2
Hamiltonian Systems on Linear Symplectic Spaces

A natural arena for Hamiltonian mechanics is a symplectic or Poisson manifold. The next few chapters concentrate on the symplectic case, while Chapter 10 introduces the Poisson case. The symplectic context focuses on the symplectic two-form $\sum dq^i \wedge dp_i$ and its infinite-dimensional analogues, while the Poisson context looks at the Poisson bracket as the fundamental object.

To facilitate an understanding of a number of points, we begin this chapter with the theory in linear spaces in which case the symplectic form becomes a skew-symmetric bilinear form that can be studied by means of linear-algebraic methods. This linear setting is already adequate for a number of interesting examples such as the wave equation and Schrödinger's equation.

Later, in Chapter 4, we make the transition to manifolds, and we generalize symplectic structures to manifolds in Chapters 5 and 6. In Chapters 7 and 8 we study the basics of Lagrangian mechanics, which are based primarily on variational principles rather than on symplectic or Poisson structures. This apparently very different approach is, however, shown to be equivalent to the Hamiltonian one under appropriate hypotheses.

2.1 Introduction

To motivate the introduction of symplectic geometry in mechanics, we briefly recall from §1.1 the classical transition from Newton's second law to the Lagrange and Hamilton equations. ***Newton's second law*** for a particle moving in Euclidean three-space $\mathbb{R}^3$, under the influence of a ***potential energy*** $V(\mathbf{q})$, is

$$\mathbf{F} = m\mathbf{a}, \tag{2.1.1}$$

where $\mathbf{q} \in \mathbb{R}^3$, $\mathbf{F}(\mathbf{q}) = -\nabla V(\mathbf{q})$ is the ***force***, m is the mass of the particle, and $\mathbf{a} = d^2\mathbf{q}/dt^2$ is the acceleration (assuming that we start in a postulated privileged coordinate frame called an ***inertial frame***)[1]. The potential energy V is introduced through the notion of work and the assumption that the force field is conservative as shown in most books on vector calculus. The introduction of the ***kinetic energy***

$$K = \frac{1}{2}m\left\|\frac{d\mathbf{q}}{dt}\right\|^2$$

is through the ***power***, or ***rate of work, equation***

$$\frac{dK}{dt} = m\,\langle \dot{\mathbf{q}}, \ddot{\mathbf{q}}\rangle = \langle \dot{\mathbf{q}}, \mathbf{F}\rangle,$$

where $\langle\, ,\rangle$ denotes the inner product on $\mathbb{R}^3$.

The ***Lagrangian*** is defined by

$$L(q^i, \dot{q}^i) = \frac{m}{2}\|\dot{\mathbf{q}}\|^2 - V(\mathbf{q}), \tag{2.1.2}$$

and one checks by direct calculation that Newton's second law is equivalent to the ***Euler–Lagrange equations***

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0, \tag{2.1.3}$$

which are second-order differential equations in q^i; the equations (2.1.3) are worthy of independent study for a general L, since they are the equations for stationary values of the ***action integral***

$$\delta \int_{t_1}^{t_2} L(q^i, \dot{q}^i)\, dt = 0, \tag{2.1.4}$$

as will be discussed in detail later. These ***variational principles*** play a fundamental role throughout mechanics—both in particle mechanics and field theory.

---

[1]Newton and subsequent workers in mechanics thought of this inertial frame as one "fixed relative to the distant stars." While this raises serious questions about what this could really mean mathematically or physically, it remains a good starting point. Deeper insight is found in Chapter 8 and in courses in general relativity.

It is easily verified that $dE/dt = 0$, where E is the ***total energy***:

$$E = \frac{1}{2}m\|\dot{\mathbf{q}}\|^2 + V(\mathbf{q}).$$

Lagrange and Hamilton observed that it is convenient to introduce the momentum $p_i = m\dot{q}^i$ and rewrite E as a function of p_i and q^i by letting

$$H(\mathbf{q}, \mathbf{p}) = \frac{\|\mathbf{p}\|^2}{2m} + V(\mathbf{q}), \tag{2.1.5}$$

for then Newton's second law is equivalent to ***Hamilton's canonical equations***

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}, \tag{2.1.6}$$

which is a *first-order* system in $(\mathbf{q}, \mathbf{p})$-space, or ***phase space***.

Matrix Notation. For a deeper understanding of Hamilton's equations, we recall some matrix notation (see Abraham, Marsden, and Ratiu [1988, Section 5.1] for more details). Let E be a real vector space and E^* its dual space. Let $e_1, \ldots, e_n$ be a basis of E with the associated dual basis for E^* denoted by $e^1, \ldots, e^n$; that is, e^i is defined by

$$\left\langle e^i, e_j \right\rangle := e^i(e_j) = \delta^i_j,$$

which equals 1 if $i = j$ and 0 if $i \neq j$. Vectors $v \in E$ are written $v = v^i e_i$ (a sum on i is understood) and covectors $\alpha \in E^*$ as $\alpha = \alpha_i e^i$; v^i and α_i are the ***components*** of v and α, respectively.

If $A : E \to F$ is a linear transformation, its ***matrix*** relative to bases $e_1, \ldots, e_n$ of E and $f_1, \ldots, f_m$ of F is denoted by $A^j{}_i$ and is defined by

$$A(e_i) = A^j{}_i f_j; \quad \text{i.e.,} \quad [A(v)]^j = A^j{}_i v^i. \tag{2.1.7}$$

Thus, the columns of the matrix of A are $A(e_1), \ldots, A(e_n)$; the upper index is the row index, and the lower index is the column index. For other linear transformations, we place the indices in their corresponding places. For example, if $A : E^* \to F$ is a linear transformation, its matrix A^{ij} satisfies $A(e^j) = A^{ij} f_i$; that is, $[A(\alpha)]^i = A^{ij}\alpha_j$.

If $B : E \times F \to \mathbb{R}$ is a bilinear form, that is, it is linear separately in each factor, its ***matrix*** B_{ij} is defined by

$$B_{ij} = B(e_i, f_j); \quad \text{i.e.,} \quad B(v, w) = v^i B_{ij} w^j. \tag{2.1.8}$$

Define the ***associated*** linear map $B^\flat : E \to F^*$ by

$$B^\flat(v)(w) = B(v, w)$$

and observe that $B^\flat(e_i) = B_{ij} f^j$. Since $B^\flat(e_i)$ is the ith column of the matrix representing the linear map $B^\flat$, it follows that *the matrix of $B^\flat$ in the bases $e_1, \ldots, e_n, f^1, \ldots, f^n$ is the transpose of B_{ij}*; that is,

$$[B^\flat]_{ji} = B_{ij}. \tag{2.1.9}$$

Let Z denote the vector space of (q,p)'s and write $z = (q,p)$. Let the coordinates q^j, p_j be collectively denoted by z^I, $I = 1, \ldots, 2n$. One reason for the notation z is that if one thinks of z as a *complex variable* $z = q + ip$, then Hamilton's equations are equivalent to the following complex form of Hamilton's equations (see Exercise 2.1-1):

$$\dot{z} = -2i \frac{\partial H}{\partial \overline{z}}, \tag{2.1.10}$$

where $\partial/\partial\overline{z} := (\partial/\partial q + i\partial/\partial p)/2$.

Symplectic and Poisson Structures. We can view Hamilton's equations (2.1.6) as follows. Think of the operation

$$\mathbf{d}H(z) = \left(\frac{\partial H}{\partial q^i}, \frac{\partial H}{\partial p_i} \right) \mapsto \left(\frac{\partial H}{\partial p_i}, -\frac{\partial H}{\partial q^i} \right) =: X_H(z), \tag{2.1.11}$$

which forms a vector field X_H, called the ***Hamiltonian vector field***, from the differential of H, as the composition of the linear map

$$R : Z^* \to Z$$

with the differential $\mathbf{d}H(z)$ of H. The matrix of R is

$$[R^{AB}] = \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{bmatrix} =: \mathbb{J}, \tag{2.1.12}$$

where we write $\mathbb{J}$ for the specific matrix (2.1.12) sometimes called the ***symplectic matrix***. Here, $\mathbf{0}$ is the $n \times n$ zero matrix and $\mathbf{1}$ is the $n \times n$ identity matrix. Thus,

$$X_H(z) = R \cdot \mathbf{d}H(z) \tag{2.1.13}$$

or, if the components of X_H are denoted by X^I, $I = 1, \ldots, 2n$,

$$X^I = R^{IJ} \frac{\partial H}{\partial z^J}, \quad \text{i.e.,} \quad X_H = \mathbb{J} \nabla H, \tag{2.1.14}$$

where ∇H is the *naive gradient* of H, that is, the row vector $\mathbf{d}H$ but regarded as a column vector.

Let $B(\alpha, \beta) = \langle \alpha, R(\beta) \rangle$ be the bilinear form associated to R, where $\langle \, , \rangle$ denotes the canonical pairing between Z^* and Z. One calls either the bilinear form B or its associated linear map R the ***Poisson structure***. The classical ***Poisson bracket*** (consistent with what we defined in Chapter 1) is defined by

$$\{F, G\} = B(\mathbf{d}F, \mathbf{d}G) = \mathbf{d}F \cdot \mathbb{J}\nabla G. \qquad (2.1.15)$$

The ***symplectic structure*** Ω is the bilinear form associated to $R^{-1} : Z \to Z^*$, that is, $\Omega(v, w) = \left\langle R^{-1}(v), w \right\rangle$, or, equivalently, $\Omega^\flat = R^{-1}$. The matrix of Ω is $\mathbb{J}$ in the sense that

$$\Omega(v, w) = v^T \mathbb{J} w. \qquad (2.1.16)$$

To unify notation we shall sometimes write

Ω	for the symplectic form,	$Z \times Z \to \mathbb{R}$	with matrix $\mathbb{J}$,
$\Omega^\flat$	for the associated linear map,	$Z \to Z^*$	with matrix $\mathbb{J}^T$,
$\Omega^\sharp$	for the inverse map $(\Omega^\flat)^{-1} = R$,	$Z^* \to Z$	with matrix $\mathbb{J}$,
B	for the Poisson form,	$Z^* \times Z^* \to \mathbb{R}$	with matrix $\mathbb{J}$.

