mathbf{D}_2\mathbf{D}_2 L(u,e) \cdot e_1 \cdot f_2 - \mathbf{D}_2\mathbf{D}_2 L(u,e) \cdot f_1 \cdot e_2, \end{aligned} \tag{7.2.8}$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ denote the first and second partial derivatives. In finite dimensions, formulae (7.2.6) and (7.2.7) or a direct pull-back of $p_i dq^i$ and $dq^i \wedge dp_i$ yields

$$\Theta_L = \frac{\partial L}{\partial \dot{q}^i} dq^i \tag{7.2.9}$$

and

$$\Omega_L = \frac{\partial^2 L}{\partial \dot{q}^i\, \partial q^j} dq^i \wedge dq^j + \frac{\partial^2 L}{\partial \dot{q}^i\, \partial \dot{q}^j} dq^i \wedge d\dot{q}^j \tag{7.2.10}$$

(a sum on all i, j is understood). As a $2n \times 2n$ skew-symmetric matrix,

$$\Omega_L = \begin{bmatrix} A & \left[\dfrac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}\right] \\ \left[-\dfrac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}\right] & 0 \end{bmatrix}, \tag{7.2.11}$$

where A is the skew-symmetrization of $\partial^2 L/(\partial \dot{q}^i\, \partial q^j)$. From these expressions, it follows that Ω_L *is (weakly) nondegenerate if and only if* the quadratic form $\mathbf{D}_2\mathbf{D}_2 L(u,e)$ *is (weakly) nondegenerate*. In this case, we say that L is a ***regular*** or ***nondegenerate*** Lagrangian. The implicit function theorem shows that *the fiber derivative is locally invertible if and only if L is regular.*

Exercises

◇ **7.2-1.** Let

$$L(q^1, q^2, q^3, \dot q^1, \dot q^2, \dot q^3) = \frac{m}{2}\left(\left(\dot q^1\right)^2 + \left(\dot q^2\right)^2 + \left(\dot q^3\right)^2\right) + q^1\dot q^1 + q^2\dot q^2 + q^3\dot q^3.$$

Calculate Θ_L, Ω_L, and the corresponding Hamiltonian.

◇ **7.2-2.** For $v \in T_qQ$, define its ***vertical lift*** $v^l \in T_v(TQ)$ to be the tangent vector to the curve $v + tv$ at $t = 0$. Show that Θ_L may be defined by

$$w \mathbin{\lrcorner} \Theta_L = v^l \mathbin{\lrcorner} \mathbf{d}L,$$

where $w \in T_v(TQ)$ satisfies $T\tau_Q \cdot w = v$, and where $w \mathbin{\lrcorner} \Theta_L = \mathbf{i}_w\Theta_L$ is the interior product. Also, show that the energy is

$$E(v) = v^l \mathbin{\lrcorner} \mathbf{d}L - L(v).$$

◇ **7.2-3** (Abstract Legendre Transform)**.** Let V be a vector bundle over a manifold S and let $L : V \to \mathbb{R}$. For $v \in V$, let

$$w = \frac{\partial L}{\partial v} \in v^*$$

denote the fiber derivative. Assume that the map $v \mapsto w$ is a local diffeomorphism and let $H : V^* \to \mathbb{R}$ be defined by

$$H(w) = \langle w, v\rangle - L(v).$$

Show that

$$v = \frac{\partial H}{\partial w}.$$

7.3 Euler–Lagrange Equations

Hyperregular Lagrangians. Given a Lagrangian L, the ***action*** of L is the map $A : TQ \to \mathbb{R}$ that is defined by $A(v) = \mathbb{F}L(v) \cdot v$, and as we defined above, the ***energy*** of L is $E = A - L$. In charts,

$$A(u, e) = \mathbf{D}_2L(u, e) \cdot e, \tag{7.3.1}$$
$$E(u, e) = \mathbf{D}_2L(u, e) \cdot e - L(u, e), \tag{7.3.2}$$

and in finite dimensions, (7.3.1) and (7.3.2) read

$$A(q^i, \dot q^i) = \dot q^i \frac{\partial L}{\partial \dot q^i} = p_i \dot q^i, \tag{7.3.3}$$
$$E(q^i, \dot q^i) = \dot q^i \frac{\partial L}{\partial \dot q^i} - L(q^i, \dot q^i) = p_i \dot q^i - L(q^i, \dot q^i). \tag{7.3.4}$$

If L is a Lagrangian such that $\mathbb{F}L : TQ \to T^*Q$ is a diffeomorphism, we say that L is a ***hyperregular*** Lagrangian. In this case, set $H = E \circ (\mathbb{F}L)^{-1}$. Then X_H and X_E are $\mathbb{F}L$-related, since $\mathbb{F}L$ is, by construction, symplectic. Thus, hyperregular Lagrangians on TQ induce Hamiltonian systems on T^*Q. Conversely, one can show that hyperregular Hamiltonians on T^*Q come from Lagrangians on TQ (see §7.4 for definitions and details).

Lagrangian Vector Fields. More generally, a vector field Z on TQ is called a ***Lagrangian vector field*** or a ***Lagrangian system*** for L if the ***Lagrangian condition***

$$\Omega_L(v)(Z(v), w) = \mathbf{d}E(v) \cdot w \tag{7.3.5}$$

holds for all $v \in T_qQ$ and $w \in T_v(TQ)$. If L is regular, so that Ω_L is a (weak) symplectic form, then there would exist at most one such Z, which would be the Hamiltonian vector field of E with respect to the (weak) symplectic form Ω_L. In this case we know that E is conserved on the flow of Z. In fact, the same result holds, even if L is degenerate:

Proposition 7.3.1. *Let Z be a Lagrangian vector field for L and let $v(t) \in TQ$ be an integral curve of Z. Then $E(v(t))$ is constant in t.*

Proof. By the chain rule,

$$\begin{aligned} \frac{d}{dt}E(v(t)) &= \mathbf{d}E(v(t)) \cdot \dot{v}(t) = \mathbf{d}E(v(t)) \cdot Z(v(t)) \\ &= \Omega_L(v(t))(Z(v(t))), Z(v(t)) = 0 \end{aligned} \tag{7.3.6}$$

by skew-symmetry of Ω_L . ■

We usually assume that Ω_L is nondegenerate, but the degenerate case comes up in the Dirac theory of constraints (see Dirac [1950, 1964], Kunzle [1969], Hanson, Regge, and Teitelboim [1976], Gotay, Nester, and Hinds [1979], references therein, and §8.5).

Second-Order Equations. The vector field Z often has a special property, namely, that Z is a second-order equation.

Definition 7.3.2. *A vector field V on TQ is called a* ***second-order equation*** *if $T\tau_Q \circ V =$ identity, where $\tau_Q : TQ \to Q$ is the canonical projection. If $c(t)$ is an integral curve of V, then $(\tau_Q \circ c)(t)$ is called the* ***base integral curve*** *of $c(t)$.*

It is easy to see that the condition for V being second-order is equivalent to the following: For any chart $U \times E$ on TQ, we can write $V(u, e) = ((u, e), (e, V_2(u, e)))$, for some map $V_2 : U \times E \to E$. Thus, the dynamics are determined by $\dot{u} = e$, and $\dot{e} = V_2(u, e)$; that is, $\ddot{u} = V_2(u, \dot{u})$, a second-order equation in the standard sense. This local computation also shows that the base integral curve uniquely determines an integral curve of V through a given initial condition in TQ.

The Euler–Lagrange Equations. From the point of view of Lagrangian vector fields, the main result concerning the Euler–Lagrange equations is the following.

Theorem 7.3.3. *Let Z be a Lagrangian system for L and suppose Z is a second-order equation. Then in a chart $U \times E$, an integral curve $(u(t), v(t)) \in U \times E$ of Z satisfies the **Euler–Lagrange equations**; that is,*

$$\frac{du(t)}{dt} = v(t),$$
$$\frac{d}{dt}\mathbf{D}_2 L(u(t), v(t)) \cdot w = \mathbf{D}_1 L(u(t), v(t)) \cdot w \tag{7.3.7}$$

for all $w \in E$. In finite dimensions, the Euler–Lagrange equations take the form

$$\frac{dq^i}{dt} = \dot{q}^i,$$
$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) = \frac{\partial L}{\partial q^i}, \quad i = 1, \ldots, n. \tag{7.3.8}$$

If L is regular, that is, Ω_L is (weakly) nondegenerate, then Z is automatically second-order, and if it is strongly nondegenerate, then

$$\frac{d^2u}{dt^2} = \frac{dv}{dt} = [\mathbf{D}_2\mathbf{D}_2 L(u,v)]^{-1}(\mathbf{D}_1 L(u,v) - \mathbf{D}_1\mathbf{D}_2 L(u,v)\cdot v), \tag{7.3.9}$$

or in finite dimensions,

$$\ddot{q}^j = G^{ij}\left(\frac{\partial L}{\partial q^i} - \frac{\partial^2 L}{\partial q^j \partial \dot{q}^i}\dot{q}^j\right), \quad i, j = 1, \ldots, n, \tag{7.3.10}$$

where $[G^{ij}]$ is the inverse of the matrix $(\partial^2 L/\partial \dot{q}^i \partial \dot{q}^j)$. Thus $u(t)$ and $q^i(t)$ are base integral curves of the Lagrangian vector field Z if and only if they satisfy the Euler–Lagrange equations.

Proof. From the definition of the energy E we have the local expression

$$\begin{aligned}\mathbf{D}E(u,e)\cdot(e_1,e_2) = \mathbf{D}_1(\mathbf{D}_2 L(u,e)\cdot e)\cdot e_1 + \mathbf{D}_2(\mathbf{D}_2 L(u,e)\cdot e)\cdot e_2 \\ - \mathbf{D}_1 L(u,e)\cdot e_1\end{aligned} \tag{7.3.11}$$

(the term $\mathbf{D}_2 L(u,e)\cdot e_2$ has canceled). Locally, we may write

$$Z(u,e) = (u, e, Y_1(u,e), Y_2(u,e)).$$

Using formula (7.2.8) for Ω_L, the condition (7.3.5) on Z may be written

$$\begin{aligned}&\mathbf{D}_1\mathbf{D}_2 L(u,e)\cdot Y_1(u,e))\cdot e_1 - \mathbf{D}_1(\mathbf{D}_2 L(u,e)\cdot e_1)\cdot Y_1(u,e)\\ &\quad + \mathbf{D}_2\mathbf{D}_2 L(u,e)\cdot Y_1(u,e)\cdot e_2 - \mathbf{D}_2\mathbf{D}_2 L(u,e)\cdot e_1\cdot Y_2(u,e)\\ &\qquad = \mathbf{D}_1(\mathbf{D}_2 L(u,e)\cdot e)\cdot e_1 - \mathbf{D}_1 L(u,e)\cdot e_1 + \mathbf{D}_2\mathbf{D}_2 L(u,e)\cdot e\cdot e_2.\end{aligned} \tag{7.3.12}$$

Thus, if Ω_L is a weak symplectic form, then $\mathbf{D}_2\mathbf{D}_2L(u,e)$ is weakly nondegenerate, so setting $e_1 = 0$ we get $Y_1(u,e) = e$; that is, Z is a second-order equation. In any case, if we assume that Z is second-order, condition (7.3.12) becomes

$$\mathbf{D}_1L(u,e)\cdot e_1 = \mathbf{D}_1(\mathbf{D}_2L(u,e)\cdot e_1)\cdot e + \mathbf{D}_2\mathbf{D}_2L(u,e)\cdot e_1\cdot Y_2(u,e) \tag{7.3.13}$$

for all $e_1 \in E$. If $(u(t), v(t))$ is an integral curve of Z, then (using dots to denote time differentiation) $\dot{u} = v$ and $\ddot{u} = Y_2(u,v)$, so (7.3.13) becomes

$$\begin{aligned}\mathbf{D}_1L(u,\dot{u})\cdot e_1 &= \mathbf{D}_1(\mathbf{D}_2L(u,\dot{u})\cdot e_1)\cdot\dot{u} + \mathbf{D}_2\mathbf{D}_2L(u,\dot{u})\cdot e_1\cdot\ddot{u}\\ &= \frac{d}{dt}\mathbf{D}_2L(u,\dot{u})\cdot e_1\end{aligned} \tag{7.3.14}$$

by the chain rule.

The last statement follows by using the chain rule on the left-hand side of Lagrange's equation and using nondegeneracy of L to solve for $\dot{v}$, that is, $\ddot{q}^j$. ■

Exercises

◇ **7.3-1.** Give an explicit example of a degenerate Lagrangian L that has a second-order Lagrangian system Z.

◇ **7.3-2.** Check directly that the validity of the expression (7.3.8) is coordinate independent. In other words, verify directly that the form of the Euler–Lagrange equations does not depend on the local coordinates chosen to describe them.