Hamilton's equations may be written

$$\dot{z} = X_H(z) = \Omega^\sharp \, \mathbf{d}H(z). \qquad (2.1.17)$$

Multiplying both sides by $\Omega^\flat$, we get

$$\Omega^\flat X_H(z) = \mathbf{d}H(z). \qquad (2.1.18)$$

In terms of the symplectic form, (2.1.18) reads

$$\Omega(X_H(z), v) = \mathbf{d}H(z) \cdot v \qquad (2.1.19)$$

for all $z, v \in Z$.

Problems such as rigid-body dynamics, quantum mechanics as a Hamiltonian system, and the motion of a particle in a rotating reference frame motivate the need to generalize these concepts. We shall do this in subsequent chapters and deal with both symplectic and Poisson structures in due course.

Exercises

◇ **2.1-1.** Writing $z = q + ip$, show that Hamilton's equations are equivalent to

$$\dot{z} = -2i \frac{\partial H}{\partial \overline{z}}.$$

Give a plausible definition of the right-hand side as part of your answer (or consult a book on complex variables theory).

◇ **2.1-2.** Write the harmonic oscillator $m\ddot{x} + kx = 0$ in the form of Euler–Lagrange equations, as Hamilton's equations, and finally, in the complex form (2.1.10).

◇ **2.1-3.** Repeat Exercise 2.1-2 for the nonlinear oscillator $m\ddot{x} + kx + \alpha x^3 = 0$.

2.2 Symplectic Forms on Vector Spaces

Let Z be a real Banach space, possibly infinite-dimensional, and let $\Omega : Z \times Z \to \mathbb{R}$ be a continuous bilinear form on Z. The form Ω is said to be ***nondegenerate*** (***or weakly nondegenerate***) if $\Omega(z_1, z_2) = 0$ for all $z_2 \in Z$ implies $z_1 = 0$. As in §2.1, the induced continuous linear mapping $\Omega^\flat : Z \to Z^*$ is defined by

$$\Omega^\flat(z_1)(z_2) = \Omega(z_1, z_2). \tag{2.2.1}$$

Nondegeneracy of Ω is equivalent to injectivity of $\Omega^\flat$, that is, to the condition "$\Omega^\flat(z) = 0$ implies $z = 0$." The form Ω is said to be ***strongly nondegenerate*** if $\Omega^\flat$ is an isomorphism, that is, $\Omega^\flat$ is onto as well as being injective. The open mapping theorem guarantees that if Z is a Banach space and $\Omega^\flat$ is one-to-one and onto, then its inverse is continuous. In most of the infinite-dimensional examples discussed in this book Ω will be only (weakly) nondegenerate.

A linear map between finite-dimensional spaces of the same dimension is one-to-one if and only if it is onto. Hence, *when Z is finite-dimensional, weak nondegeneracy and strong nondegeneracy are equivalent*. If Z is finite-dimensional, the matrix elements of Ω relative to a basis $\{e_I\}$ are defined by

$$\Omega_{IJ} = \Omega(e_I, e_J).$$

If $\{e^J\}$ denotes the basis for Z^* that is dual to $\{e_I\}$, that is, $\langle e^J, e_I \rangle = \delta_I^J$, and if we write $z = z^I e_I$ and $w = w^I e_I$, then

$$\Omega(z, w) = z^I \Omega_{IJ} w^J \quad \text{(sum over } I, J).$$

Since the matrix of $\Omega^\flat$ relative to the bases $\{e_I\}$ and $\{e^J\}$ equals the transpose of the matrix of Ω relative to $\{e_I\}$, that is $(\Omega^\flat)_{JI} = \Omega_{IJ}$, nondegeneracy is equivalent to $\det[\Omega_{IJ}] \neq 0$. In particular, if Ω is skew and nondegenerate, then Z is even-dimensional, since the determinant of a skew-symmetric matrix with an odd number of rows (and columns) is zero.

Definition 2.2.1. *A **symplectic form** Ω on a vector space Z is a nondegenerate skew-symmetric bilinear form on Z. The pair (Z, Ω) is called a **symplectic vector space**. If Ω is strongly nondegenerate, (Z, Ω) is called a **strong symplectic vector space**.*

Examples

We now develop some basic examples of symplectic forms.

(a) Canonical Forms. Let W be a vector space, and let $Z = W \times W^*$. Define the ***canonical symplectic form*** Ω on Z by

$$\Omega((w_1, \alpha_1), (w_2, \alpha_2)) = \alpha_2(w_1) - \alpha_1(w_2), \tag{2.2.2}$$

where $w_1, w_2 \in W$ and $\alpha_1, \alpha_2 \in W^*$.

More generally, let W and W' be two vector spaces in duality, that is, there is a weakly nondegenerate pairing $\langle \, , \rangle : W' \times W \to \mathbb{R}$. Then on $W \times W'$,

$$\Omega((w_1, \alpha_1), (w_2, \alpha_2)) = \langle \alpha_2, w_1 \rangle - \langle \alpha_1, w_2 \rangle \tag{2.2.3}$$

is a weak symplectic form. ♦

(b) The Space of Functions. Let $\mathcal{F}(\mathbb{R}^3)$ be the space of smooth functions $\varphi : \mathbb{R}^3 \to \mathbb{R}$, and let $\mathrm{Den}_c(\mathbb{R}^3)$ be the space of smooth densities on $\mathbb{R}^3$ with compact support. We write a density $\pi \in \mathrm{Den}_c(\mathbb{R}^3)$ as a function $\pi' \in \mathcal{F}(\mathbb{R}^3)$ with compact support times the volume element d^3x on $\mathbb{R}^3$ as $\pi = \pi' \, d^3x$. The spaces $\mathcal{F}$ and Den_c are in weak nondegenerate duality by the pairing $\langle \varphi, \pi \rangle = \int \varphi \pi' \, d^3x$. Therefore, from (2.2.3) we get the symplectic form Ω on the vector space $Z = \mathcal{F}(\mathbb{R}^3) \times \mathrm{Den}_c(\mathbb{R}^3)$:

$$\Omega((\varphi_1, \pi_1), (\varphi_2, \pi_2)) = \int_{\mathbb{R}^3} \varphi_1 \pi_2 - \int_{\mathbb{R}^3} \varphi_2 \pi_1. \tag{2.2.4}$$

We choose densities with compact support so that the integrals in this formula will be finite. Other choices of spaces could be used as well. ♦

(c) Finite-Dimensional Canonical Form. Suppose that W is a real vector space of dimension n. Let $\{e_i\}$ be a basis of W, and let $\{e^i\}$ be the dual basis of W^*. With $Z = W \times W^*$ and defining $\Omega : Z \times Z \to \mathbb{R}$ as in (2.2.2), one computes that the matrix of Ω in the basis

$$\{(e_1, 0), \ldots, (e_n, 0), (0, e^1), \ldots, (0, e^n)\}$$

is

$$\mathbb{J} = \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{bmatrix}, \tag{2.2.5}$$

where $\mathbf{1}$ and $\mathbf{0}$ are the $n \times n$ identity and zero matrices. ♦

(d) Symplectic Form Associated to an Inner Product Space. If $(W, \langle \, , \rangle)$ is a real inner product space, W is in duality with itself, so we obtain a symplectic form on $Z = W \times W$ from (2.2.3):

$$\Omega((w_1, w_2), (z_1, z_2)) = \langle z_2, w_1 \rangle - \langle z_1, w_2 \rangle . \tag{2.2.6}$$

As a special case of (2.2.6), let $W = \mathbb{R}^3$ with the usual inner product

$$\langle \mathbf{q}, \mathbf{v} \rangle = \mathbf{q} \cdot \mathbf{v} = \sum_{i=1}^{3} q^i v^i.$$

The corresponding symplectic form on $\mathbb{R}^6$ is given by

$$\Omega((\mathbf{q}_1, \mathbf{v}_1), (\mathbf{q}_2, \mathbf{v}_2)) = \mathbf{v}_2 \cdot \mathbf{q}_1 - \mathbf{v}_1 \cdot \mathbf{q}_2, \tag{2.2.7}$$

where $\mathbf{q}_1, \mathbf{q}_2, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$. This coincides with Ω defined in Example (c) for $W = \mathbb{R}^3$, provided that $\mathbb{R}^3$ is identified with $(\mathbb{R}^3)^*$. ♦

Bringing Ω to canonical form using elementary linear algebra results in the following statement. *If (Z, Ω) is a p-dimensional symplectic vector space, then p is even. Furthermore, Z is, as a vector space, isomorphic to one of the standard examples, namely $W \times W^*$, and there is a basis of W in which the matrix of Ω is $\mathbb{J}$. Such a basis is called* ***canonical***, *as are the corresponding coordinates.* See Exercise 2.2-3.

(e) Symplectic Form on $\mathbb{C}^n$. Write elements of complex n-space $\mathbb{C}^n$ as n-tuples $z = (z_1, \ldots, z_n)$ of complex numbers. The ***Hermitian inner product*** is

$$\langle z, w \rangle = \sum_{j=1}^{n} z_j \overline{w}_j = \sum_{j=1}^{n} (x_j u_j + y_j v_j) + i \sum_{j=1}^{n} (u_j y_j - v_j x_j),$$

where $z_j = x_j + i y_j$ and $w_j = u_j + i v_j$. Thus, $\operatorname{Re} \langle z, w \rangle$ is the real inner product and $-\operatorname{Im} \langle z, w \rangle$ is the symplectic form if $\mathbb{C}^n$ is identified with $\mathbb{R}^n \times \mathbb{R}^n$. ♦

(f) Quantum-Mechanical Symplectic Form. We now discuss an interesting symplectic vector space that arises in quantum mechanics, as we shall further explain in Chapter 3. Recall that a ***Hermitian inner product*** $\langle \, , \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ on a complex Hilbert space $\mathcal{H}$ is linear in its first argument and antilinear in its second, and $\langle \psi_1, \psi_2 \rangle$ is the complex conjugate of $\langle \psi_2, \psi_1 \rangle$, where $\psi_1, \psi_2 \in \mathcal{H}$.

Set

$$\Omega(\psi_1, \psi_2) = -2\hbar \operatorname{Im} \langle \psi_1, \psi_2 \rangle,$$

where $\hbar$ is Planck's constant. One checks that Ω is a strong symplectic form on $\mathcal{H}$.

There is another view of this symplectic form motivated by the preceding Example (d) that is interesting. Let $\mathcal{H}$ be the complexification of a real Hilbert space H, so the complex Hilbert space $\mathcal{H}$ is identified with $H \times H$, and the Hermitian inner product is given by

$$\langle (u_1, u_2), (v_1, v_2) \rangle = \langle u_1, v_1 \rangle + \langle u_2, v_2 \rangle + i(\langle u_2, v_1 \rangle - \langle u_1, v_2 \rangle).$$

The imaginary part of this form coincides with that in (2.2.6).

There is yet another view related to the interpretation of a wave function ψ and its conjugate $\bar{\psi}$ being conjugate variables. Namely, we consider the embedding of $\mathcal{H}$ into $\mathcal{H} \times \mathcal{H}^*$ via $\psi \mapsto (i\psi, \psi)$. Then one checks that the restriction of $\hbar$ times the canonical symplectic form (2.2.6) on $\mathcal{H} \times \mathcal{H}^*$, namely,

$$((\psi_1, \varphi_1), (\psi_2, \varphi_2)) \mapsto \hbar \operatorname{Re}[\langle \varphi_2, \psi_1 \rangle - \langle \varphi_1, \psi_2 \rangle],$$

coincides with Ω. ♦

Exercises

◇ **2.2-1.** Verify that the formula for the symplectic form for $\mathbb{R}^{2n}$ as a matrix, namely,

$$\mathbb{J} = \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{bmatrix},$$

coincides with the definition of the symplectic form as the canonical form on $\mathbb{R}^{2n}$ regarded as the product $\mathbb{R}^n \times (\mathbb{R}^n)^*$.

◇ **2.2-2.** Let (Z, Ω) be a finite-dimensional symplectic vector space and let $V \subset Z$ be a linear subspace. Assume that V is symplectic; that is, Ω restricted to $V \times V$ is nondegenerate. Let

$$V^\Omega = \{\, z \in Z \mid \Omega(z, v) = 0 \text{ for all } v \in V \,\}.$$

Show that V^Ω is symplectic and $Z = V \oplus V^\Omega$.

◇ **2.2-3.** Find a canonical basis for a symplectic form Ω on Z as follows. Let $e_1 \in Z$, $e_1 \neq 0$. Find $e_2 \in Z$ with $\Omega(e_1, e_2) \neq 0$. By rescaling e_2, assume $\Omega(e_1, e_2) = 1$. Let V be the span of e_1 and e_2. Apply Exercise 2.2-2 and repeat this construction on V^Ω.

◇ **2.2-4.** Let (Z, Ω) be a finite-dimensional symplectic vector space and $V \subset Z$ a subspace. Define V^Ω as in Exercise 2.2-2. Show that Z/V^Ω and V^* are isomorphic vector spaces.

2.3 Canonical Transformations, or Symplectic Maps

To motivate the definition of symplectic maps (synonymous with canonical transformations), start with Hamilton's equations

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}, \tag{2.3.1}$$

and a transformation $\varphi : Z \to Z$ of phase space to itself. Write

$$(\tilde{q}, \tilde{p}) = \varphi(q, p),$$

that is,

$$\tilde{z} = \varphi(z). \tag{2.3.2}$$

Assume that $z(t) = (q(t), p(t))$ satisfies Hamilton's equations, that is,

$$\dot{z}(t) = X_H(z(t)) = \Omega^\sharp \, \mathbf{d}H(z(t)), \tag{2.3.3}$$

where $\Omega^\sharp : Z^* \to Z$ is the linear map with matrix $\mathbb{J}$ whose entries we denote by B^{JK}. By the chain rule, $\tilde{z} = \varphi(z)$ satisfies

$$\dot{\tilde{z}}^I = \frac{\partial \varphi^I}{\partial z^J} \dot{z}^J =: A^I{}_J \dot{z}^J \tag{2.3.4}$$

(sum on J). Substituting (2.3.3) into (2.3.4), employing coordinate notation, and using the chain rule, we conclude that

$$\dot{\tilde{z}}^I = A^I{}_J B^{JK} \frac{\partial H}{\partial z^K} = A^I{}_J B^{JK} A^L{}_K \frac{\partial H}{\partial \tilde{z}^L}. \tag{2.3.5}$$

Thus, the equations (2.3.5) are Hamiltonian if and only if

$$A^I{}_J B^{JK} A^L{}_K = B^{IL}, \tag{2.3.6}$$

which in matrix notation reads

$$A \mathbb{J} A^T = \mathbb{J}. \tag{2.3.7}$$

In terms of composition of linear maps, (2.3.6) means

$$A \circ \Omega^\sharp \circ A^T = \Omega^\sharp, \tag{2.3.8}$$

since the matrix of $\Omega^\sharp$ in canonical coordinates is $\mathbb{J}$ (see §2.1). A transformation satisfying (2.3.6) is called a ***canonical transformation,*** a ***symplectic transformation***, or a ***Poisson transformation***.[2]

Taking determinants of (2.3.7) shows that $\det A = \pm 1$ (we will see in Chapter 9 that $\det A = 1$ is the only possibility) and in particular that A is invertible; taking the inverse of (2.3.8) gives

$$(A^T)^{-1} \circ \Omega^\flat \circ A^{-1} = \Omega^\flat,$$

that is,

$$A^T \circ \Omega^\flat \circ A = \Omega^\flat, \tag{2.3.9}$$

[2] In Chapter 10, where Poisson structures can be different from symplectic ones, we will see that (2.3.8) generalizes to the Poisson context.

which has the matrix form

$$A^T \mathbb{J} A = \mathbb{J}, \tag{2.3.10}$$

since the matrix of $\Omega^\flat$ in canonical coordinates is $-\mathbb{J}$ (see §2.1). Note that (2.3.7) and (2.3.10) are equivalent (the inverse of one gives the other). As bilinear forms, (2.3.9) reads

$$\Omega(\mathbf{D}\varphi(z) \cdot z_1, \mathbf{D}\varphi(z) \cdot z_2) = \Omega(z_1, z_2), \tag{2.3.11}$$

where $\mathbf{D}\varphi$ is the derivative of φ (the Jacobian matrix in finite dimensions). With (2.3.11) as a guideline, we next write the general condition for a map to be symplectic.

Definition 2.3.1. *If (Z, Ω) and (Y, Ξ) are symplectic vector spaces, a smooth map $f : Z \to Y$ is called **symplectic** or **canonical** if it preserves the symplectic forms, that is, if*

$$\Xi(\mathbf{D}f(z) \cdot z_1, \mathbf{D}f(z) \cdot z_2) = \Omega(z_1, z_2) \tag{2.3.12}$$

for all $z, z_1, z_2 \in Z$.

We next introduce some notation that will help us write (2.3.12) in a compact and efficient way.

Pull-Back Notation

We introduce a convenient notation for these sorts of transformations.

$\varphi^* f$ ***pull-back of a function***: $\varphi^* f = f \circ \varphi$.

$\varphi_* g$ ***push-forward of a function***: $\varphi_* g = g \circ \varphi^{-1}$.

$\varphi_* X$ ***push-forward of a vector field*** X by φ:

$$(\varphi_* X)(\varphi(z)) = \mathbf{D}\varphi(z) \cdot X(z);$$

in components,

$$(\varphi_* X)^I = \frac{\partial \varphi^I}{\partial z^J} X^J.$$

$\varphi^* Y$ ***pull-back of a vector field*** Y by φ: $\varphi^* Y = (\varphi^{-1})_* Y$

$\varphi^* \Omega$ ***pull-back of a bilinear form*** Ω on Z gives a bilinear form $\varphi^* \Omega$ depending on the point $z \in Z$:

$$(\varphi^* \Omega)_z(z_1, z_2) = \Omega(\mathbf{D}\varphi(z) \cdot z_1, \mathbf{D}\varphi(z) \cdot z_2);$$

in components,

$$(\varphi^*\Omega)_{IJ} = \frac{\partial \varphi^K}{\partial z^I}\frac{\partial \varphi^L}{\partial z^J}\Omega_{KL};$$

$\varphi_*\Xi$ ***push-forward of a bilinear form*** Ξ by φ equals the pull-back by the inverse: $\varphi_*\Xi = (\varphi^{-1})^*\Xi$.

In this pull-back notation, (2.3.12) reads $(f^*\Xi)_z = \Omega_z$, or $f^*\Xi = \Omega$ for short.

The Symplectic Group. It is simple to verify that if (Z, Ω) is a finite-dimensional symplectic vector space, the set of all linear symplectic mappings $T : Z \to Z$ forms a group under composition. It is called the ***symplectic group*** and is denoted by $\mathrm{Sp}(Z, \Omega)$. As we have seen, in a canonical basis, a matrix A is symplectic if and only if

$$A^T \mathbb{J} A = \mathbb{J}, \tag{2.3.13}$$

where A^T is the transpose of A. For $Z = W \times W^*$ and a canonical basis, if A has the matrix

$$A = \begin{bmatrix} A_{qq} & A_{qp} \\ A_{pq} & A_{pp} \end{bmatrix}, \tag{2.3.14}$$

then one checks (Exercise 2.3-2) that (2.3.13) *is equivalent to either of the following two conditions*:

(1) $A_{qq}A_{qp}^T$ and $A_{pp}A_{pq}^T$ are symmetric and $A_{qq}A_{pp}^T - A_{qp}A_{pq}^T = \mathbf{1}$;

(2) $A_{pq}^T A_{qq}$ and $A_{qp}^T A_{pp}$ are symmetric and $A_{qq}^T A_{pp} - A_{pq}^T A_{pq} = \mathbf{1}$.