7.4 Hyperregular Lagrangians and Hamiltonians

Above, we said that a smooth Lagrangian $L : TQ \to \mathbb{R}$ is ***hyperregular*** if $\mathbb{F}L : TQ \to T^*Q$ is a diffeomorphism. From (7.2.8) or (7.2.11) it follows that the symmetric bilinear form $\mathbf{D}_2\mathbf{D}_2L(u,e)$ is strongly nondegenerate. As before, let $\pi_Q : T^*Q \to Q$ and $\tau_Q : TQ \to Q$ denote the canonical projections.

Proposition 7.4.1. *Let L be a hyperregular Lagrangian on TQ and let $H = E\circ(\mathbb{F}L)^{-1} \in \mathcal{F}(T^*Q)$, where E is the energy of L. Then the Lagrangian vector field Z on TQ and the Hamiltonian vector field X_H on T^*Q are $\mathbb{F}L$-related, that is,*

$$(\mathbb{F}L)^*X_H = Z.$$

Furthermore, if $c(t)$ *is an integral curve of* Z *and* $d(t)$ *an integral curve of* X_H *with* $\mathbb{F}L(c(0)) = d(0)$*, then*

$$\mathbb{F}L(c(t)) = d(t) \quad \textit{and} \quad (\tau_Q \circ c)(t) = (\pi_Q \circ d)(t).$$

The curve $(\tau_Q \circ c)(t)$ *is called the* ***base integral curve*** *of* $c(t)$*, and similarly,* $(\pi_Q \circ d)(t)$ *is the* ***base integral curve*** *of* $d(t)$*.*

Proof. For $v \in TQ$ and $w \in T_v(TQ)$, we have

$$\begin{aligned}\Omega(\mathbb{F}L(v))(T_v\mathbb{F}L(Z(v)), T_v\mathbb{F}L(w)) &= ((\mathbb{F}L)^*\Omega)(v)(Z(v), w)\\ &= \Omega_L(v)(Z(v), w)\\ &= \mathbf{d}E(v) \cdot w\\ &= \mathbf{d}(H \circ \mathbb{F}L)(v) \cdot w\\ &= \mathbf{d}H(\mathbb{F}L(v)) \cdot T_v\mathbb{F}L(w)\\ &= \Omega(\mathbb{F}L(v))(X_H(\mathbb{F}L(v)), T_v\mathbb{F}L(w)),\end{aligned}$$

so that by weak nondegeneracy of Ω and the fact that $T_v\mathbb{F}L$ is an isomorphism, it follows that

$$T_v\mathbb{F}L(Z(v)) = X_H(\mathbb{F}L(v)).$$

Thus $T\mathbb{F}L \circ Z = X_H \circ \mathbb{F}L$, that is, $Z = (\mathbb{F}L)^*X_H$.

If φ_t denotes the flow of Z and ψ_t the flow of X_H, the relation $Z = (\mathbb{F}L)^*X_H$ is equivalent to $\mathbb{F}L \circ \varphi_t = \psi_t \circ \mathbb{F}L$. Thus, if $c(t) = \varphi_t(v)$, then

$$\mathbb{F}L(c(t)) = \psi_t(\mathbb{F}L(v))$$

is an integral curve of X_H that at $t = 0$ passes through $\mathbb{F}L(v) = \mathbb{F}L(c(0))$, whence $\psi_t(\mathbb{F}L(v)) = d(t)$ by uniqueness of integral curves of smooth vector fields. Finally, since $\tau_Q = \pi_Q \circ \mathbb{F}L$, we get

$$(\tau_Q \circ c)(t) = (\pi_Q \circ \mathbb{F}L \circ c)(t) = (\pi_Q \circ d)(t). \qquad \blacksquare$$

The Action. We claim that the action A of L is related to the Lagrangian vector field Z of L by

$$A(v) = \langle \Theta_L(v), Z(v) \rangle, \quad v \in TQ. \tag{7.4.1}$$

We prove this formula under the assumption that Z is a second-order equation, even if L is not regular. In fact,

$$\begin{aligned}\langle \Theta_L(v), Z(v) \rangle &= \langle ((\mathbb{F}L)^*\Theta)(v), Z(v) \rangle\\ &= \langle \Theta(\mathbb{F}L(v)),\ T_v\mathbb{F}L(Z(v)) \rangle\\ &= \langle \mathbb{F}L(v), T\pi_Q \cdot T_v\mathbb{F}L(Z(v)) \rangle\\ &= \langle \mathbb{F}L(v), T_v(\pi_Q \circ \mathbb{F}L)(Z(v)) \rangle\\ &= \langle \mathbb{F}L(v), T_v\tau_Q(Z(v)) \rangle = \langle \mathbb{F}L(v), v \rangle = A(v),\end{aligned}$$

by definition of a second-order equation and the definition of the action. If L is hyperregular and $H = E \circ (\mathbb{F}L)^{-1}$, then

$$A \circ (\mathbb{F}L)^{-1} = \langle \Theta, X_H \rangle . \tag{7.4.2}$$

Indeed, by (7.4.1), the properties of push-forward, and the previous proposition, we have

$$A \circ (\mathbb{F}L)^{-1} = (\mathbb{F}L)_* A = (\mathbb{F}L)_*(\langle \Theta_L, Z \rangle) = \langle (\mathbb{F}L)_* \Theta_L, (\mathbb{F}L)_* Z \rangle = \langle \Theta, X_H \rangle .$$

If $H : T^*Q \to \mathbb{R}$ is a smooth Hamiltonian, the function $G : T^*Q \to \mathbb{R}$ given by $G = \langle \Theta, X_H \rangle$ is called the ***action*** of H. Thus, (7.4.2) says that the push-forward of the action A of L equals the action G of $H = E \circ (\mathbb{F}L)^{-1}$.

Hyperregular Hamiltonians. A Hamiltonian H is called ***hyperregular*** if $\mathbb{F}H : T^*Q \to TQ$, defined by

$$\mathbb{F}H(\alpha) \cdot \beta = \left.\frac{d}{ds}\right|_{s=0} H(\alpha + s\beta), \tag{7.4.3}$$

where $\alpha, \beta \in T_q^*Q$, is a diffeomorphism; here we must assume that either the model space E of Q is reflexive, so that $T_q^{**}Q = T_qQ$ for all $q \in Q$, or what is more reasonable, that $\mathbb{F}H(\alpha)$ lies in $T_qQ \subset T_q^{**}Q$. As in the case of Lagrangians, hyperregularity of H implies the strong nondegeneracy of $\mathbf{D}_2\mathbf{D}_2 H(u, \alpha)$, and the curve $s \mapsto \alpha + s\beta$ appearing in (7.4.3) can be replaced by an arbitrary smooth curve $\alpha(s)$ in T_q^*Q such that

$$\alpha(0) = \alpha \quad \text{and} \quad \alpha'(0) = \beta.$$

Proposition 7.4.2. (i) *Let $H \in \mathcal{F}(T^*Q)$ be a hyperregular Hamiltonian and define*

$$E = H \circ (\mathbb{F}H)^{-1}, \quad A = G \circ (\mathbb{F}H)^{-1}, \quad \textit{and} \quad L = A - E \in \mathcal{F}(TQ).$$

Then L is a hyperregular Lagrangian and $\mathbb{F}L = \mathbb{F}H^{-1}$. Furthermore, A is the action of L, and E the energy of L.

(ii) *Let $L \in \mathcal{F}(TQ)$ be a hyperregular Lagrangian and define*

$$H = E \circ (\mathbb{F}L)^{-1}.$$

Then H is a hyperregular Hamiltonian and $\mathbb{F}H = (\mathbb{F}L)^{-1}$.

Proof. (i) Locally, $G(u, \alpha) = \langle \alpha, \mathbf{D}_2 H(u, \alpha) \rangle$, so that

$$A(u, \mathbf{D}_2 H(u, \alpha)) = (A \circ \mathbb{F}H)(u, \alpha) = G(u, \alpha) = \langle \alpha, \mathbf{D}_2 H(u, \alpha) \rangle ,$$

whence

$$(L \circ \mathbb{F}H)(u, \alpha) = L(u, \mathbf{D}_2 H(u, \alpha)) = \langle \alpha, \mathbf{D}_2 H(u, \alpha) \rangle - H(u, \alpha).$$

Let $e = \mathbf{D}_2(\mathbf{D}_2 H(u,\alpha)) \cdot \beta$, and let $e(s) = \mathbf{D}_2 H(u, \alpha + s\beta)$ be a curve that at $s = 0$ passes through $e(0) = \mathbf{D}_2 H(u,\alpha)$ and whose derivative at $s = 0$ equals $e'(0) = \mathbf{D}_2(\mathbf{D}_2 H(u,\alpha)) \cdot \beta = e$. Therefore,

$$\begin{aligned}
&\langle (\mathbb{F}L \circ \mathbb{F}H)(u,\alpha), e\rangle \\
&\qquad = \langle \mathbb{F}L(u, \mathbf{D}_2 H(u,\alpha)), e\rangle \\
&\qquad = \left.\frac{d}{dt}\right|_{s=0} L(u, e(s)) = \left.\frac{d}{dt}\right|_{s=0} L(u, \mathbf{D}_2 H(u, \alpha + s\beta)) \\
&\qquad = \left.\frac{d}{dt}\right|_{s=0} [\langle \alpha + s\beta, \mathbf{D}_2 H(u, \alpha + s\beta)\rangle - H(u, \alpha + s\beta)] \\
&\qquad = \langle \alpha, \mathbf{D}_2(\mathbf{D}_2 H(u,\alpha)) \cdot \beta\rangle = \langle \alpha, e\rangle .
\end{aligned}$$

Since $\mathbf{D}_2\mathbf{D}_2 H(u,\alpha)$ is strongly nondegenerate, this implies that $e \in E$ is arbitrary and hence $\mathbb{F}L \circ \mathbb{F}H =$ identity. Since $\mathbb{F}H$ is a diffeomorphism, this says that $\mathbb{F}L = (\mathbb{F}H)^{-1}$ and hence that L is hyperregular.

To see that A is the action of L, note that since $\mathbb{F}H^{-1} = \mathbb{F}L$, we have by definition of G,

$$A = G \circ (\mathbb{F}H)^{-1} = \langle \Theta, X_H\rangle \circ \mathbb{F}L,$$

which by (7.4.2) implies that A is the action of L. Therefore, $E = A - L$ is the energy of L.

(ii) Locally, since we define $H = E \circ (\mathbb{F}L)^{-1}$, we have

$$\begin{aligned}
(H \circ \mathbb{F}L)(u,e) &= H(u, \mathbf{D}_2 L(u,e)) \\
&= A(u,e) - L(u,e) \\
&= \mathbf{D}_2 L(u,e) \cdot e - L(u,e)
\end{aligned}$$

and proceed as before. Let

$$\alpha = \mathbf{D}_2(\mathbf{D}_2 L(u,e)) \cdot f,$$

where $f \in E$ and $\alpha(s) = \mathbf{D}_2 L(u, e + sf)$; then

$$\alpha(0) = \mathbf{D}_2 L(u,e) \quad \text{and} \quad \alpha'(0) = \alpha,$$

so that

$$\begin{aligned}
\langle \alpha, (\mathbb{F}H \circ \mathbb{F}L)(u,e)\rangle &= \langle \alpha, \mathbb{F}H(u, \mathbf{D}_2 L(u,e))\rangle \\
&= \left.\frac{d}{ds}\right|_{s=0} H(u, \alpha(s)) \\
&= \left.\frac{d}{ds}\right|_{s=0} H(u, \mathbf{D}_2 L(u, e + sf)) \\
&= \left.\frac{d}{ds}\right|_{s=0} [\langle \mathbf{D}_2 L(u, e + sf), e + sf\rangle - L(u, e + sf)] \\
&= \langle \mathbf{D}_2(\mathbf{D}_2 L(u,e)) \cdot f, e\rangle = \langle \alpha, e\rangle ,
\end{aligned}$$

which shows, by strong nondegeneracy of $\mathbf{D}_2\mathbf{D}_2L$, that $\mathbb{F}H \circ \mathbb{F}L = \text{identity}$. Since $\mathbb{F}L$ is a diffeomorphism, it follows that $\mathbb{F}H = (\mathbb{F}L)^{-1}$ and H is hyperregular. ■

The main result is summarized in the following.