In infinite dimensions $\mathrm{Sp}(Z, \Omega)$ is, by definition, the set of elements of $\mathrm{GL}(Z)$ (the group of invertible bounded linear operators of Z to Z) that leave Ω fixed.

Symplectic Orthogonal Complements. If (Z, Ω) is a (weak) symplectic space and E and F are subspaces of Z, we define

$$E^\Omega = \{\, z \in Z \mid \Omega(z, e) = 0 \text{ for all } e \in E \,\},$$

called the ***symplectic orthogonal complement*** of E. We leave it to the reader to check that

(i) E^Ω is closed;

(ii) $E \subset F$ implies $F^\Omega \subset E^\Omega$;

(iii) $E^\Omega \cap F^\Omega = (E + F)^\Omega$;

(iv) if Z is finite-dimensional, then $\dim E + \dim E^\Omega = \dim Z$ (to show this, use the fact that $E^\Omega = \ker(i^* \circ \Omega^\flat)$, where $i : E \to Z$ is the inclusion and $i^* : Z^* \to E^*$ is its dual, $i^*(\alpha) = \alpha \circ i$, which is surjective; alternatively, use Exercise 2.2-4);

(v) if Z is finite-dimensional, $E^{\Omega\Omega} = E$ (this is also true in infinite dimensions if E is closed); and

(vi) if E and F are closed, then $(E \cap F)^\Omega = E^\Omega + F^\Omega$ (to prove this use (iii) and (v)).

Exercises

◇ **2.3-1.** Show that a transformation $\varphi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is symplectic in the sense that its derivative matrix $A = \mathbf{D}\varphi(z)$ satisfies the condition $A^T \mathbb{J} A = \mathbb{J}$ if and only if the condition

$$\Omega(Az_1, Az_2) = \Omega(z_1, z_2)$$

holds for all $z_1, z_2 \in \mathbb{R}^{2n}$.

◇ **2.3-2.** Let $Z = W \times W^*$, let $A : Z \to Z$ be a linear transformation, and, using canonical coordinates, write the matrix of A as

$$A = \begin{bmatrix} A_{qq} & A_{qp} \\ A_{pq} & A_{pp} \end{bmatrix}.$$

Show that A being symplectic is equivalent to either of the two following conditions:

(i) $A_{qq}A_{qp}^T$ and $A_{pp}A_{pq}^T$ are symmetric and $A_{qq}A_{pp}^T - A_{qp}A_{pq}^T = \mathbf{1}$;

(ii) $A_{pq}^T A_{qq}$ and $A_{qp}^T A_{pp}$ are symmetric and $A_{qq}^T A_{pp} - A_{pq}^T A_{qp} = \mathbf{1}$. (Here $\mathbf{1}$ denotes the $n \times n$ identity.)

◇ **2.3-3.** Let f be a given function of $\mathbf{q} = (q^1, q^2, \ldots, q^n)$. Define the map $\varphi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ by $\varphi(\mathbf{q}, \mathbf{p}) = (\mathbf{q}, \mathbf{p} + \mathbf{d}f(\mathbf{q}))$. Show that φ is a canonical (symplectic) transformation.

◇ **2.3-4.**

(a) Let $A \in \mathrm{GL}(n, \mathbb{R})$ be an invertible linear transformation. Show that the map $\varphi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ given by $(\mathbf{q}, \mathbf{p}) \mapsto (A\mathbf{q}, (A^{-1})^T \mathbf{p})$ is a canonical transformation.

(b) If $\mathbf{R}$ is a rotation in $\mathbb{R}^3$, show that the map $(\mathbf{q}, \mathbf{p}) \mapsto (\mathbf{R}\mathbf{q}, \mathbf{R}\mathbf{p})$ is a canonical transformation.

◇ **2.3-5.** Let (Z, Ω) be a finite-dimensional symplectic vector space. A subspace $E \subset Z$ is called ***isotropic***, ***coisotropic***, and ***Lagrangian*** if $E \subset E^{\Omega}$, $E^{\Omega} \subset E$, and $E = E^{\Omega}$, respectively. Note that E is Lagrangian if and only if it is isotropic and coisotropic at the same time. Show that:

(a) An isotropic (coisotropic) subspace E is Lagrangian if and only if $\dim E = \dim E^{\Omega}$. In this case necessarily $2 \dim E = \dim Z$.

(b) Every isotropic (coisotropic) subspace is contained in (contains) a Lagrangian subspace.

(c) An isotropic (coisotropic) subspace is Lagrangian if and only if it is a maximal isotropic (minimal coisotropic) subspace.

2.4 The General Hamilton Equations

The concrete form of Hamilton's equations we have already encountered is a special case of a construction on symplectic spaces. Here, we discuss this formulation for systems whose phase space is linear; in subsequent sections we will generalize the setting to phase spaces that are symplectic manifolds and in Chapter 10 to spaces where only a Poisson bracket is given. These generalizations will all be important in our study of specific examples.

Definition 2.4.1. *Let (Z, Ω) be a symplectic vector space. A vector field $X : Z \to Z$ is called* ***Hamiltonian*** *if*

$$\Omega^{\flat}(X(z)) = \mathbf{d}H(z), \tag{2.4.1}$$

for all $z \in Z$, for some C^1 function $H : Z \to \mathbb{R}$. Here $\mathbf{d}H(z) = \mathbf{D}H(z)$ is alternative notation for the derivative of H. If such an H exists, we write $X = X_H$ and call H a ***Hamiltonian function****, or* ***energy function****, for the vector field X.*

In a number of important examples, especially infinite-dimensional ones, H need not be defined on all of Z. We shall briefly discuss in §3.3 some of the technicalities involved.

If Z is finite-dimensional, nondegeneracy of Ω implies that $\Omega^{\flat} : Z \to Z^*$ is an isomorphism, which guarantees that X_H exists for any given function H. However, if Z is infinite-dimensional and Ω is only weakly nondegenerate, we do not know *a priori* that X_H exists for a given H. If it does exist, it is unique, since $\Omega^{\flat}$ is one-to-one.

The set of Hamiltonian vector fields on Z is denoted by $\mathfrak{X}_{\mathrm{Ham}}(Z)$, or simply $\mathfrak{X}_{\mathrm{Ham}}$. Thus, $X_H \in \mathfrak{X}_{\mathrm{Ham}}$ is the vector field determined by the condition

$$\Omega(X_H(z), \delta z) = \mathbf{d}H(z) \cdot \delta z \quad \text{for all } z, \delta z \in Z. \tag{2.4.2}$$

If X is a vector field, the ***interior product*** $\mathbf{i}_X\Omega$ (also denoted by $X \,\lrcorner\, \Omega$) is defined to be the dual vector (also called, a ***one-form***) given at a point $z \in Z$ as follows:

$$(\mathbf{i}_X\Omega)_z \in Z^*; \quad (\mathbf{i}_X\Omega)_z(v) := \Omega(X(z), v),$$

for all $v \in Z$. Then condition (2.4.1) or (2.4.2) may be written as

$$\mathbf{i}_X\Omega = \mathbf{d}H; \quad \text{i.e.,} \quad X \,\lrcorner\, \Omega = \mathbf{d}H. \tag{2.4.3}$$

To express H in terms of X_H and Ω, we integrate the identity

$$\mathbf{d}H(tz) \cdot z = \Omega(X_H(tz), z)$$

from $t = 0$ to $t = 1$. The fundamental theorem of calculus gives

$$\begin{aligned} H(z) - H(0) &= \int_0^1 \frac{dH(tz)}{dt} dt = \int_0^1 \mathbf{d}H(tz) \cdot z \, dt \\ &= \int_0^1 \Omega(X_H(tz), z) \, dt. \end{aligned} \tag{2.4.4}$$

Let us now abstract the calculation we did in arriving at (2.3.7).

Proposition 2.4.2. *Let (Z, Ω) and (Y, Ξ) be symplectic vector spaces and $f : Z \to Y$ a diffeomorphism. Then f is a symplectic transformation if and only if for all Hamiltonian vector fields X_H on Y, we have $f_* X_{H \circ f} = X_H$, that is,*

$$\mathbf{D}f(z) \cdot X_{H\circ f}(z) = X_H(f(z)). \tag{2.4.5}$$

Proof. Note that for $v \in Z$,

$$\begin{aligned} \Omega(X_{H\circ f}(z), v) &= \mathbf{d}(H \circ f)(z) \cdot v = \mathbf{d}H(f(z)) \cdot \mathbf{D}f(z) \cdot v \\ &= \Xi(X_H(f(z)), \mathbf{D}f(z) \cdot v). \end{aligned} \tag{2.4.6}$$

If f is symplectic, then

$$\Xi(\mathbf{D}f(z) \cdot X_{H\circ f}(z), \mathbf{D}f(z) \cdot v) = \Omega(X_{H\circ f}(z), v),$$

and thus by nondegeneracy of Ξ and the fact that $\mathbf{D}f(z) \cdot v$ is an arbitrary element of Y (because f is a diffeomorphism and hence $\mathbf{D}f(z)$ is an isomorphism), (2.4.5) holds. Conversely, if (2.4.5) holds, then (2.4.6) implies

$$\Xi(\mathbf{D}f(z) \cdot X_{H\circ f}(z), \mathbf{D}f(z) \cdot v) = \Omega(X_{H\circ f}(z), v)$$

for any $v \in Z$ and any C^1 map $H : Y \to \mathbb{R}$. However, $X_{H\circ f}(z)$ equals an arbitrary element $w \in Z$ for a correct choice of the Hamiltonian function H, namely, $(H \circ f)(z) = \Omega(w, z)$. Thus, f is symplectic. ■

Definition 2.4.3. ***Hamilton's equations*** *for H is the system of differential equations defined by X_H. Letting $c : \mathbb{R} \to Z$ be a curve, they are the equations*

$$\frac{dc(t)}{dt} = X_H(c(t)). \tag{2.4.7}$$

The Classical Hamilton Equations. We now relate the abstract form (2.4.7) to the classical form of Hamilton's equations. In the following, an n-tuple $(q^1, \ldots, q^n)$ will be denoted simply by (q^i).

Proposition 2.4.4. *Suppose that (Z, Ω) is a $2n$-dimensional symplectic vector space, and let $(q^i, p_i) = (q^1, \ldots, q^n, p_1, \ldots, p_n)$ denote canonical coordinates, with respect to which Ω has matrix $\mathbb{J}$. Then in this coordinate system, $X_H : Z \to Z$ is given by*

$$X_H = \left(\frac{\partial H}{\partial p_i}, -\frac{\partial H}{\partial q^i} \right) = \mathbb{J} \cdot \nabla H. \tag{2.4.8}$$

Thus, Hamilton's equations in canonical coordinates are

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i}. \tag{2.4.9}$$

More generally, if $Z = V \times V'$, $\langle \cdot\,, \cdot \rangle : V \times V' \to \mathbb{R}$ is a weakly nondegenerate pairing, and $\Omega((e_1, \alpha_1), (e_2, \alpha_2)) = \langle \alpha_2, e_1 \rangle - \langle \alpha_1, e_2 \rangle$, then

$$X_H(e, \alpha) = \left(\frac{\delta H}{\delta \alpha}, -\frac{\delta H}{\delta e} \right), \tag{2.4.10}$$

where $\delta H / \delta \alpha \in V$ and $\delta H / \delta e \in V'$ are the ***partial functional derivatives*** *defined by*

$$\mathbf{D}_2 H(e, \alpha) \cdot \beta = \left\langle \beta, \frac{\delta H}{\delta \alpha} \right\rangle \tag{2.4.11}$$

for any $\beta \in V'$ and similarly for $\delta H / \delta e$; in (2.4.10) *it is assumed that the functional derivatives exist.*

Proof. If $(f, \beta) \in V \times V'$, then

$$\begin{aligned} \Omega\left(\left(\frac{\delta H}{\delta \alpha}, -\frac{\delta H}{\delta e} \right), (f, \beta) \right) &= \left\langle \beta, \frac{\delta H}{\delta \alpha} \right\rangle + \left\langle \frac{\delta H}{\delta e}, f \right\rangle \\ &= \mathbf{D}_2 H(e, \alpha) \cdot \beta + \mathbf{D}_1 H(e, \alpha) \cdot f \\ &= \langle \mathbf{d} H(e, \alpha), (f, \beta) \rangle . \end{aligned}$$ ■

Proposition 2.4.5 (Conservation of Energy)**.** *Let $c(t)$ be an integral curve of X_H. Then $H(c(t))$ is constant in t. If φ_t denotes the flow of X_H, that is, $\varphi_t(z)$ is the solution of (2.4.7) with initial conditions $z \in Z$, then $H \circ \varphi_t = H$.*

Proof. By the chain rule,

$$\begin{aligned} \frac{d}{dt} H(c(t)) &= \mathbf{d} H(c(t)) \cdot \frac{d}{dt} c(t) = \Omega\left(X_H(c(t)), \frac{d}{dt} c(t) \right) \\ &= \Omega\left(X_H(c(t)), X_H(c(t)) \right) = 0, \end{aligned}$$

where the final equality follows from the skew-symmetry of Ω. ■

Exercises

◇ **2.4-1.** Let the skew-symmetric bilinear form Ω on $\mathbb{R}^{2n}$ have the matrix

$$\begin{bmatrix} \mathbf{B} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{bmatrix},$$

where $\mathbf{B} = [B_{ij}]$ is a skew-symmetric $n \times n$ matrix, and $\mathbf{1}$ is the identity matrix.

(a) Show that Ω is nondegenerate and hence a symplectic form on $\mathbb{R}^{2n}$.

(b) Show that Hamilton's equations with respect to Ω are, in standard coordinates,

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i} - B_{ij}\frac{\partial H}{\partial p_j}.$$

2.5 When Are Equations Hamiltonian?

Having seen how to derive Hamilton's equations on (Z, Ω) given H, it is natural to consider the converse: When is a given set of equations

$$\frac{dz}{dt} = X(z), \tag{2.5.1}$$

where $X : Z \to Z$ is a given vector field, Hamilton's equations for some H? If X is linear, the answer is given by the following.

Proposition 2.5.1. *Let the vector field $A : Z \to Z$ be linear. Then A is Hamiltonian if and only if A is Ω-skew, that is,*

$$\Omega(Az_1, z_2) = -\Omega(z_1, Az_2)$$

for all $z_1, z_2 \in Z$. Furthermore, in this case one can take $H(z) = \frac{1}{2}\Omega(Az, z)$.

Proof. Differentiating the defining relation

$$\Omega(X_H(z), v) = \mathbf{d}H(z) \cdot v \tag{2.5.2}$$

with respect to z in the direction u and using bilinearity of Ω, one gets

$$\Omega(\mathbf{D}X_H(z) \cdot u, v) = \mathbf{D}^2 H(z)(v, u). \tag{2.5.3}$$

From this and the symmetry of the second partial derivatives, we get

$$\begin{aligned} \Omega(\mathbf{D}X_H(z) \cdot u, v) &= \mathbf{D}^2 H(z)(u, v) = \Omega(\mathbf{D}X_H(z) \cdot v, u) \\ &= -\Omega(u, \mathbf{D}X_H(z) \cdot v). \end{aligned} \tag{2.5.4}$$

If $A = X_H$ for some H, then $\mathbf{D}X_H(z) = A$, and (2.5.4) becomes $\Omega(Au, v) = -\Omega(u, Av)$; hence A is Ω-skew.

Conversely, suppose that A is Ω-skew. Defining $H(z) = \frac{1}{2}\Omega(Az, z)$, we claim that $A = X_H$. Indeed,

$$\begin{aligned}\mathbf{d}H(z) \cdot u &= \tfrac{1}{2}\Omega(Au, z) + \tfrac{1}{2}\Omega(Az, u)\\ &= -\tfrac{1}{2}\Omega(u, Az) + \tfrac{1}{2}\Omega(Az, u)\\ &= \tfrac{1}{2}\Omega(Az, u) + \tfrac{1}{2}\Omega(Az, u) = \Omega(Az, u).\end{aligned}$$ ■

In canonical coordinates, where Ω has matrix $\mathbb{J}$, Ω-skewness of A is equivalent to symmetry of the matrix $\mathbb{J}A$; that is, $\mathbb{J}A + A^T\mathbb{J} = 0$. The vector space of all linear transformations of Z satisfying this condition is denoted by $\mathfrak{sp}(Z, \Omega)$, and its elements are called ***infinitesimal symplectic transformations***. In canonical coordinates, if $Z = W \times W^*$ and if A has the matrix

$$A = \begin{bmatrix} A_{qq} & A_{qp} \\ A_{pq} & A_{pp} \end{bmatrix}, \tag{2.5.5}$$

then one checks that A *is infinitesimally symplectic if and only if* A_{qp} *and* A_{pq} *are both symmetric and* $A_{qq}^T + A_{pp} = \mathbf{0}$ (see Exercise 2.5-1).

In the complex linear case, we use Example (f) in §2.2 ($2\hbar$ times the negative imaginary part of a Hermitian inner product $\langle\, ,\rangle$ is the symplectic form) to arrive at the following.

Corollary 2.5.2. *Let $\mathcal{H}$ be a complex Hilbert space with Hermitian inner product $\langle\, ,\rangle$ and let $\Omega(\psi_1, \psi_2) = -2\hbar \operatorname{Im}\langle\psi_1, \psi_2\rangle$. Let $A : \mathcal{H} \to \mathcal{H}$ be a complex linear operator. There exists an $H : \mathcal{H} \to \mathbb{R}$ such that $A = X_H$ if and only if iA is symmetric or, equivalently, satisfies*

$$\langle iA\psi_1, \psi_2\rangle = \langle\psi_1, iA\psi_2\rangle\,. \tag{2.5.6}$$

In this case, H may be taken to be $H(\psi) = \hbar\langle iA\psi, \psi\rangle$. We let $H_{\rm op} = i\hbar A$, and thus Hamilton's equation $\dot{\psi} = A\psi$ becomes the ***Schrödinger equation***[3]

$$i\hbar\frac{\partial\psi}{\partial t} = H_{\rm op}\psi. \tag{2.5.7}$$

Proof. The operator A is Ω-skew if and only if the condition

$$\operatorname{Im}\langle A\psi_1, \psi_2\rangle = -\operatorname{Im}\langle\psi_1, A\psi_2\rangle$$

[3] Strictly speaking, equation (2.5.6) is required to hold only on the domain of the operator A, which need not be all of $\mathcal{H}$. We shall ignore these issues for simplicity. This example is continued in §2.6 and in §3.2.

holds for all $\psi_1, \psi_2 \in \mathcal{H}$. Replacing ψ_1 by $i\psi_1$ and using the relation $\operatorname{Im}(iz) = \operatorname{Re} z$, this condition is equivalent to $\operatorname{Re}\langle A\psi_1, \psi_2\rangle = -\operatorname{Re}\langle \psi_1, A\psi_2\rangle$. Since

$$\langle iA\psi_1, \psi_2\rangle = -\operatorname{Im}\langle A\psi_1, \psi_2\rangle + i\ \operatorname{Re}\langle A\psi_1, \psi_2\rangle \tag{2.5.8}$$

and

$$\langle \psi_1, iA\psi_2\rangle = +\operatorname{Im}\langle \psi_1, A\psi_2\rangle - i\ \operatorname{Re}\langle \psi_1, A\psi_2\rangle \tag{2.5.9}$$

we see that Ω-skewness of A is equivalent to iA being symmetric. Finally,

$$\hbar\langle iA\psi, \psi\rangle = \hbar \operatorname{Re} i\langle A\psi, \psi\rangle = -\hbar \operatorname{Im}\langle A\psi, \psi\rangle = \frac{1}{2}\Omega(A\psi, \psi),$$

and the corollary follows from Proposition 2.5.1. ■

For nonlinear differential equations, the analogue of Proposition 2.5.1 is the following.