Theorem 7.4.3. *Hyperregular Lagrangians $L \in \mathcal{F}(TQ)$ and hyperregular Hamiltonians $H \in \mathcal{F}(T^*Q)$ correspond in a bijective manner by the preceding constructions. The following diagram commutes:*

$$
\begin{tikzcd}
T(T^*Q) \arrow[rr, shift left, "T\mathbb{F}H"] & & T(TQ) \arrow[ll, shift left, "T\mathbb{F}L"] & \\
& \mathbb{R} & & \\
T^*Q \arrow[uu, "X_H"] \arrow[ur, "G"] \arrow[rr, shift right, "\mathbb{F}H"'] \arrow[dr, "H"'] & & TQ \arrow[uu, "X_E"'] \arrow[ul, "A"'] \arrow[ll, shift right, "\mathbb{F}L"'] \arrow[r, "L"] \arrow[dl, "E"] & \mathbb{R} \\
& \mathbb{R} & &
\end{tikzcd}
$$

Proof. Let L be a hyperregular Lagrangian and let H be the associated hyperregular Hamiltonian, that is,

$$H = E \circ (\mathbb{F}L)^{-1} = (A - L) \circ (\mathbb{F}L)^{-1} = G - L \circ \mathbb{F}H$$

by Propositions 7.4.1 and 7.4.2. From H we construct a Lagrangian L' by

$$
\begin{aligned}
L' &= G \circ (\mathbb{F}H)^{-1} - H \circ (\mathbb{F}H)^{-1} \\
&= G \circ (\mathbb{F}H)^{-1} - (G - L \circ \mathbb{F}H) \circ (\mathbb{F}H)^{-1} = L.
\end{aligned}
$$

Conversely, if H is a given hyperregular Hamiltonian, then the associated Lagrangian L is hyperregular and is given by

$$L = G \circ (\mathbb{F}H)^{-1} - H \circ (\mathbb{F}H)^{-1} = A - H \circ \mathbb{F}L.$$

Thus, the corresponding hyperregular Hamiltonian induced by L is

$$
\begin{aligned}
H' &= E \circ (\mathbb{F}L)^{-1} = (A - L) \circ (\mathbb{F}L)^{-1} \\
&= A \circ (\mathbb{F}L)^{-1} - (A - H \circ \mathbb{F}L) \circ (\mathbb{F}L)^{-1} = H.
\end{aligned}
$$

The commutativity of the two diagrams is now a direct consequence of the above and Propositions 7.4.1 and 7.4.2. ■

Neighborhood Theorem for Regular Lagrangians. We now prove an important theorem for regular Lagrangians that concerns the structure of solutions near a given one.

Definition 7.4.4. *Let $\overline{q}(t)$ be a given solution of the Euler–Lagrange equations, $\overline{t}_1 \leq t \leq \overline{t}_2$. Let $\overline{q}_1 = \overline{q}\left(\overline{t}_1\right)$ and $\overline{q}_2 = \overline{q}\left(\overline{t}_2\right)$. We say that $\overline{q}(t)$ is a **nonconjugate solution** if there is a neighborhood $\mathcal{U}$ of the curve $\overline{q}(t)$ and neighborhoods $\mathcal{U}_1 \subset \mathcal{U}$ of $\overline{q}_1$ and $\mathcal{U}_2 \subset \mathcal{U}$ of $\overline{q}_2$ such that for all $q_1 \in \mathcal{U}_1$ and $q_2 \in \mathcal{U}_2$ and t_1 close to $\overline{t}_1$, t_2 close to $\overline{t}_2$, there exists a unique solution $q(t)$, $t_1 \leq t \leq t_2$, of the Euler–Lagrange equations satisfying the following conditions: $q\left(t_1\right) = q_1$, $q\left(t_2\right) = q_2$, and $q(t) \in \mathcal{U}$. See Figure* 7.4.1.

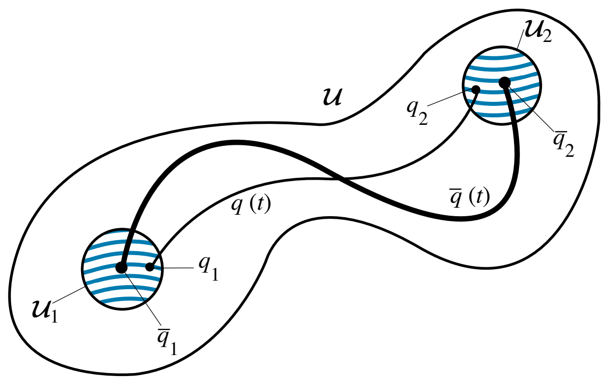

Figure 7.4.1. Neighborhood theorem

To determine conditions guaranteeing that a solution is nonconjugate, we shall use the following observation. Let $\overline{v}_1 = \dot{\overline{q}}\left(t_1\right)$ and $\overline{v}_2 = \dot{\overline{q}}\left(t_2\right)$. Let F_t be the flow of the Euler–Lagrange equations on TQ. By construction of $F_t(q, v)$, we have $F_{t_2}\left(\overline{q}_1, \overline{v}_1\right) = \left(\overline{q}_2, \overline{v}_2\right)$.

Next, we attempt to apply the implicit function theorem to the flow map. We want to solve

$$\left(\pi_Q \circ F_{t_2}\right)\left(q_1, v_1\right) = q_2$$

for v_1, where we regard q_1, t_1, t_2 as parameters. To do this, we form the linearization

$$w_2 := T_{v_1}\left(\pi_Q \circ F_{\overline{t}_2}\right)\left(\overline{q}_1, \overline{v}_1\right) \cdot w_1 .$$

We require that $w_1 \mapsto w_2$ be invertible. The right-hand side of this equation suggests forming the curve

$$w(t) := T_{v_1} \pi_Q F_t\left(\overline{q}_1, \overline{v}_1\right) \cdot w_1, \tag{7.4.4}$$

which is the solution of the linearized, or first variation, equation of the Euler–Lagrange equations satisfied by $F_t(\overline{q}_1, \overline{v}_1)$. Let us work out the equation satisfied by

$$w(t) := T_{v_1}\pi_Q F_t(\overline{q}_1, \overline{v}_1) \cdot w_1$$

in coordinates. Start with a solution $q(t)$ of the Euler–Lagrange equations

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

Given the curve of initial conditions $\varepsilon \mapsto (q_1, v_1 + \varepsilon w_1)$, we get corresponding solutions $(q_\varepsilon(t), \dot{q}_\varepsilon(t))$, whose derivative with respect to ε we denoted by $(u(t), \dot{u}(t))$. Differentiation of the Euler–Lagrange equations with respect to ε gives

$$\frac{d}{dt}\left(\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} \cdot \dot{u}^j + \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j} \cdot u^j\right) - \frac{\partial^2 L}{\partial q^i \partial q^j} \cdot u^j - \frac{\partial^2 L}{\partial q^i \partial \dot{q}^j} \cdot u^j = 0, \quad (7.4.5)$$

which is a second-order equation for u^j. This equation evaluated along $\overline{q}(t)$ is called the ***Jacobi equation*** along $\overline{q}(t)$. This equation, taken from $\overline{q}(\overline{t}_1)$ to $\overline{q}(\overline{t}_2)$ with initial conditions

$$u(t_1) = 0 \quad \text{and} \quad \dot{u}(t_1) = w_1,$$

defines the desired linear map $w_1 \mapsto w_2$; that is, $w_2 = \dot{u}(\overline{t}_2)$.

Theorem 7.4.5. *Assume that L is a regular Lagrangian. If the linear map $w_1 \mapsto w_2$ is an isomorphism, then $\overline{q}(t)$ is nonconjugate.*

Proof. This follows directly from the implicit function theorem. Under the hypothesis that $w_1 \mapsto w_2$ is invertible, there are neighborhoods $\mathcal{U}_1$ of $\overline{q}_1$, $\mathcal{U}_2$ of $\overline{q}_2$ and neighborhoods of $\overline{t}_1$ and $\overline{t}_2$ as well as a smooth function $v_1 = v_1(t_1, t_2, q_1, q_2)$ defined on the product of these four neighborhoods such that

$$(\pi_Q \circ F_{t_2})\,(q_1, v_1(t_1, t_2, q_1, q_2)) = q_2 \quad (7.4.6)$$

is an identity. Then

$$q(t) := (\pi_Q \circ F_t)(q_1, v_1(t_1, t_2, q_1, q_2))$$

is a solution of the Euler–Lagrange equations with initial conditions

$$(q_1, v_1(t_1, t_2, q_1, q_2)) \,\text{at}\, t = t_1.$$

Moreover, $q(t_2) = q_2$ by (7.4.6). The fact that v_1 is close to $\overline{v}_1$ means that the geodesic found lies in a neighborhood of the curve $\overline{q}(t)$; this produces the neighborhood $\mathcal{U}$. ■

If q_1 and q_2 are close and if t_2 is not much different from t_1, then by continuity, $\dot{u}(t)$ is approximately constant over $[t_1, t_2]$, so that

$$w_2 = \dot{u}(t_2) = (t_2 - t_1)\dot{u}(t_1) + O(t_2 - t_1)^2 = (t_2 - t_1)w_1 + O(t_2 - t_1)^2.$$

Thus, in these circumstances, the map $w_1 \mapsto w_2$ is invertible. Therefore, we get the following corollary.

Corollary 7.4.6. *Let $L : TQ \times \mathbb{R} \to \mathbb{R}$ be a given C^2 regular Lagrangian and let $v_q \in TQ$ and $t_1 \in \mathbb{R}$. Then the solution of the Euler–Lagrange equations with initial condition v_q at $t = t_1$ is nonconjugate for a sufficiently small time interval $[t_1, t_2]$.*

The term "nonconjugate" comes from the study of geodesics, which are considered in the next section.

Exercises

◇ **7.4-1.** Write down the Lagrangian and the equations of motion for a spherical pendulum with S^2 as configuration space. Convert the equations to Hamiltonian form using the Legendre transformation. Find the conservation law corresponding to angular momentum about the axis of gravity by "bare hands" methods.

◇ **7.4-2.** Let $L(q, \dot{q}) = \frac{1}{2}m(q)\dot{q}^2 - V(q)$ on $T\mathbb{R}$, where $m(q) > 0$ and $V(q)$ are smooth. Show that *any* two points $q_1, q_2 \in \mathbb{R}$ can be joined by a solution of the Euler–Lagrange equations. (Hint: Consider the energy equation.)

7.5 Geodesics

Let Q be a weak pseudo-Riemannian manifold whose metric evaluated at $q \in Q$ is denoted interchangeably by $\langle \cdot \, , \cdot \rangle$ or $g(q)$ or g_q. Consider on TQ the Lagrangian given by the kinetic energy of the metric, that is,

$$L(v) = \tfrac{1}{2} \langle v, v \rangle_q \,, \tag{7.5.1}$$

or in finite dimensions

$$L(v) = \tfrac{1}{2} g_{ij} v^i v^j. \tag{7.5.2}$$

The fiber derivative of L is given for $v, w \in T_qQ$ by

$$\mathbb{F}L(v) \cdot w = \langle v, w \rangle \tag{7.5.3}$$

or in finite dimensions by

$$\mathbb{F}L(v) \cdot w = g_{ij} v^i w^j, \quad \text{i.e.,} \quad p_i = g_{ij} \dot{q}^j. \tag{7.5.4}$$

From this equation we see that in any chart U for Q,

$$\mathbf{D}_2\mathbf{D}_2L(q,v)\cdot(e_1,e_2) = \langle e_1, e_2\rangle_q\,,$$

where $\langle\,,\rangle_q$ denotes the inner product on E induced by the chart. Thus, L is automatically weakly nondegenerate. Note that the action is given by $A = 2L$, so $E = L$.

The Lagrangian vector field Z in this case is denoted by $S : TQ \to T^2Q$ and is called the ***Christoffel map*** or ***geodesic spray*** of the metric $\langle\,,\rangle_q$. Thus, S is a second-order equation and hence has a local expression of the form

$$S(q,v) = ((q,v),(v,\gamma(q,v))) \tag{7.5.5}$$

in a chart on Q. To determine the map $\gamma : U \times E \to E$ from Lagrange's equations, note that

$$\mathbf{D}_1L(q,v)\cdot w = \tfrac{1}{2}\mathbf{D}_q\langle v,v\rangle_q\cdot w \quad\text{and}\quad \mathbf{D}_2L(q,v)\cdot w = \langle v,w\rangle_q\,, \tag{7.5.6}$$

so that the Euler–Lagrange equations (7.3.7) are

$$\dot q = v, \tag{7.5.7}$$

$$\frac{d}{dt}(\langle v,w\rangle_q) = \tfrac{1}{2}\mathbf{D}_q\langle v,v\rangle_q\cdot w. \tag{7.5.8}$$

Keeping w fixed and expanding the left-hand side of (7.5.8) yields

$$\mathbf{D}_q\langle v,w\rangle_q\cdot\dot q + \langle\dot v,w\rangle_q\,. \tag{7.5.9}$$

Taking into account $\dot q = v$, we get

$$\langle\ddot q,w\rangle_q = \tfrac{1}{2}\mathbf{D}_q\langle v,v\rangle_q\cdot w - \mathbf{D}_q\langle v,w\rangle_q\cdot v. \tag{7.5.10}$$

Hence $\gamma : U \times E \to E$ is defined by the equality

$$\langle\gamma(q,v),w\rangle_q = \tfrac{1}{2}\mathbf{D}_q\langle v,v\rangle_q\cdot w - \mathbf{D}_q\langle v,w\rangle_q\cdot v; \tag{7.5.11}$$

note that $\gamma(q,v)$ is a quadratic form in v. If Q is finite-dimensional, we define the ***Christoffel symbols*** Γ^i_{jk} by putting

$$\gamma^i(q,v) = -\Gamma^i_{jk}(q)v^jv^k \tag{7.5.12}$$

and demanding that $\Gamma^i_{jk} = \Gamma^i_{kj}$. With this notation, the relation (7.5.11) is equivalent to

$$-g_{il}\Gamma^i_{jk}v^jv^kw^l = \frac{1}{2}\frac{\partial g_{jk}}{\partial q^l}v^jv^kw^l - \frac{\partial g_{jl}}{\partial q^k}v^jw^lv^k. \tag{7.5.13}$$

Taking into account the symmetry of Γ^i_{jk}, this gives

$$\Gamma^h_{jk} = \frac{1}{2} g^{hl} \left(\frac{\partial g_{jl}}{\partial q^k} + \frac{\partial g_{kl}}{\partial q^j} - \frac{\partial g_{jk}}{\partial q^l} \right). \tag{7.5.14}$$

In infinite dimensions, since the metric $\langle \, , \rangle$ is only weakly nondegenerate, (7.5.11) guarantees the uniqueness of γ but not its existence. It exists whenever the Lagrangian vector field S exists.

The integral curves of S projected to Q are called ***geodesics*** of the metric g. By (7.5.5), their basic governing equation has the local expression

$$\ddot{q} = \gamma(q, \dot{q}), \tag{7.5.15}$$

which in finite dimensions reads

$$\ddot{q}^i + \Gamma^i_{jk} \dot{q}^j \dot{q}^k = 0, \tag{7.5.16}$$

where $i, j, k = 1, \ldots, n$ and, as usual, there is a sum on j and k. Note that the definition of γ makes sense in both the finite- and infinite-dimensional cases, whereas the Christoffel symbols Γ^i_{jk} are literally defined only for finite-dimensional manifolds. Working intrinsically with g provides a way to deal with geodesics of weak Riemannian (and pseudo-Riemannian) metrics on infinite-dimensional manifolds.

Taking the Lagrangian approach as basic, we see that the Γ^i_{jk} live as geometric objects in $T(TQ)$. This is because they encode the principal part of the Lagrangian vector field Z. If one writes down the transformation properties of Z on $T(TQ)$ in natural charts, the classical transformation rule for the Γ^i_{jk} results:

$$\overline{\Gamma}^k_{ij} = \frac{\partial q^p}{\partial \overline{q}^i} \frac{\partial q^m}{\partial \overline{q}^j} \Gamma^r_{pm} \frac{\partial \overline{q}^k}{\partial q^r} + \frac{\partial \overline{q}^k}{\partial q^l} \frac{\partial^2 q^l}{\partial \overline{q}^i \, \partial \overline{q}^j}, \tag{7.5.17}$$

where $(q^1, \ldots, q^n), (\overline{q}^1, \ldots, \overline{q}^n)$ are two different coordinate systems on an open set of Q. We leave this calculation to the reader.

The Lagrangian approach leads naturally to invariant manifolds for the geodesic flow. For example, for each real $e > 0$, let

$$\Sigma_e = \{ \, v \in TQ \mid \|v\| = e \, \}$$

be the ***pseudo-sphere bundle*** of radius $\sqrt{e}$ in TQ. Then Σ_e is a smooth submanifold of TQ invariant under the geodesic flow. Indeed, if we show that Σ_e is a smooth submanifold, its invariance under the geodesic flow, that is, under the flow of Z, follows by conservation of energy. To show that Σ_e is a smooth submanifold we prove that e is a regular value of L for $e > 0$. This is done locally by (7.5.6):

$$\begin{aligned} \mathbf{D}L(u, v) \cdot (w_1, w_2) &= \mathbf{D}_1 L(u, v) \cdot w_1 + \mathbf{D}_2 L(u, v) \cdot w_2 \\ &= \tfrac{1}{2} \mathbf{D}_u \langle v, v \rangle_u \cdot w_1 + \langle v, w_2 \rangle_u \\ &= \langle v, w_2 \rangle_u \,, \end{aligned} \tag{7.5.18}$$

since $\langle v, v\rangle = 2e =$ constant. By weak nondegeneracy of the pseudo-metric $\langle\,,\rangle$, this shows that $\mathbf{D}L(u, v) : E \times E \to \mathbb{R}$ is a surjective linear map, that is, e is a regular value of L.

Convex Neighborhoods and Conjugate Points. We proved in the last section that short arcs of solutions of the Euler–Lagrange equations are nonconjugate. In the special case of geodesics one can do somewhat better by exploiting the fact, evident from the quadratic nature of (7.5.16), that if $q(t)$ is a solution and $\alpha > 0$, then so is $q(\alpha t)$, so one can "rescale" solutions simply by changing the size of the initial velocity. One finds that locally there are *convex* neighborhoods, that is, neighborhoods U such that for any $q_1, q_2 \in U$ there is a unique geodesic (up to a scaling) joining q_1, q_2 and lying in U. In Riemannian geometry there is another important result, the ***Hopf–Rinow theorem***, stating that any two points (in the same connected component) can be joined by some geodesic.

As one follows a geodesic from a given point, there is a first point after which *nearby* geodesics fail to be unique. These are ***conjugate points***. They are the zeros of the Jacobi equation discussed earlier. For example, on a great circle on a sphere, pairs of antipodal points are conjugate.

In certain circumstances one can "reduce" the Euler–Lagrange problem to one of geodesics: See the discussion of the Jacobi metric in §7.7.

Covariant derivatives. We now reconcile the above approach to geodesics via Lagrangian systems to a common approach in differential geometry. Define the ***covariant derivative***

$$\nabla : \mathfrak{X}(Q) \times \mathfrak{X}(Q) \to \mathfrak{X}(Q), \quad (X, Y) \mapsto \nabla_X Y$$

locally by

$$(\nabla_X Y)(u) = -\gamma(u)(X(u), Y(u)) + \mathbf{D}Y(u) \cdot X(u), \tag{7.5.19}$$

where X, Y are the local representatives of X and Y, and $\gamma(u) : E \times E \to E$ denotes the symmetric bilinear form defined by the polarization of $\gamma(u, v)$, which is a quadratic form in v. In local coordinates, the preceding equation becomes

$$\nabla_X Y = X^j Y^k \Gamma^i_{jk} \frac{\partial}{\partial q^i} + X^j \frac{\partial Y^k}{\partial q^j} \frac{\partial}{\partial q^k}. \tag{7.5.20}$$

It is straightforward to check that this definition is chart independent and that ∇ satisfies the following conditions:

(i) ∇ is $\mathbb{R}$-bilinear;

(ii) for $f : Q \to \mathbb{R}$,

$$\nabla_{fX} Y = f \nabla_X Y \quad \text{and} \quad \nabla_X fY = f \nabla_X Y + X[f]Y;$$

and

(iii) for vector fields X and Y,

$$\begin{aligned}(\nabla_X Y - \nabla_Y X)(u) &= \mathbf{D}Y(u)\cdot X(u) - \mathbf{D}X(u)\cdot Y(u)\\ &= [X,Y](u).\end{aligned} \tag{7.5.21}$$

In fact, these three properties characterize covariant derivative operators. The particular covariant derivative determined by (7.5.14) is called the ***Levi-Civita covariant derivative***. If $c(t)$ is a curve in Q and $X \in \mathfrak{X}(Q)$, the ***covariant derivative of*** X ***along*** c is defined by

$$\frac{DX}{Dt} = \nabla_u X, \tag{7.5.22}$$

where u is a vector field coinciding with $\dot{c}(t)$ at $c(t)$. This is possible, since by (7.5.19) or (7.5.20), $\nabla_X Y$ depends only on the point values of X. Explicitly, in a local chart, we have

$$\frac{DX}{Dt}(c(t)) = -\gamma_{c(t)}(u(c(t)), X(c(t))) + \frac{d}{dt}X(c(t)), \tag{7.5.23}$$

which shows that DX/Dt depends only on $\dot{c}(t)$ and not on how $\dot{c}(t)$ is extended to a vector field. In finite dimensions,

$$\left(\frac{DX}{Dt}\right)^i = \Gamma^i_{jk}(c(t))\dot{c}^j(t)X^k(c(t)) + \frac{d}{dt}X^i(c(t)). \tag{7.5.24}$$

The vector field X is called ***autoparallel*** or ***parallel transported*** along c if $DX/Dt = 0$. Thus $\dot{c}$ is autoparallel along c if and only if

$$\ddot{c}(t) - \gamma(t)(\dot{c}(t), \dot{c}(t)) = 0,$$

that is, $c(t)$ is a geodesic. In finite dimensions, this reads

$$\ddot{c}^i + \Gamma^i_{jk}\dot{c}^j\dot{c}^k = 0.$$

Exercises

◇ **7.5-1.** Consider the Lagrangian

$$L_\epsilon(x,y,z,\dot{x},\dot{y},\dot{z}) = \frac{1}{2}\left(\dot{x}^2+\dot{y}^2+\dot{z}^2\right) - \frac{1}{2\epsilon}\left[1-\left(x^2+y^2+z^2\right)\right]^2$$

for a particle in $\mathbb{R}^3$. Let $\gamma_\epsilon(t)$ be the curve in $\mathbb{R}^3$ obtained by solving the Euler–Lagrange equations for L_ϵ with the initial conditions $\mathbf{x}_0, \mathbf{v}_0 = \dot{\gamma}_\epsilon(0)$. Show that

$$\lim_{\epsilon\to 0}\gamma_\epsilon(t)$$

is a great circle on the two-sphere S^2, provided that $\mathbf{x}_0$ has length one and that $\mathbf{x}_0 \cdot \mathbf{v}_0 = 0$.

◇ **7.5-2.** Write out the geodesic equations in terms of q^i and p_i and check directly that Hamilton's equations are satisfied.

7.6 The Kaluza–Klein Approach to Charged Particles

In §6.7 we studied the motion of a charged particle in a magnetic field as a Hamiltonian system. Here we show that this description is the reduction of a larger and, in some sense, simpler system called the ***Kaluza–Klein system***.[1]

Physically, we are motivated as follows: Since charge is a basic conserved quantity, we would like to introduce a new cyclic variable whose conjugate momentum is the charge.[2] For a charged particle, the resultant system is in fact geodesic motion!

Recall from §6.7 that if $\mathbf{B} = \nabla \times \mathbf{A}$ is a given magnetic field on $\mathbb{R}^3$, then with respect to canonical variables $(\mathbf{q}, \mathbf{p})$, the Hamiltonian is

$$H(\mathbf{q}, \mathbf{p}) = \frac{1}{2m} \left\| \mathbf{p} - \frac{e}{c}\mathbf{A} \right\|^2 . \tag{7.6.1}$$

First we claim that we can obtain (7.6.1) via the Legendre transform if we choose

$$L(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2} m \left\| \dot{\mathbf{q}} \right\|^2 + \frac{e}{c}\mathbf{A} \cdot \dot{\mathbf{q}}. \tag{7.6.2}$$

Indeed, in this case,

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}} = m\dot{\mathbf{q}} + \frac{e}{c}\mathbf{A} \tag{7.6.3}$$

and

$$\begin{aligned} H(\mathbf{q}, \mathbf{p}) &= \mathbf{p} \cdot \dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}) \\ &= \left(m\dot{\mathbf{q}} + \frac{e}{c}\mathbf{A} \right) \cdot \dot{\mathbf{q}} - \frac{1}{2} m \left\| \dot{\mathbf{q}} \right\|^2 - \frac{e}{c}\mathbf{A} \cdot \dot{\mathbf{q}} \\ &= \frac{1}{2} m \left\| \dot{\mathbf{q}} \right\|^2 = \frac{1}{2m} \left\| \mathbf{p} - \frac{e}{c}\mathbf{A} \right\|^2 . \end{aligned} \tag{7.6.4}$$

Thus, the Euler–Lagrange equations for (7.6.2) reproduce the equations for a particle in a magnetic field.[3]

Let the configuration space be

$$Q_K = \mathbb{R}^3 \times S^1 \tag{7.6.5}$$

---

[1] After learning reduction theory (see Abraham and Marsden [1978] or Marsden [1992]), the reader can revisit this construction, but here all the constructions are done directly.