Proposition 2.5.3. *Let $X : Z \to Z$ be a (smooth) vector field on a symplectic vector space (Z, Ω). Then $X = X_H$ for some $H : Z \to \mathbb{R}$ if and only if $\mathbf{D}X(z)$ is Ω-skew for all z.*

Proof. We have seen the "only if" part in the proof of Proposition 2.5.1. Conversely, if $\mathbf{D}X(z)$ is Ω-skew, define[4]

$$H(z) = \int_0^1 \Omega(X(tz), z)\, dt + \text{constant}; \tag{2.5.10}$$

we claim that $X = X_H$. Indeed,

$$\begin{aligned}
\mathbf{d}H(z)\cdot v &= \int_0^1 [\Omega(\mathbf{D}X(tz)\cdot tv, z) + \Omega(X(tz), v)]\, dt \\
&= \int_0^1 [\Omega(t\mathbf{D}X(tz)\cdot z, v) + \Omega(X(tz), v)]\, dt \\
&= \Omega\left(\int_0^1 [t\mathbf{D}X(tz)\cdot z + X(tz)]\, dt, v\right) \\
&= \Omega\left(\int_0^1 \frac{d}{dt}[tX(tz)]\, dt, v\right) = \Omega(X(z), v).
\end{aligned}$$

■

An interesting characterization of Hamiltonian vector fields involves the Cayley transform. Let (Z, Ω) be a symplectic vector space and $A : Z \to Z$ a

[4]Looking ahead to Chapter 4 on differential forms, one can check that (2.5.10) for H is reproduced by the proof of the Poincaré lemma applied to the one-form $\mathbf{i}_X\Omega$. That $\mathbf{D}X(z)$ is Ω-skew is equivalent to $\mathbf{d}(\mathbf{i}_X\Omega) = 0$.

linear transformation such that $I - A$ is invertible. Then A *is Hamiltonian if and only if its* ***Cayley transform*** $C = (I + A)(I - A)^{-1}$ *is symplectic*. See Exercise 2.5-2. For applications, see Laub and Meyer [1974], Paneitz [1981], Feng [1986], and Austin and Krishnaprasad [1993]. The Cayley transform is useful in some Hamiltonian numerical algorithms, as this last reference and Marsden [1992] show.

Exercises

◇ **2.5-1.** Let $Z = W \times W^*$ and use a canonical basis to write the matrix of the linear map $A : Z \to Z$ as

$$A = \begin{bmatrix} A_{qq} & A_{qp} \\ A_{pq} & A_{pp} \end{bmatrix}.$$

Show that A is infinitesimally symplectic, that is, $\mathbb{J}A + A^T\mathbb{J} = \mathbf{0}$, if and only if A_{qp} and A_{pq} are both symmetric and $A_{qq}^T + A_{pp} = \mathbf{0}$.

◇ **2.5-2.** Let (Z, Ω) be a symplectic vector space. Let $A : Z \to Z$ be a linear map and assume that $(I - A)$ is invertible. Show that A is Hamiltonian if and only if its Cayley transform

$$(I + A)(I - A)^{-1}$$

is symplectic. Give an example of a linear Hamiltonian vector field such that $(I - A)$ is not invertible.

◇ **2.5-3.** Suppose that (Z, Ω) is a finite-dimensional symplectic vector space and let $\varphi : Z \to Z$ be a linear symplectic map with $\det \varphi = 1$ (as mentioned in the text, this assumption is superfluous, as will be shown later). If λ is an eigenvalue of multiplicity k, then so is $1/\lambda$. Prove this using the characteristic polynomial of φ.

◇ **2.5-4.** Suppose that (Z, Ω) is a finite-dimensional symplectic vector space and let $A : Z \to Z$ be a Hamiltonian vector field.

(a) Show that the ***generalized kernel*** of A, defined to be the set

$$\{ z \in Z \mid A^k z = 0 \text{ for some integer } k \geq 1 \},$$

is a symplectic subspace.

(b) In general, the literal kernel $\ker A$ is not a symplectic subspace of (Z, Ω). Give a counter example.

2.6 Hamiltonian Flows

This subsection discusses flows of Hamiltonian vector fields a little further. The next subsection gives the abstract definition of the Poisson bracket, relates it to the classical definitions, and then shows how it may be used in describing the dynamics. Later on, Poisson brackets will play an increasingly important role.

Let X_H be a Hamiltonian vector field on a symplectic vector space (Z, Ω) with Hamiltonian $H : Z \to \mathbb{R}$. The ***flow*** of X_H is the collection of maps $\varphi_t : Z \to Z$ satisfying

$$\frac{d}{dt}\varphi_t(z) = X_H(\varphi_t(z)) \tag{2.6.1}$$

for each $z \in Z$ and real t and $\varphi_0(z) = z$. Here and in the following, all statements concerning the map $\varphi_t : Z \to Z$ are to be considered only for those z and t such that $\varphi_t(z)$ is defined, as determined by differential equations theory.

Linear Flows. First consider the case in which A is a (bounded) *linear* vector field. The flow of A may be written as $\varphi_t = e^{tA}$; that is, the solution of $dz/dt = Az$ with initial condition z_0 is given by $z(t) = \varphi_t(z_0) = e^{tA}z_0$.

Proposition 2.6.1. *The flow φ_t of a linear vector field $A : Z \to Z$ consists of (linear) canonical transformations if and only if A is Hamiltonian.*

Proof. For all $u, v \in Z$ we have

$$\begin{aligned}\frac{d}{dt}(\varphi_t^*\Omega)(u,v) &= \frac{d}{dt}\Omega(\varphi_t(u),\varphi_t(v))\\ &= \Omega\left(\frac{d}{dt}\varphi_t(u),\varphi_t(v)\right) + \Omega\left(\varphi_t(u),\frac{d}{dt}\varphi_t(v)\right)\\ &= \Omega(A\varphi_t(u),\varphi_t(v)) + \Omega(\varphi_t(u),A\varphi_t(v)).\end{aligned}$$

Therefore, A is Ω-skew, that is, A is Hamiltonian, if and only if each φ_t is a linear canonical transformation. ■

Nonlinear Flows. For nonlinear flows, there is a corresponding result.

Proposition 2.6.2. *The flow φ_t of a (nonlinear) Hamiltonian vector field X_H consists of canonical transformations. Conversely, if the flow of a vector field X consists of canonical transformations, then it is Hamiltonian.*

Proof. Let φ_t be the flow of a vector field X. By (2.6.1) and the chain rule,

$$\frac{d}{dt}[\mathbf{D}\varphi_t(z)\cdot v] = \mathbf{D}\left[\frac{d}{dt}\varphi_t(z)\right]\cdot v = \mathbf{D}X(\varphi_t(z))\cdot(\mathbf{D}\varphi_t(z)\cdot v),$$

which is called the ***first variation equation***. Using this, we get

$$\begin{aligned}\frac{d}{dt}\Omega(\mathbf{D}\varphi_t(z)\cdot u, \mathbf{D}\varphi_t(z)\cdot v) &= \Omega(\mathbf{D}X(\varphi_t(z))\cdot[\mathbf{D}\varphi_t(z)\cdot u], \mathbf{D}\varphi_t(z)\cdot v)\\ &\quad + \Omega(\mathbf{D}\varphi_t(z)\cdot u, \mathbf{D}X(\varphi_t(z))\cdot[\mathbf{D}\varphi_t(z)\cdot v]).\end{aligned}$$

If $X = X_H$, then $\mathbf{D}X_H(\varphi_t(z))$ is Ω-skew by Proposition 2.5.3, so

$$\Omega(\mathbf{D}\varphi_t(z) \cdot u, \mathbf{D}\varphi_t(z) \cdot v) = \text{constant}.$$

At $t = 0$ this equals $\Omega(u, v)$, so $\varphi_t^* \Omega = \Omega$. Conversely, if φ_t is canonical, this calculation shows that $\mathbf{D}X(\varphi_t(z))$ is Ω-skew, whence by Proposition 2.5.3, $X = X_H$ for some H. ■

Later on, we give another proof of Proposition 2.6.2 using differential forms.

Example: The Schrödinger Equation

Recall that if $\mathcal{H}$ is a complex Hilbert space, a complex linear map $U : \mathcal{H} \to \mathcal{H}$ is called ***unitary*** if it preserves the Hermitian inner product.

Proposition 2.6.3. *Let $A : \mathcal{H} \to \mathcal{H}$ be a complex linear map on a complex Hilbert space $\mathcal{H}$. The flow φ_t of A is canonical, that is, consists of canonical transformations with respect to the symplectic form Ω defined in Example* (f) *of* §2.2*, if and only if φ_t is unitary.*

Proof. By definition,

$$\Omega(\psi_1, \psi_2) = -2\hbar \operatorname{Im} \langle \psi_1, \psi_2 \rangle ,$$

so

$$\Omega(\varphi_t \psi_1, \varphi_t \psi_2) = -2\hbar \operatorname{Im} \langle \varphi_t \psi_1, \varphi_t \psi_2 \rangle$$

for $\psi_1, \psi_2 \in \mathcal{H}$. Thus, φ_t is canonical if and only if $\operatorname{Im} \langle \varphi_t \psi_1, \varphi_t \psi_2 \rangle = \operatorname{Im} \langle \psi_1, \psi_2 \rangle$, and this in turn is equivalent to unitarity by complex linearity of φ_t, since $\langle \psi_1, \psi_2 \rangle = -\operatorname{Im} \langle i\psi_1, \psi_2 \rangle + i \operatorname{Im} \langle \psi_1, \psi_2 \rangle$. ■

This shows that the flow of the ***Schrödinger equation*** $\dot{\psi} = A\psi$ is canonical and unitary and so preserves the probability amplitude of any wave function that is a solution. In other words, we have

$$\langle \varphi_t \psi, \varphi_t \psi \rangle = \langle \psi, \psi \rangle ,$$

where φ_t is the flow of A. Later we shall see how this conservation of the norm also results from a symmetry-induced conservation law.