[2] This process is applicable to other situations as well; for example, in fluid dynamics one can profitably introduce a variable conjugate to the conserved mass density or entropy; see Marsden, Ratiu, and Weinstein [1984a, 1984b].

[3] If an electric field $E = -\nabla\varphi$ is also present, one simply subtracts $e\varphi$ from L, treating $e\varphi$ as a potential energy, as in the next section.

with variables $(\mathbf{q}, \theta)$. Define $A = \mathbf{A}^\flat$, a one-form on $\mathbb{R}^3$, and consider the one-form

$$\omega = A + \mathbf{d}\theta \tag{7.6.6}$$

on Q_K called the ***connection one-form***. Let the ***Kaluza–Klein Lagrangian*** be defined by

$$\begin{aligned} L_K(\mathbf{q}, \dot{\mathbf{q}}, \theta, \dot{\theta}) &= \frac{1}{2}m\|\dot{\mathbf{q}}\|^2 + \frac{1}{2}\left\|\left\langle \omega, (\mathbf{q}, \dot{\mathbf{q}}, \theta, \dot{\theta})\right\rangle\right\|^2 \\ &= \frac{1}{2}m\|\dot{\mathbf{q}}\|^2 + \frac{1}{2}(\mathbf{A}\cdot\dot{\mathbf{q}} + \dot{\theta})^2. \end{aligned} \tag{7.6.7}$$

The corresponding momenta are

$$\mathbf{p} = m\dot{\mathbf{q}} + (\mathbf{A}\cdot\dot{\mathbf{q}} + \dot{\theta})\mathbf{A} \tag{7.6.8}$$

and

$$p = \mathbf{A}\cdot\dot{\mathbf{q}} + \dot{\theta}. \tag{7.6.9}$$

Since L_K is quadratic and positive definite in $\dot{\mathbf{q}}$ and $\dot{\theta}$, the Euler–Lagrange equations are the geodesic equations on $\mathbb{R}^3 \times S^1$ for the metric for which L_K is the kinetic energy. Since p is constant in time, as can be seen from the Euler–Lagrange equation for $(\theta, \dot{\theta})$, we can define the ***charge*** e by setting

$$p = \frac{e}{c}; \tag{7.6.10}$$

then (7.6.8) coincides with (7.6.3). The corresponding Hamiltonian on T^*Q_K endowed with the canonical symplectic form is

$$H_K(\mathbf{q}, \mathbf{p}, \theta, p) = \frac{1}{2m}\|\mathbf{p} - p\mathbf{A}\|^2 + \frac{1}{2}p^2. \tag{7.6.11}$$

With (7.6.10), (7.6.11) differs from (7.6.1) by the constant $p^2/2$.

These constructions generalize to the case of a particle in a Yang–Mills field, where ω becomes the ***connection*** of a Yang–Mills field and its ***curvature*** measures the field strength that, for an electromagnetic field, reproduces the relation $\mathbf{B} = \nabla \times \mathbf{A}$. Also, the possibility of putting the interaction in the Hamiltonian, or via a momentum shift, into the symplectic structure, also generalizes. We refer to Wong [1970], Sternberg [1977], Weinstein [1978a], and Montgomery [1984] for details and further references. Finally, we remark that the relativistic context is the most natural in which to introduce the full electromagnetic field. In that setting the construction we have given for the magnetic field will include both electric and magnetic effects. Consult Misner, Thorne, and Wheeler [1973] for additional information.

Exercises

◇ **7.6-1.** The bob on a spherical pendulum has a charge e, mass m, and moves under the influence of a constant gravitational field with acceleration g, and a magnetic field $\mathbf{B}$. Write down the Lagrangian, the Euler–Lagrange equations, and the variational principle for this system. Transform the system to Hamiltonian form. Find a conserved quantity if the field $\mathbf{B}$ is symmetric about the axis of gravity.

7.7 Motion in a Potential Field

We now generalize geodesic motion to include potentials $V : Q \to \mathbb{R}$. Recall that the ***gradient*** of V is the vector field grad $V = \nabla V$ defined by the equality

$$\langle \operatorname{grad} V(q), v \rangle_q = \mathbf{d}V(q) \cdot v, \tag{7.7.1}$$

for all $v \in T_qQ$. In finite dimensions, this definition becomes

$$(\operatorname{grad} V)^i = g^{ij} \frac{\partial V}{\partial q^j}. \tag{7.7.2}$$

Define the (weakly nondegenerate) Lagrangian $L(v) = \frac{1}{2} \langle v, v \rangle_q - V(q)$. A computation similar to the one in §7.5 shows that the Euler–Lagrange equations are

$$\ddot{q} = \gamma(q, \dot{q}) - \operatorname{grad} V(q), \tag{7.7.3}$$

or in finite dimensions,

$$\ddot{q}^i + \Gamma^i_{jk} \dot{q}^j \dot{q}^k + g^{il} \frac{\partial V}{\partial q^l} = 0. \tag{7.7.4}$$

The action of L is given by

$$A(v) = \langle v, v \rangle_q , \tag{7.7.5}$$

so that the energy is

$$E(v) = A(v) - L(v) = \tfrac{1}{2} \langle v, v \rangle_q + V(q). \tag{7.7.6}$$

The equations (7.7.3) written as

$$\dot{q} = v, \quad \dot{v} = \gamma(q, v) - \operatorname{grad} V(q) \tag{7.7.7}$$

are thus Hamiltonian with Hamiltonian function E with respect to the symplectic form Ω_L.

Invariant Form. There are several ways to write equations (7.7.7) in invariant form. Perhaps the simplest is to use the language of covariant derivatives from the last section and to write

$$\frac{D\dot{c}}{Dt} = -\nabla V \tag{7.7.8}$$

or, what is perhaps better,

$$g^\flat \frac{D\dot{c}}{Dt} = -\mathbf{d}V, \tag{7.7.9}$$

where $g^\flat : TQ \to T^*Q$ is the map associated to the Riemannian metric. This last equation is the geometric way of writing $m\mathbf{a} = \mathbf{F}$.

Another method uses the following terminology:

Definition 7.7.1. *Let $v, w \in T_qQ$. The* ***vertical lift*** *of w with respect to v is defined by*

$$\operatorname{ver}(w, v) = \left.\frac{d}{dt}\right|_{t=0} (v + tw) \in T_v(TQ).$$

The ***horizontal part*** *of a vector $U \in T_v(TQ)$ is $T_v\tau_Q(U) \in T_qQ$. A vector field is called* ***vertical*** *if its horizontal part is zero.*

In charts, if $v = (u, e),\ w = (u, f)$, and $U = ((u, e), (e_1, e_2))$, this definition says that

$$\operatorname{ver}(w, v) = ((u, e), (0, f)) \quad \text{and} \quad T_v\tau_Q(U) = (u, e_1).$$

Thus, U is vertical iff $e_1 = 0$. Thus, *any vertical vector $U \in T_v(TQ)$ is the vertical lift of some vector w (which in a natural local chart is (u, e_2)) with respect to v.*

If S denotes the geodesic spray of the metric $\langle\,,\rangle$ on TQ, equations (7.7.7) say that the Lagrangian vector field Z defined by $L(v) = \frac{1}{2}\langle v, v\rangle_q - V(q)$, where $v \in T_qQ$, is given by

$$Z = S - \operatorname{ver}(\nabla V), \tag{7.7.10}$$

that is,

$$Z(v) = S(v) - \operatorname{ver}((\nabla V)(q), v). \tag{7.7.11}$$

Remarks. In general, there is *no* canonical way to take the *vertical part* of a vector $U \in T_v(TQ)$ without extra structure. Having such a structure is what one means by a ***connection***. In case Q is pseudo-Riemannian, such a projection can be constructed in the following manner. Suppose, in natural charts, that $U = ((u, e), (e_1, e_2))$. Define

$$U_{\text{ver}} = ((u, e),\ (0, \gamma(u)(e_1, e_2) + e_2))$$

where $\gamma(u)$ is the bilinear symmetric form associated to the quadratic form $\gamma(u, e)$ in e. ♦

We conclude with some miscellaneous remarks connecting motion in a potential field with geodesic motion. We confine ourselves to the finite-dimensional case for simplicity.

Definition 7.7.2. *Let $g = \langle\, , \rangle$ be a pseudo-Riemannian metric on Q and let $V : Q \to \mathbb{R}$ be bounded above. If $e > V(q)$ for all $q \in Q$, define the* ***Jacobi metric*** *g_e by $g_e = (e - V)g$, that is,*

$$g_e(v, w) = (e - V(q)) \langle v, w\rangle$$

for all $v, w \in T_qQ$.

Theorem 7.7.3. *Let Q be finite-dimensional. The base integral curves of the Lagrangian $L(v) = \frac{1}{2} \langle v, v\rangle - V(q)$ with energy e are the same as geodesics of the Jacobi metric with energy 1, up to a reparametrization.*

The proof is based on the following proposition of separate interest.

Proposition 7.7.4. *Let (P, Ω) be a (finite-dimensional) symplectic manifold, $H, K \in \mathcal{F}(P)$, and assume that $\Sigma = H^{-1}(h) = K^{-1}(k)$ for $h, k \in \mathbb{R}$ regular values of H and K, respectively. Then the integral curves of X_H and X_K on the invariant submanifold Σ of both X_H and X_K coincide up to a reparametrization.*

Proof. From $\Omega(X_H(z), v) = \mathbf{d}H(z) \cdot v$, we see that

$$X_H(z) \in (\ker \mathbf{d}H(z))^{\Omega} = (T_z\Sigma)^{\Omega},$$

the symplectic orthogonal complement of $T_z\Sigma$. Since

$$\dim P = \dim T_z\Sigma + \dim(T_z\Sigma)^{\Omega}$$

(see §2.3) and since $T_z\Sigma$ has codimension one, $(T_z\Sigma)^{\Omega}$ has dimension one. Thus, the nonzero vectors $X_H(z)$ and $X_K(z)$ are multiples of each other at every point $z \in \Sigma$, that is, there is a smooth nowhere-vanishing function $\lambda : \Sigma \to \mathbb{R}$ such that $X_H(z) = \lambda(z)X_K(z)$ for all $z \in \Sigma$. Let $c(t)$ be the integral curve of X_K with initial condition $c(0) = z_0 \in \Sigma$. The function

$$\varphi \mapsto \int_0^{\varphi} \frac{dt}{(\lambda \circ c)(t)}$$

is a smooth monotone function and therefore has an inverse $t \mapsto \varphi(t)$. If $d(t) = (c \circ \varphi)(t)$, then $d(0) = z_0$ and

$$\begin{aligned} d'(t) &= \varphi'(t)c'(\varphi(t)) = \frac{1}{t'(\varphi)} X_K(c(\varphi(t))) = (\lambda \circ c)(\varphi)X_K(d(t)) \\ &= \lambda(d(t))X_K(d(t)) = X_H(d(t)), \end{aligned}$$

that is, the integral curve of X_H through z_0 is obtained by reparametrizing the integral curve of X_K through z_0. ■

Proof of Theorem 7.7.3. Let H be the Hamiltonian for L, namely

$$H(q,p) = \frac{1}{2}\|p\|^2 + V(q),$$

and let H_e be that for the Jacobi metric:

$$H_e(q,p) = \frac{1}{2}(e - V(q))^{-1}\|p\|^2.$$

The factor $(e - V(q))^{-1}$ occurs because the inverse metric is used for the momenta. Clearly, $H = e$ defines the same set as $H_e = 1$, so the result follows from Proposition 7.7.4 if we show that e is a regular value of H and 1 is a regular value of H_e. Note that if $(q,p) \in H^{-1}(e)$, then $p \neq 0$, since $e > V(q)$ for all $q \in Q$. Therefore, $\mathbb{F}H(q,p) \neq 0$ for any $(q,p) \in H^{-1}(e)$, and hence $\mathbf{d}H(q,p) \neq 0$, that is, e is a regular value of H. Since

$$\mathbb{F}H_e(q,\dot{p}) = \frac{1}{2}(e - V(q))^{-1}\mathbb{F}H(q,p),$$

this also shows that

$$\mathbb{F}H_e(q,p) \neq 0 \quad \text{for all} \quad (q,p) \in H^{-1}(e) = H_e^{-1}(1),$$

and thus 1 is a regular value of H_e. ■

7.8 The Lagrange–d'Alembert Principle

In this section we study a generalization of Lagrange's equations for mechanical systems with exterior forces. A special class of such forces is dissipative forces, which will be studied at the end of this section.