2.7 Poisson Brackets

Definition 2.7.1. *Given a symplectic vector space (Z, Ω) and two functions $F, G : Z \to \mathbb{R}$, the **Poisson bracket** $\{F, G\} : Z \to \mathbb{R}$ of F and G is defined by*

$$\{F, G\}(z) = \Omega(X_F(z), X_G(z)). \tag{2.7.1}$$

Using the definition of a Hamiltonian vector field, we find that equivalent expressions are

$$\{F, G\}(z) = \mathbf{d}F(z) \cdot X_G(z) = -\mathbf{d}G(z) \cdot X_F(z). \tag{2.7.2}$$

In (2.7.2) we write $\pounds_{X_G} F = \mathbf{d}F \cdot X_G$ for the derivative of F in the direction X_G.

Lie Derivative Notation. The ***Lie derivative*** of f along X, $\pounds_X f = \mathbf{d}f \cdot X$, is the ***directional derivative*** of f in the direction X. In coordinates it is given by

$$\pounds_X f = \frac{\partial f}{\partial z^I} X^I \quad \text{(sum on } I\text{)}.$$

Functions F, G such that $\{F, G\} = 0$ are said to be in ***involution*** or to ***Poisson commute***.

Examples

Now we turn to some examples of Poisson brackets.

(a) Canonical Bracket. Suppose that Z is $2n$-dimensional. Then in canonical coordinates $(q^1, \ldots, q^n, p_1, \ldots, p_n)$ we have

$$\begin{aligned} \{F, G\} &= \left[\frac{\partial F}{\partial p_i}, -\frac{\partial F}{\partial q^i} \right] \mathbb{J} \begin{bmatrix} \dfrac{\partial G}{\partial p_i} \\ -\dfrac{\partial G}{\partial q^i} \end{bmatrix} \\ &= \frac{\partial F}{\partial q^i} \frac{\partial G}{\partial p_i} - \frac{\partial F}{\partial p_i} \frac{\partial G}{\partial q^i} \qquad \text{(sum on } i\text{)}. \end{aligned} \tag{2.7.3}$$

From this we get the ***fundamental Poisson brackets***

$$\{q^i, q^j\} = 0, \quad \{p_i, p_j\} = 0, \quad \text{and} \quad \{q^i, p_j\} = \delta^i_j. \tag{2.7.4}$$

In terms of the Poisson structure, that is, the bilinear form B from §2.1, the Poisson bracket takes the form

$$\{F, G\} = B(\mathbf{d}F, \mathbf{d}G). \tag{2.7.5}$$

♦

(b) The Space of Functions. Let (Z, Ω) be defined as in Example (b) of §2.2 and let $F, G : Z \to \mathbb{R}$. Using equations (2.4.10) and (2.7.1) above, we get

$$\begin{aligned}\{F, G\} = \Omega(X_F, X_G) &= \Omega\left(\left(\frac{\delta F}{\delta \pi}, -\frac{\delta F}{\delta \varphi}\right), \left(\frac{\delta G}{\delta \pi}, -\frac{\delta G}{\delta \varphi}\right)\right) \\ &= \int_{\mathbb{R}^3}\left(\frac{\delta G}{\delta \pi}\frac{\delta F}{\delta \varphi} - \frac{\delta F}{\delta \pi}\frac{\delta G}{\delta \varphi}\right) d^3x. \end{aligned} \tag{2.7.6}$$

This example will be used in the next chapter when we study classical field theory. ♦

The Jacobi–Lie Bracket. The ***Jacobi–Lie bracket*** $[X, Y]$ of two vector fields X and Y on a vector space Z is defined by demanding that

$$\mathbf{d}f \cdot [X, Y] = \mathbf{d}(\mathbf{d}f \cdot Y) \cdot X - \mathbf{d}(\mathbf{d}f \cdot X) \cdot Y$$

for all real-valued functions f. In Lie derivative notation, this reads

$$\pounds_{[X,Y]} f = \pounds_X \pounds_Y f - \pounds_Y \pounds_X f.$$

One checks that this condition becomes, in vector analysis notation,

$$[X, Y] = (X \cdot \nabla)Y - (Y \cdot \nabla)X,$$

and in coordinates,

$$[X, Y]^J = X^I \frac{\partial}{\partial z^I} Y^J - Y^I \frac{\partial}{\partial z^I} X^J.$$

Proposition 2.7.2. *Let $[\,,]$ denote the Jacobi–Lie bracket of vector fields, and let $F, G \in \mathcal{F}(Z)$. Then*

$$X_{\{F,G\}} = -[X_F, X_G]. \tag{2.7.7}$$

Proof. We calculate as follows:

$$\begin{aligned}\Omega(X_{\{F,G\}}(z), u) &= \mathbf{d}\{F, G\}(z) \cdot u = \mathbf{d}(\Omega(X_F(z), X_G(z))) \cdot u \\ &= \Omega(\mathbf{D}X_F(z) \cdot u, X_G(z)) + \Omega(X_F(z), \mathbf{D}X_G(z) \cdot u) \\ &= \Omega(\mathbf{D}X_F(z) \cdot X_G(z), u) - \Omega(\mathbf{D}X_G(z) \cdot X_F(z), u) \\ &= \Omega(\mathbf{D}X_F(z) \cdot X_G(z) - \mathbf{D}X_G(z) \cdot X_F(z), u) \\ &= \Omega(-[X_F, X_G](z), u).\end{aligned}$$

Weak nondegeneracy of Ω implies the result. ■

Jacobi's Identity. We are now ready to prove the Jacobi identity in a fairly general context.

Proposition 2.7.3. *Let (Z, Ω) be a symplectic vector space. Then the Poisson bracket $\{\,,\}: \mathcal{F}(Z) \times \mathcal{F}(Z) \to \mathcal{F}(Z)$ makes $\mathcal{F}(Z)$ into a* ***Lie algebra****. That is, this bracket is real bilinear, skew-symmetric, and satisfies* ***Jacobi's identity****, that is,*

$$\{F, \{G, H\}\} + \{G, \{H, F\}\} + \{H, \{F, G\}\} = 0.$$

Proof. To verify Jacobi's identity note that for $F, G, H : Z \to \mathbb{R}$, we have

$$\begin{aligned} \{F, \{G, H\}\} &= -\pounds_{X_F}\{G, H\} = \pounds_{X_F}\pounds_{X_G} H, \\ \{G, \{H, F\}\} &= -\pounds_{X_G}\{H, F\} = -\pounds_{X_G}\pounds_{X_F} H, \end{aligned}$$

and

$$\{H, \{F, G\}\} = \pounds_{X_{\{F,G\}}} H,$$

so that

$$\{F, \{G, H\}\} + \{G, \{H, F\}\} + \{H, \{F, G\}\} = \pounds_{X_{\{F,G\}}} H + \pounds_{[X_F, X_G]} H.$$

The result thus follows by (2.7.7). ■

From Proposition 2.7.2 we see that the Jacobi–Lie bracket of two Hamiltonian vector fields is again Hamiltonian. Thus, we obtain the following corollary.

Corollary 2.7.4. *The set of Hamiltonian vector fields $\mathfrak{X}_{\mathrm{Ham}}(Z)$ forms a Lie subalgebra of $\mathfrak{X}(Z)$.*

Next, we characterize symplectic maps in terms of brackets.

Proposition 2.7.5. *Let $\varphi : Z \to Z$ be a diffeomorphism. Then φ is symplectic if and only if it preserves Poisson brackets, that is,*

$$\{\varphi^* F, \varphi^* G\} = \varphi^*\{F, G\} \tag{2.7.8}$$

for all $F, G : Z \to \mathbb{R}$.

Proof. We use the identity

$$\varphi^*(\pounds_X f) = \pounds_{\varphi^* X}(\varphi^* f),$$

which follows from the chain rule. Thus,

$$\varphi^*\{F, G\} = \varphi^* \pounds_{X_G} F = \pounds_{\varphi^* X_G}(\varphi^* F)$$

and

$$\{\varphi^* F, \varphi^* G\} = \pounds_{X_{G \circ \varphi}}(\varphi^* F).$$

Thus, φ preserves Poisson brackets if and only if $\varphi^* X_G = X_{G \circ \varphi}$ for every $G : Z \to \mathbb{R}$, that is, if and only if φ is symplectic by Proposition 2.4.2. ■

Proposition 2.7.6. *Let X_H be a Hamiltonian vector field on Z, with Hamiltonian H and flow φ_t. Then for $F : Z \to \mathbb{R}$,*

$$\begin{aligned} \frac{d}{dt}(F \circ \varphi_t) &= \{F \circ \varphi_t, H\} \\ &= \{F, H\} \circ \varphi_t. \end{aligned} \tag{2.7.9}$$

Proof. By the chain rule and the definition of X_F,

$$\begin{aligned} \frac{d}{dt}[(F \circ \varphi_t)(z)] &= \mathbf{d}F(\varphi_t(z)) \cdot X_H(\varphi_t(z)) \\ &= \Omega(X_F(\varphi_t(z)), X_H(\varphi_t(z))) \\ &= \{F, H\}(\varphi_t(z)). \end{aligned}$$

By Proposition 2.6.2 and (2.7.8), this equals

$$\{F \circ \varphi_t, H \circ \varphi_t\}(z) = \{F \circ \varphi_t, H\}(z)$$

by conservation of energy. ■

Corollary 2.7.7. *Let $F, G : Z \to \mathbb{R}$. Then F is constant along integral curves of X_G if and only if G is constant along integral curves of X_F, and this is true if and only if $\{F, G\} = 0$.*

Proposition 2.7.8. *Let $A, B : Z \to Z$ be linear Hamiltonian vector fields with corresponding energy functions*

$$H_A(z) = \tfrac{1}{2}\Omega(Az, z) \quad \textit{and} \quad H_B(z) = \tfrac{1}{2}\Omega(Bz, z).$$

Letting

$$[A, B] = A \circ B - B \circ A$$

be the operator commutator, *we have*

$$\{H_A, H_B\} = H_{[A,B]}. \tag{2.7.10}$$

Proof. By definition, $X_{H_A} = A$, and so

$$\{H_A, H_B\}(z) = \Omega(Az, Bz).$$

Since A and B are Ω-skew, we get

$$\begin{aligned} \{H_A, H_B\}(z) &= \tfrac{1}{2}\Omega(ABz, z) - \tfrac{1}{2}\Omega(BAz, z) \\ &= \tfrac{1}{2}\Omega([A, B]z, z) \qquad (2.7.11) \\ &= H_{[A,B]}(z). \end{aligned}$$ ■

2.8 A Particle in a Rotating Hoop

In this subsection we take a break from the abstract theory to do an example the "old-fashioned" way. This and other examples will also serve as excellent illustrations of the theory we are developing.