Force Fields. Let $L : TQ \to \mathbb{R}$ be a Lagrangian function, let Z be the Lagrangian vector field associated to L, assumed to be a second-order equation, and denote by $\tau_Q : TQ \to Q$ the canonical projection. Recall that a vector field Y on TQ is called *vertical* if $T\tau_Q \circ Y = 0$. Such a vector field Y defines a one-form Δ^Y on TQ by contraction with Ω_L:

$$\Delta^Y = -\mathbf{i}_Y \Omega_L = Y \mathbin{\lrcorner} \Omega_L.$$

Proposition 7.8.1. *If Y is vertical, then Δ^Y is a **horizontal one-form**, that is, $\Delta^Y(U) = 0$ for any vertical vector field U on TQ. Conversely, given a horizontal one-form Δ on TQ, and assuming that L is regular, the vector field Y on TQ, defined by $\Delta = -\mathbf{i}_Y\Omega_L$, is vertical.*

Proof. This follows from a straightforward calculation in local coordinates. We use the fact that a vector field $Y(u,e) = (Y_1(u,e), Y_2(u,e))$ is vertical if and only if the first component Y_1 is zero, and the local formula for Ω_L derived earlier:

$$\begin{aligned}\Omega_L(u,e)(Y_1,Y_2),(U_1,U_2)) \\ &= \mathbf{D}_1(\mathbf{D}_2L(u,e)\cdot Y_1)\cdot U_1 - \mathbf{D}_1(\mathbf{D}_2L(u,e)\cdot U_1)\cdot Y_1 \\ &\quad + \mathbf{D}_2\mathbf{D}_2L(u,e)\cdot Y_1\cdot U_2 - \mathbf{D}_2\mathbf{D}_2L(u,e)\cdot U_1\cdot Y_2. \qquad (7.8.1)\end{aligned}$$

This shows that $(\mathbf{i}_Y\Omega_L)(U) = 0$ for all vertical U is equivalent to

$$\mathbf{D}_2\mathbf{D}_2L(u,e)(U_2,Y_1) = 0.$$

If Y is vertical, this is clearly true. Conversely, if L is regular and the last displayed equation is true, then $Y_1 = 0$, so Y is vertical. ■

Proposition 7.8.2. *Any fiber-preserving map $F : TQ \to T^*Q$ over the identity induces a horizontal one-form $\tilde{F}$ on TQ by*

$$\tilde{F}(v)\cdot V_v = \langle F(v), T_v\tau_Q(V_v)\rangle\,, \qquad (7.8.2)$$

where $v \in TQ$ and $V_v \in T_v(TQ)$. Conversely, formula (7.8.2) *defines, for any horizontal one-form $\tilde{F}$, a fiber-preserving map F over the identity. Any such F is called a* ***force field****, and thus, in the regular case, any vertical vector field Y is induced by a force field.*

Proof. Given F, formula (7.8.2) clearly defines a smooth one-form $\tilde{F}$ on TQ. If V_v is vertical, then the right-hand side of formula (7.8.2) vanishes, and so $\tilde{F}$ is a horizontal one-form. Conversely, given a horizontal one-form $\tilde{F}$ on TQ and given $v, w \in T_qQ$, let $V_v \in T_v(TQ)$ be such that $T_v\tau(V_v) = w$. Then define F by formula (7.8.2); that is, $\langle F(v), w\rangle = \tilde{F}(v)\cdot V_v$. Since $\tilde{F}$ is horizontal, we see that F is well-defined, and its expression in charts shows that it is smooth. ■

Treating Δ^Y as the exterior force one-form acting on a mechanical system with a Lagrangian L, we now will write the governing equations of motion.

The Lagrange–d'Alembert Principle. First, we recall the definition from Vershik and Faddeev [1981] and Wang and Krishnaprasad [1992].

Definition 7.8.3. *The* ***Lagrangian force*** *associated with a Lagrangian L and a given second-order vector field* (*the ultimate equations of motion*) *X is the horizontal one-form on TQ defined by*

$$\Phi_L(X) = \mathbf{i}_X\Omega_L - \mathbf{d}E. \qquad (7.8.3)$$

Given a horizontal one-form ω (*referred to as the* ***exterior force one-form***), *the* ***local Lagrange–d'Alembert principle*** *associated with the second-order vector field X on TQ states that*

$$\Phi_L(X) + \omega = 0. \qquad (7.8.4)$$

It is easy to check that $\Phi_L(X)$ is indeed horizontal if X is second-order. Conversely, if L is regular and if $\Phi_L(X)$ is horizontal, then X is second-order.

One can also formulate an equivalent principle in terms of variational principles.

Definition 7.8.4. *Given a Lagrangian L and a force field F, as defined in Proposition* 7.8.2*, the* ***integral Lagrange–d'Alembert principle*** *for a curve $q(t)$ in Q is*

$$\delta \int_a^b L(q(t), \dot{q}(t))\, dt + \int_a^b F(q(t), \dot{q}(t)) \cdot \delta q\, dt = 0, \tag{7.8.5}$$

where the variation is given by the usual expression

$$\begin{aligned} \delta \int_a^b L(q(t), \dot{q}(t))\, dt &= \int_a^b \left(\frac{\partial L}{\partial q^i} \delta q^i + \frac{\partial L}{\partial \dot{q}^i} \frac{d}{dt} \delta q^i \right) dt \\ &= \int_a^b \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} \right) \delta q^i\, dt \end{aligned} \tag{7.8.6}$$

for a given variation δq (vanishing at the endpoints).

The two forms of the Lagrange–d'Alembert principle are in fact equivalent. This will follow from the fact that both give the Euler–Lagrange equations with forcing in local coordinates (provided that Z is second-order). We shall see this in the following development.

Proposition 7.8.5. *Let the exterior force one-form ω be associated to a vertical vector field Y, that is, let $\omega = \Delta^Y = -\mathbf{i}_Y \Omega_L$. Then $X = Z + Y$ satisfies the local Lagrange–d'Alembert principle. Conversely, if, in addition, L is regular, the only second-order vector field X satisfying the local Lagrange–d'Alembert principle is $X = Z + Y$.*

Proof. For the first part, the equality $\Phi_L(X) + \omega = 0$ is a simple verification. For the converse, we already know that X is a solution, and uniqueness is guaranteed by regularity. ■

To develop the differential equations associated to $X = Z + Y$, we take $\omega = -\mathbf{i}_Y \Omega_L$ and note that in a coordinate chart, $Y(q, v) = (0, Y_2(q, v))$, since Y is vertical, that is, $Y_1 = 0$. From the local formula for Ω_L, we get

$$\omega(q, v) \cdot (u, w) = \mathbf{D}_2 \mathbf{D}_2 L(q, v) \cdot Y_2(q, v) \cdot u. \tag{7.8.7}$$

Letting $X(q, v) = (v, X_2(q, v))$, one finds that

$$\begin{aligned} &\Phi_L(X)(q, v) \cdot (u, w) \\ &\quad = (-\mathbf{D}_1(\mathbf{D}_2 L(q, v)\cdot) \cdot v - \mathbf{D}_2 \mathbf{D}_2 L(q, v) \cdot X_2(q, v) + \mathbf{D}_1 L(q, v)) \cdot u. \end{aligned} \tag{7.8.8}$$

Thus, the local Lagrange–d'Alembert principle becomes

$$(-\mathbf{D}_1(\mathbf{D}_2 L(q,v)\cdot)\cdot v - \mathbf{D}_2\mathbf{D}_2 L(q,v)\cdot X_2(q,v) + \mathbf{D}_1 L(q,v) \\ + \mathbf{D}_2\mathbf{D}_2 L(q,v)\cdot Y_2(q,v)) = 0. \quad (7.8.9)$$

Setting $v = dq/dt$ and $X_2(q,v) = dv/dt$, the preceding relation and the chain rule give

$$\frac{d}{dt}\mathbf{D}_2 L(q,v) - \mathbf{D}_1 L(q,v) = \mathbf{D}_2\mathbf{D}_2 L(q,v)\cdot Y_2(q,v), \quad (7.8.10)$$

which in finite dimensions reads

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot q^i}\right) - \frac{\partial L}{\partial q^i} = \frac{\partial^2 L}{\partial \dot q^i\,\partial \dot q^j} Y^j(q^k, \dot q^k). \quad (7.8.11)$$

The force one-form Δ^Y is therefore given by

$$\Delta^Y(q^k, \dot q^k) = \frac{\partial^2 L}{\partial \dot q^i\,\partial \dot q^j} Y^j(q^k, \dot q^k)\, dq^i, \quad (7.8.12)$$

and the corresponding force field is

$$F^Y = \left(q^i, \frac{\partial^2 L}{\partial \dot q^i\,\partial \dot q^j} Y^j(q^k, \dot q^k)\right). \quad (7.8.13)$$

Thus, the condition for an integral curve takes the form of the standard Euler–Lagrange equations with forces:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot q^i}\right) - \frac{\partial L}{\partial q^i} = F^Y_i(q^k, \dot q^k). \quad (7.8.14)$$

Since the integral Lagrange–d'Alembert principle gives the same equations, it follows that the two principles are equivalent. From now on, we will refer to either one as simply the ***Lagrange–d'Alembert principle.***

We summarize the results obtained so far in the following:

Theorem 7.8.6. *Given a regular Lagrangian and a force field $F : TQ \to T^*Q$, for a curve $q(t)$ in Q the following are equivalent:*

(i) *$q(t)$ satisfies the local Lagrange–d'Alembert principle;*

(ii) *$q(t)$ satisfies the integral Lagrange–d'Alembert principle; and*

(iii) *$q(t)$ is the base integral curve of the second-order equation $Z + Y$, where Y is the vertical vector field on TQ inducing the force field F by* (7.8.13)*, and Z is the Lagrangian vector field on L.*

The Lagrange–d'Alembert principle plays a crucial role in ***nonholonomic mechanics***, such as mechanical systems with rolling constraints. See, for example, Bloch, Krishnaprasad, Marsden, and Murray [1996] and references therein.

Dissipative Forces. Let E denote the energy defined by L, that is, $E = A - L$, where $A(v) = \langle \mathbb{F}L(v), v \rangle$ is the action of L.

Definition 7.8.7. *A vertical vector field Y on TQ is called **weakly dissipative** if $\langle \mathbf{d}E, Y \rangle \leq 0$ at all points of TQ. If the inequality is strict off the zero section of TQ, then Y is called **dissipative**. A **dissipative Lagrangian system** on TQ is a vector field $Z + Y$, for Z a Lagrangian vector field and Y a dissipative vector field.*

Corollary 7.8.8. *A vertical vector field Y on TQ is dissipative if and only if the force field F^Y that it induces satisfies $\left\langle F^Y(v), v \right\rangle < 0$ for all nonzero $v \in TQ$ (≤ 0 for the weakly dissipative case).*

Proof. Let Y be a vertical vector field. By Proposition 7.8.1, Y induces a horizontal one-form $\Delta^Y = -\mathbf{i}_Y \Omega_L$ on TQ, and by Proposition 7.8.2 , Δ^Y in turn induces a force field F^Y given by

$$\left\langle F^Y(v), w \right\rangle = \Delta^Y(v) \cdot V_v = -\Omega_L(v)(Y(v), V_v), \tag{7.8.15}$$

where $T\tau_Q(V_v) = w$ and $V_v \in T_v(TQ)$. If Z denotes the Lagrangian system defined by L, we get

$$\begin{aligned}
(\mathbf{d}E \cdot Y)(v) &= (\mathbf{i}_Z \Omega_L)(Y)(v) = \Omega_L(Z, Y)(v) \\
&= -\Omega_L(v)(Y(v), Z(v)) \\
&= \left\langle F^Y(v), T_v\tau(Z(v)) \right\rangle \\
&= \left\langle F^Y(v), v \right\rangle,
\end{aligned}$$

since Z is a second-order equation. Thus, $\mathbf{d}E \cdot Y < 0$ if and only if $\left\langle F^Y(v), v \right\rangle < 0$ for all $v \in TQ$. ■

Definition 7.8.9. *Given a dissipative vector field Y on TQ, let $F^Y : TQ \to T^*Q$ be the induced force field. If there is a function $R : TQ \to \mathbb{R}$ such that F^Y is the fiber derivative of $-R$, then R is called a **Rayleigh dissipation function**.*

Note that in this case, $\mathbf{D}_2 R(q, v) \cdot v > 0$ for the dissipativity of Y. Thus, if R is linear in the fiber variable, the Rayleigh dissipation function takes on the classical form $\langle \mathcal{R}(q)v, v \rangle$, where $\mathcal{R}(q) : TQ \to T^*Q$ is a bundle map over the identity that defines a symmetric positive definite form on each fiber of TQ.

Finally, if the force field is given by a Rayleigh dissipation function R, then the Euler–Lagrange equations with forcing become

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = -\frac{\partial R}{\partial \dot{q}^i}. \tag{7.8.16}$$

Combining Corollary 7.8.8 with the fact that the differential of E along Z is zero, we find that under the flow of the Euler–Lagrange equations with forcing of Rayleigh dissipation type, we have

$$\frac{d}{dt}E(q,v) = F(v)\cdot v = -\mathbb{F}R(q,v)\cdot v < 0. \tag{7.8.17}$$

Exercises

◇ **7.8-1.** What is the power or rate of work equation (see §2.1) for a system with forces on a Riemannian manifold?

◇ **7.8-2.** Write the equations for a ball in a rotating hoop, including friction, in the language of this section (see §2.8). Compute the Rayleigh dissipation function.

◇ **7.8-3.** Consider a Riemannian manifold Q and a potential function $V : Q \to \mathbb{R}$. Let K denote the kinetic energy function and let $\omega = -\mathbf{d}V$. Show that the Lagrange–d'Alembert principle for K with external forces given by the one-form ω produces the same dynamics as the standard kinetic minus potential Lagrangian.

7.9 The Hamilton–Jacobi Equation

In §6.5 we studied generating functions of canonical transformations. Here we link them with the flow of a Hamiltonian system via the Hamilton–Jacobi equation. In this section we approach Hamilton–Jacobi theory from the point of view of extended phase space. In the next chapter we will have another look at Hamilton–Jacobi theory from the variational point of view, as it was originally developed by Jacobi [1866]. In particular, we will show in that section, roughly speaking, that the integral of the Lagrangian along solutions of the Euler–Lagrange equations, but thought of as a function of the endpoints, satisfies the Hamilton–Jacobi equation.

Canonical Transformations and Generating Functions. We consider a symplectic manifold P and form the ***extended phase space*** $P\times\mathbb{R}$. For our purposes in this section, we will use the following definition. A ***time-dependent canonical transformation*** is a diffeomorphism

$$\rho : P\times\mathbb{R} \to P\times\mathbb{R}$$

of the form

$$\rho(z,t) = (\rho_t(z), t),$$

where for each $t\in\mathbb{R}$, $\rho_t : P \to P$ is a symplectic diffeomorphism.

In this section we will specialize to the case of cotangent bundles, so assume that $P = T^*Q$ for a configuration manifold Q. For each fixed t, let $S_t : Q \times Q \to \mathbb{R}$ be the generating function for a time-dependent symplectic map, as described in §6.5. Thus, we get a function $S : Q \times Q \times \mathbb{R} \to \mathbb{R}$ defined by $S(q_1, q_2, t) = S_t(q_1, q_2)$. As explained in §6.5, one has to be aware that in general, generating functions are defined only locally, and indeed, the global theory of generating functions and the associated global Hamilton–Jacobi theory is more sophisticated. We will give a brief (optional) introduction to this general theory at the end of this section. See also Abraham and Marsden [1978, Section 5.3] for more information and references. Since our goal in the first part of this section is to give an *introductory presentation of the theory*, we will do many of the calculations in coordinates.

Recall that in local coordinates, the conditions for a generating function are written as follows. If the transformation ψ has the local expression

$$\psi : (q^i, p_i, t) \mapsto (\overline{q}^i, \overline{p}_i, t),$$

with inverse denoted by

$$\phi : (\overline{q}^i, \overline{p}_i, t) \mapsto (q^i, p_i, t),$$

and if $S(q^i, \overline{q}^i, t)$ is a generating function for ψ, we have the relations

$$\overline{p}_i = -\frac{\partial S}{\partial \overline{q}^i} \quad \text{and} \quad p_i = \frac{\partial S}{\partial q^i}. \tag{7.9.1}$$

From (7.9.1) it follows that

$$\begin{aligned} p_i\, dq^i &= \overline{p}_i\, d\overline{q}^i + \frac{\partial S}{\partial q^i} dq^i + \frac{\partial S}{\partial \overline{q}^i} d\overline{q}^i \\ &= \overline{p}_i\, d\overline{q}^i - \frac{\partial S}{\partial t} dt + \mathbf{d}S, \end{aligned} \tag{7.9.2}$$

where $\mathbf{d}S$ is the differential of S as a function on $Q \times Q \times \mathbb{R}$:

$$\mathbf{d}S = \frac{\partial S}{\partial q^i} dq^i + \frac{\partial S}{\partial \overline{q}^i} d\overline{q}^i + \frac{\partial S}{\partial t} dt.$$

Let $K : T^*Q \times \mathbb{R} \to \mathbb{R}$ be an arbitrary function. From (7.9.2) we get the following basic relationship:

$$p_i\, dq^i - K(q^i, p_i, t)\, dt = \overline{p}_i\, d\overline{q}^i - \overline{K}(\overline{q}^i, \overline{p}_i, t)\, dt + \mathbf{d}S(q^i, \overline{q}^i, t), \tag{7.9.3}$$

where $\overline{K}(\overline{q}^i, \overline{p}_i, t) = K(q^i, p_i, t) + \partial S(q^i, \overline{q}^i, t)/\partial t$. If we define

$$\Theta_K = p_i\, dq^i - K\, dt, \tag{7.9.4}$$

then (7.9.3) is equivalent to

$$\Theta_K = \psi^* \Theta_{\overline{K}} + \psi^* \mathbf{d}S, \tag{7.9.5}$$

where $\psi : T^*Q \times \mathbb{R} \to Q \times Q \times \mathbb{R}$ is the map

$$(q^i, p_i, t) \mapsto (q^i, \overline{q}^i(q^j, p_j, t), t).$$

By taking the exterior derivative of (7.9.3) (or (7.9.5)), it follows that

$$dq^i \wedge dp_i + dK \wedge dt = d\overline{q}^i \wedge d\overline{p}_i + d\overline{K} \wedge dt. \tag{7.9.6}$$

This may be written as

$$\Omega_K = \psi^* \Omega_{\overline{K}}, \tag{7.9.7}$$

where $\Omega_K = -\mathbf{d}\Theta_K = dq^i \wedge dp_i + dK \wedge dt$.

Recall from Exercise 6.2-3 that given a time-dependent function K and associated time-dependent vector field X_K on T^*Q, the vector field $\tilde{X}_K = (X_K, 1)$ on $T^*Q \times \mathbb{R}$ is uniquely determined (among all vector fields with a 1 in the second component) by the equation $\mathbf{i}_{\tilde{X}_K} \Omega_K = 0$. From this relation and (7.9.7), we get

$$0 = \psi_*(\mathbf{i}_{\tilde{X}_K} \Omega_K) = \mathbf{i}_{\psi_*(\tilde{X}_K)} \psi_* \Omega_K = \mathbf{i}_{\psi_*(\tilde{X}_K)} \Omega_{\overline{K}}.$$

Since ψ is the identity in the second component, that is, it preserves time, the vector field $\psi_*(\tilde{X}_K)$ has a 1 in the second component, and therefore by uniqueness of such vector fields we get the identity

$$\psi_*(\tilde{X}_K) = \tilde{X}_{\overline{K}}. \tag{7.9.8}$$

The Hamilton–Jacobi Equation. The data we shall need are a Hamiltonian H and a generating function S, as above.

Definition 7.9.1. *Given a time-dependent Hamiltonian H and a transformation ψ with generating function S as above, we say that the* ***Hamilton–Jacobi equation*** *holds if*

$$H\left(q^1, \ldots, q^n, \frac{\partial S}{\partial q^1}, \ldots, \frac{\partial S}{\partial q^n}, t\right) + \frac{\partial S}{\partial t}(q^i, \overline{q}^i, t) = 0, \tag{7.9.9}$$

in which $\partial S/\partial q^i$ are evaluated at $(q^i, \overline{q}^i, t)$ and in which the $\overline{q}^i$ are regarded as constants.

The Hamilton–Jacobi equation may be regarded as a nonlinear partial differential equation for the function S relative to the variables $(q^1, \ldots, q^n, t)$ depending parametrically on $(\overline{q}^1, \ldots, \overline{q}^n)$.

Definition 7.9.2. *We say that the map ψ* ***transforms a vector field $\tilde{X}$ to equilibrium*** *if*

$$\psi_* \tilde{X} = (0, 1). \tag{7.9.10}$$

If ψ transforms $\tilde{X}$ to equilibrium, then the integral curves of $\tilde{X}$ with initial conditions (q_0^i, p_i^0, t_0) are given by

$$(q^i(t), p_i(t), t) = \psi^{-1}(\overline{q}^i(q_0^i, p_i^0, t_0), \overline{p}_i(q_0^i, p_i^0, t_0), t + t_0), \qquad (7.9.11)$$

since the integral curves of the constant vector field $(0, 1)$ are just straight lines in the t-direction and since ψ maps integral curves of $\tilde{X}$ to those of $(0, 1)$. In other words, if a map transforms a vector field $\tilde{X}$ to equilibrium, the integral curves of $\tilde{X}$ are represented by straight lines in the image space, and so the vector field has been "integrated."

Notice that if ϕ is the inverse of ψ, then ϕ_t is the flow of the vector field X in the usual sense.

Theorem 7.9.3 (Hamilton–Jacobi)**.**

(i) *Suppose that S satisfies the Hamilton–Jacobi equation for a given time-dependent Hamiltonian H and that S generates a time-dependent canonical transformation ψ. Then ψ transforms $\tilde{X}_H$ to equilibrium. Thus, as explained above, the solution of Hamilton's equations for H are given in terms of ψ by* (7.9.11)*.*

(ii) *Conversely, if ψ is a time-dependent canonical transformation with generating function S that transforms $\tilde{X}_H$ to equilibrium, then there is a function $\hat{S}$, which differs from S only by a function of t that also generates ψ, and satisfies the Hamilton–Jacobi equation for H.*

Proof. To prove (i), assume that S satisfies the Hamilton–Jacobi equation. As we explained above, this means that $\overline{H} = 0$. From (7.9.8) we get

$$\psi_* \tilde{X}_H = \tilde{X}_{\overline{H}} = (0, 1).$$

This proves the first statement.

To prove the converse (ii), assume that

$$\psi_* \tilde{X}_H = (0, 1),$$

and so, again by (7.9.8),

$$\tilde{X}_{\overline{H}} = \tilde{X}_0 = (0, 1),$$

which means that $\overline{H}$ is a constant relative to the variables $(\overline{q}^i, \overline{p}_i)$ (its Hamiltonian vector field at each instant of time is zero) and thus $\overline{H} = f(t)$, a function of time only. We can then modify S to $\hat{S} = S - F$, where $F(t) = \int^t f(s) ds$. This function, differing from S by a function of time alone, generates the same map ψ. Since

$$0 = \overline{H} - f(t) = H + \partial S/\partial t - dF/dt = H + \partial \hat{S}/\partial t,$$

and $\partial S/\partial q^i = \partial \hat{S}/\partial q^i$, we see that $\hat{S}$ satisfies the Hamilton–Jacobi equation for H. ■

Remarks.

1. In general, the function S develops *singularities*, or *caustics*, as time increases, so it must be used with care. This process is, however, fundamental in geometric optics and in quantization. Moreover, one has to be careful with the sense in which S generates the identity at $t = 0$, as it might have singular behavior in t.

2. Here is another link between the Lagrangian and Hamiltonian view of the Hamilton–Jacobi theory. Define S for t close to a fixed time t_0 by the ***action integral***

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

where $q^i(s)$ is the solution of the Euler–Lagrange equation equaling $\overline{q}^i$ at time t_0 and equaling q^i at time t. We will show in §8.2 that S satisfies the Hamilton–Jacobi equation. See Arnold [1989, Section 4.6] and Abraham and Marsden [1978, Section 5.2] for more information.

3. If H is time-independent and W satisfies the time-independent Hamilton–Jacobi equation

$$H\left(q^i, \frac{\partial W}{\partial q^i}\right) = E,$$

then $S(q^i, \overline{q}^i, t) = W(q^i, \overline{q}^i) - tE$ satisfies the time-dependent Hamilton–Jacobi equation, as is easily checked. When using this remark, it is important to remember that E is not really a "constant," but it equals $H(\overline{q}, \overline{p})$, the energy evaluated at $(\overline{q}, \overline{p})$, which will eventually be the initial conditions. We emphasize that one must generate the time t-map using S rather than W.

4. The Hamilton–Jacobi equation is fundamental in the study of the quantum–classical relationship, which is described in the Internet supplement for Chapter 7.

5. The action function S is a key tool used in the proof of the ***Liouville–Arnold theorem***, which gives the existence of action angle coordinates for systems with integrals in involution; see Arnold [1989] and Abraham and Marsden [1978] for details.

6. The Hamilton–Jacobi equation plays an important role in the development of numerical integrators that preserve the symplectic structure (see de Vogelaére [1956], Channell [1983], Feng [1986], Channell and Scovel [1990], Ge and Marsden [1988], Marsden [1992], and Wendlandt and Marsden [1997]).

7. The method of separation of variables. It is sometimes possible to simplify and even solve the Hamilton–Jacobi equation by what is often called the method of separation of variables. Assume that in the Hamilton–Jacobi equation the coordinate q^1 and the term $\partial S/\partial q^1$ appear jointly in some expression $f(q^1, \partial S/\partial q^1)$ that does not involve $q^2, \ldots, q^n, t$. That is, we can write H in the form

$$H\left(q^1, q^2, \ldots, q^n, p_1, p_2, \ldots, p_n\right) = \tilde{H}(f(q^1, p_1), q^2, \ldots, q^n, p_2, \ldots, p_n)$$

for some smooth functions f and $\tilde{H}$. Then one seeks a solution of the Hamilton–Jacobi equation in the form

$$S(q^i, \overline{q}^i, t) = S_1(q^1, \overline{q}^1) + \tilde{S}(q^2, \ldots, q^n, \overline{q}^2, \ldots, \overline{q}^n).$$

We then note that if S_1 solves

$$f\left(q^1, \frac{\partial S_1}{\partial q^1}\right) = C(\overline{q}^1)$$

for an arbitrary function $C(\overline{q}^1)$ and if $\tilde{S}$ solves

$$\tilde{H}\left(C(\overline{q}^1), q^2, \ldots, q^n, \frac{\partial \tilde{S}}{\partial q^2}, \ldots, \frac{\partial \tilde{S}}{\partial q^n}\right) + \frac{\partial \tilde{S}}{\partial t} = 0,$$

then S solves the original Hamilton–Jacobi equation. In this way, one of the variables is eliminated, and one tries to repeat the procedure.

A closely related situation occurs when H is independent of time and one seeks a solution of the form

$$S(q^i, \overline{q}^i, t) = W(q^i, \overline{q}^i) + S_1(t).$$

The resulting equation for S_1 has the solution $S_1(t) = -Et$, and the remaining equation for W is the time-independent Hamilton–Jacobi equation as in Remark 3.

If q^1 is a cyclic variable, that is, if H does not depend explicitly on q^1, then we can choose $f(q^1, p_1) = p_1$, and correspondingly, we can choose $S_1(q^1) = C(\overline{q}^1)q^1$. In general, if there are k cyclic coordinates $q^1, q^2, \ldots, q^k$, we seek a solution to the Hamilton–Jacobi equation of the form

$$S(q^i, \overline{q}^i, t) = \sum_{j=1}^{k} C_j(\overline{q}^j)q^j + \tilde{S}(q^{k+1}, \ldots, q^n, \overline{q}^{k+1}, \ldots, \overline{q}^n, t),$$

with $p_i = C_i(\overline{q}^i)$, $i = 1, \ldots, k$, being the momenta conjugate to the cyclic variables. ♦

The Geometry of Hamilton–Jacobi Theory (Optional). Now we describe briefly and informally some additional geometry connected with the Hamilton–Jacobi equation (7.9.9). For each $x = (q^i, t) \in \tilde{Q} := Q \times \mathbb{R}$, $\mathbf{d}S(x)$ is an element of the cotangent bundle $T^*\tilde{Q}$. We suppress the dependence of S on $\overline{q}^i$ for the moment, since it does not play an immediate role. As x varies in $\tilde{Q}$, the set $\{\, \mathbf{d}S(x) \mid x \in \tilde{Q} \,\}$ defines a submanifold of $T^*\tilde{Q}$ that in terms of coordinates is given by $p_j = \partial S/\partial q^j$ and $p = \partial S/\partial t$; here the variables conjugate to q^i are denoted by p_i and that conjugate to t is denoted by p. We will write $\xi_i = p_i$ for $i = 1, 2, \ldots, n$ and $\xi_{n+1} = p$. We call this submanifold the *range*, or *graph*, of $\mathbf{d}S$ (either term is appropriate, depending on whether one thinks of $\mathbf{d}S$ as a mapping or as a section of a bundle) and denote it by $\operatorname{graph} \mathbf{d}S \subset T^*\tilde{Q}$. The restriction of the canonical symplectic form on $T^*\tilde{Q}$ to $\operatorname{graph} \mathbf{d}S$ is zero, since

$$\sum_{j=1}^{n+1} dx^j \wedge d\xi_j = \sum_{j=1}^{n+1} dx^j \wedge d\frac{\partial S}{\partial x_j} = \sum_{j,k=1}^{n+1} dx^j \wedge dx^k \, \frac{\partial^2 S}{\partial x^j \partial x^k} = 0.$$

Moreover, the dimension of the submanifold $\operatorname{graph} \mathbf{d}S$ is half of the dimension of the symplectic manifold $T^*\tilde{Q}$. Such a submanifold is called ***Lagrangian***, as we already mentioned in connection with generating functions (§6.5). What is important here is that the projection from $\operatorname{graph} \mathbf{d}S$ to $\tilde{Q}$ is a diffeomorphism, and even more, the converse holds: If $\Lambda \subset T^*\tilde{Q}$ is a Lagrangian submanifold of $T^*\tilde{Q}$ such that the projection on $\tilde{Q}$ is a diffeomorphism in a neighborhood of a point $\lambda \in \Lambda$, then in some neighborhood of λ we can write $\Lambda = \operatorname{graph} \mathbf{d}\varphi$ for some function φ. To show this, notice that because the projection is a diffeomorphism, Λ is given (around λ) as a submanifold of the form $(x^j, \rho_j(x))$. The condition for Λ to be Lagrangian requires that on Λ,

$$\sum_{j=1}^{n+1} dx^j \wedge d\xi_j = 0,$$

that is,

$$\sum_{j=1}^{n+1} dx^j \wedge d\rho_j(x) = 0, \quad \text{i.e.,} \quad \frac{\partial \rho_j}{\partial x^k} - \frac{\partial \rho_k}{\partial x^j} = 0;$$

thus, there is a φ such that $\rho_j = \partial\varphi/\partial x^j$, which is the same as $\Lambda = \operatorname{graph} \mathbf{d}\varphi$. The conclusion of these remarks is that Lagrangian submanifolds of $T^*\tilde{Q}$ are natural generalizations of graphs of differentials of functions on $\tilde{Q}$. Note that Lagrangian submanifolds are defined even if the projection to $\tilde{Q}$ is not a diffeomorphism. For more information on Lagrangian manifolds and generating functions, see Abraham and Marsden [1978], Weinstein [1977], and Guillemin and Sternberg [1977].

From the point of view of Lagrangian submanifolds, *the graph of the differential of a solution of the Hamilton–Jacobi equation is a Lagrangian submanifold of* $T^*\tilde{Q}$ *that is contained in the surface* $\tilde{H}_0 \subset T^*\tilde{Q}$ *defined by the equation* $\tilde{H} := p + H(q^i, p_i, t) = 0$. Here, as above, $p = \xi_{n+1}$ is the momentum conjugate to t. This point of view allows one to include solutions that are singular in the usual context. This is not the only benefit: We also get more insight in the content of the Hamilton–Jacobi Theorem 7.9.3. The tangent space to $\tilde{H}_0$ has dimension 1 less than the dimension of the symplectic manifold $T^*\tilde{Q}$, and it is given by the set of vectors X such that $(dp + \mathbf{d}H)(X) = 0$. If a vector Y is in the symplectic orthogonal of $T_{(x,\xi)}(\tilde{H}_0)$, that is,

$$\sum_{j=1}^{n+1}(dx^j \wedge d\xi_j)(X, Y) = 0$$

for all $X \in T_{(x,\xi)}(\tilde{H}_0)$, then Y is a multiple of the vector field

$$X_{\tilde{H}} = \frac{\partial}{\partial t} - \frac{\partial H}{\partial t}\frac{\partial}{\partial p} + X_H$$

evaluated at (x, ξ). Moreover, the integral curves of $X_{\tilde{H}}$ projected to (q^i, p_i) are the solutions of Hamilton's equations for H.

The key observation that links Hamilton's equations and the Hamilton–Jacobi equation is that *the vector field* $X_{\tilde{H}}$*, which is obviously tangent to* $\tilde{H}_0$*, is, moreover, tangent to any Lagrangian submanifold contained in* $\tilde{H}_0$ (the reason for this is a very simple algebraic fact given in Exercise 7.9-3). This is the same as saying that a solution of Hamilton's equations for $\tilde{H}$ is either disjoint from a Lagrangian submanifold contained in $\tilde{H}_0$ or completely contained in it. This gives a way to construct a solution of the Hamilton–Jacobi equation starting from an initial condition at $t = t_0$. Namely, take a Lagrangian submanifold Λ_0 in T^*Q and embed it in $T^*\tilde{Q}$ at $t = t_0$ using

$$(q^i, p_i) \mapsto (q^i, t = t_0, p_i, p = -H(q^i, p_i, t_0)).$$

The result is an isotropic submanifold $\tilde{\Lambda}_0 \subset T^*\tilde{Q}$, that is, a submanifold on which the canonical form vanishes. Now take all integral curves of $X_{\tilde{H}}$ whose initial conditions lie in $\tilde{\Lambda}_0$. The collection of these curves spans a manifold Λ whose dimension is one higher than $\tilde{\Lambda}_0$. It is obtained by flowing $\tilde{\Lambda}_0$ along $X_{\tilde{H}}$; that is, $\Lambda = \cup_t \Lambda_t$, where $\Lambda_t = \Phi_t(\tilde{\Lambda}_0)$ and Φ_t is the flow of $X_{\tilde{H}}$. Since $X_{\tilde{H}}$ is tangent to $\tilde{H}_0$ and $\Lambda_0 \subset \tilde{H}_0$, we get $\Lambda_t \subset \tilde{H}_0$ and hence $\Lambda \subset \tilde{H}_0$. Since the flow Φ_t of $X_{\tilde{H}}$ is a canonical map, it leaves the symplectic form of $T^*\tilde{Q}$ invariant and therefore takes an isotropic submanifold into an isotropic one; in particular, Λ_t is an isotropic submanifold of $T^*\tilde{Q}$. The tangent space of Λ at some $\lambda \in \Lambda_t$ is a direct sum of the tangent space of

Λ_t and the subspace generated by $X_{\tilde{H}}$. Since the first subspace is contained in $T_\lambda \tilde{H}_0$ and the second is symplectically orthogonal to $T_\lambda \tilde{H}_0$, we see that Λ is also an isotropic submanifold of $T^*\tilde{Q}$. But its dimension is half that of $T^*\tilde{Q}$, and therefore Λ is a Lagrangian submanifold contained in $\tilde{H}_0$, that is, it is a solution of the Hamilton–Jacobi equation with initial condition Λ_0 at $t = t_0$.

Using the above point of view it is easy to understand the singularities of a solution of the Hamilton–Jacobi equation. They correspond to those points of the Lagrangian manifold solution where the projection to $\tilde{Q}$ is not a local diffeomorphism. These singularities might be present in the initial condition (that is, Λ_0 might not locally project diffeomorphically to Q), or they might appear at later times by folding the submanifolds Λ_t as t varies. The projection of such a singular point to $\tilde{Q}$ is called a ***caustic point*** of the solution. Caustic points are of fundamental importance in geometric optics and the semiclassical approximation of quantum mechanics. We refer to Abraham and Marsden [1978, Section 5.3] and Guillemin and Sternberg [1984] for further information.

Exercises

◇ **7.9-1.** Solve the Hamilton–Jacobi equation for the harmonic oscillator. Check directly the validity of the Hamilton–Jacobi theorem (connecting the solution of the Hamilton–Jacobi equation and the flow of the Hamiltonian vector field) for this case.

◇ **7.9-2.** Verify by *direct calculation* the following. Let $W(q, \overline{q})$ and

$$H(q,p) = \frac{p^2}{2m} + V(q)$$

be given, where $q, p \in \mathbb{R}$. Show that for $p \neq 0$,

$$\frac{1}{2m}(W_q)^2 + V = E$$

and $\dot{q} = p/m$ if and only if $(q, W_q(q, \overline{q}))$ satisfies Hamilton's equation with energy E.

◇ **7.9-3.** Let (V, Ω) be a symplectic vector space and $W \subset V$ be a linear subspace. Recall from §2.4 that

$$W^\Omega = \{\, v \in V \mid \Omega(v, w) = 0 \text{ for all } w \in W \,\}$$

denotes the symplectic orthogonal of W. A subspace $L \subset V$ is called ***Lagrangian*** if $L = L^\Omega$. Show that if $L \subset W$ is a Lagrangian subspace, then $W^\Omega \subset L$.

◇ **7.9-4.** Solve the Hamilton–Jacobi equation for a central force field. Check directly the validity of the Hamilton–Jacobi theorem.