Derivation of the Equations. Consider a particle constrained to move on a circular hoop; for example a bead sliding in a Hula-Hoop. The particle is assumed to have mass m and to be acted on by gravitational and frictional forces, as well as constraint forces that keep it on the hoop. The hoop itself is spun about a vertical axis with constant angular velocity ω, as in Figure 2.8.1.

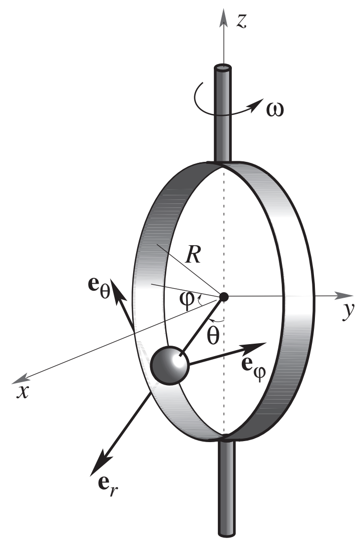

FIGURE 2.8.1. A particle moving in a hoop rotating with angular velocity ω.

The position of the particle in space is specified by the angles θ and φ, as shown in Figure 2.8.1. We can take $\varphi = \omega t$, so the position of the particle becomes determined by θ alone. Let the orthonormal frame along the coordinate directions $\mathbf{e}_\theta$, $\mathbf{e}_\varphi$, and $\mathbf{e}_r$ be as shown.

The forces acting on the particle are:

1. Friction, proportional to the velocity of the particle relative to the hoop: $-\nu R\dot{\theta}\mathbf{e}_\theta$, where $\nu \geq 0$ is a constant.[5]

---

[5]This is a "law of friction" that is more like a viscous fluid friction than a sliding friction in which ν is the ratio of the tangential force to the normal force; in any actual experimental setup (e.g., involving rolling spheres) a realistic modeling of the friction is not a trivial task; see, for example, Lewis and Murray [1995].

2. Gravity: $-mg\mathbf{k}$.

3. Constraint forces in the directions $\mathbf{e}_r$ and $\mathbf{e}_\varphi$ to keep the particle in the hoop.

The equations of motion are derived from Newton's second law $\mathbf{F} = m\mathbf{a}$. To get them, we need to calculate the acceleration $\mathbf{a}$; here $\mathbf{a}$ means the acceleration relative to the *fixed inertial frame* xyz in space; it does not mean $\ddot{\theta}$. Relative to this xyz coordinate system, we have

$$\begin{aligned} x &= R\sin\theta\cos\varphi, \\ y &= R\sin\theta\sin\varphi, \\ z &= -R\cos\theta. \end{aligned} \tag{2.8.1}$$

Calculating the second derivatives using $\varphi = \omega t$ and the chain rule gives

$$\begin{aligned} \ddot{x} &= -\omega^2 x - \dot{\theta}^2 x + (R\cos\theta\cos\varphi)\ddot{\theta} - 2R\omega\dot{\theta}\cos\theta\sin\varphi, \\ \ddot{y} &= -\omega^2 y - \dot{\theta}^2 y + (R\cos\theta\sin\varphi)\ddot{\theta} + 2R\omega\dot{\theta}\cos\theta\cos\varphi, \\ \ddot{z} &= -z\dot{\theta}^2 + (R\sin\theta)\ddot{\theta}. \end{aligned} \tag{2.8.2}$$

If $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, denote unit vectors along the x, y, and z axes, respectively, we have the easily verified relation

$$\mathbf{e}_\theta = (\cos\theta\cos\varphi)\mathbf{i} + (\cos\theta\sin\varphi)\mathbf{j} + \sin\theta\mathbf{k}. \tag{2.8.3}$$

Now consider the vector equation $\mathbf{F} = m\mathbf{a}$, where $\mathbf{F}$ is the sum of the three forces described earlier and

$$\mathbf{a} = \ddot{x}\mathbf{i} + \ddot{y}\mathbf{j} + \ddot{z}\mathbf{k}. \tag{2.8.4}$$

The $\mathbf{e}_\varphi$ and $\mathbf{e}_r$ components of $\mathbf{F} = m\mathbf{a}$ tell us only what the constraint forces must be; the equation of motion comes from the $\mathbf{e}_\theta$ component:

$$\mathbf{F}\cdot\mathbf{e}_\theta = m\mathbf{a}\cdot\mathbf{e}_\theta. \tag{2.8.5}$$

Using (2.8.3), the left side of (2.8.5) is

$$\mathbf{F}\cdot\mathbf{e}_\theta = -\nu R\dot{\theta} - mg\sin\theta, \tag{2.8.6}$$

while from (2.8.2), (2.8.3), and (2.8.4), the right side of (2.8.5) is

$$\begin{aligned} m\mathbf{a}\cdot\mathbf{e}_\theta &= m\{\ddot{x}\cos\theta\cos\varphi + \ddot{y}\cos\theta\sin\varphi + \ddot{z}\sin\theta\} \\ &= m\{\cos\theta\cos\varphi[-\omega^2 x - \dot{\theta}^2 x + (R\cos\theta\cos\varphi)\ddot{\theta} - 2R\omega\dot{\theta}\cos\theta\sin\varphi] \\ &\quad + \cos\theta\sin\varphi[-\omega^2 y - \dot{\theta}^2 y + (R\cos\theta\sin\varphi)\ddot{\theta} + 2R\omega\dot{\theta}\cos\theta\cos\varphi] \\ &\quad + \sin\theta[-z\dot{\theta}^2 + (R\sin\theta)\ddot{\theta}]\}. \end{aligned}$$

Using (2.8.1), this simplifies to

$$m\mathbf{a} \cdot \mathbf{e}_\theta = mR\{\ddot{\theta} - \omega^2 \sin\theta \cos\theta\}. \tag{2.8.7}$$

Comparing (2.8.5), (2.8.6), and (2.8.7), we get

$$\ddot{\theta} = \omega^2 \sin\theta \cos\theta - \frac{\nu}{m}\dot{\theta} - \frac{g}{R}\sin\theta \tag{2.8.8}$$

as our final equation of motion. Several remarks concerning it are in order:

(i) If $\omega = 0$ and $\nu = 0$, (2.8.8) reduces to the ***pendulum equation***

$$R\ddot{\theta} + g \sin\theta = 0.$$

In fact, our system can be viewed just as well as a ***whirling pendulum***.

(ii) For $\nu = 0$, (2.8.8) is Hamiltonian. This is readily verified using the variables $q = \theta$, $p = mR^2\dot{\theta}$, the canonical bracket structure

$$\{F, K\} = \frac{\partial F}{\partial q}\frac{\partial K}{\partial p} - \frac{\partial K}{\partial q}\frac{\partial F}{\partial p}, \tag{2.8.9}$$

and the Hamiltonian

$$H = \frac{p^2}{2mR^2} - mgR\cos\theta - \frac{mR^2\omega^2}{2}\sin^2\theta. \tag{2.8.10}$$

Derivation as Euler–Lagrange Equations. We now use Lagrangian methods to derive (2.8.8). In Figure 2.8.1, the velocity is

$$\mathbf{v} = R\dot{\theta}\mathbf{e}_\theta + (\omega R \sin\theta)\mathbf{e}_\varphi,$$

so the kinetic energy is

$$T = \tfrac{1}{2}m\|\mathbf{v}\|^2 = \tfrac{1}{2}m(R^2\dot{\theta}^2 + [\omega R \sin\theta]^2), \tag{2.8.11}$$

while the potential energy is

$$V = -mgR\cos\theta. \tag{2.8.12}$$

Thus, the Lagrangian is given by

$$L = T - V = \frac{1}{2}mR^2\dot{\theta}^2 + \frac{mR^2\omega^2}{2}\sin^2\theta + mgR\cos\theta, \tag{2.8.13}$$

and the Euler–Lagrange equations, namely,

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

(see §1.1 or §2.1), become

$$mR^2\ddot{\theta} = mR^2\omega^2 \sin\theta\cos\theta - mgR\sin\theta,$$

which are the same equations we derived by hand in (2.8.8) for $\nu = 0$. The Legendre transform gives $p = mR^2\dot{\theta}$ and the Hamiltonian (2.8.10). Notice that this Hamiltonian is *not* the kinetic plus potential energy of the particle. In fact, if one postulated this, then Hamilton's equations would give the *incorrect equations*. This has to do with deeper covariance properties of the Lagrangian versus Hamiltonian equations.

Equilibria. The ***equilibrium solutions*** are solutions satisfying $\dot{\theta} = 0$, $\ddot{\theta} = 0$; (2.8.8) gives

$$R\omega^2 \sin\theta\cos\theta = g\sin\theta. \tag{2.8.14}$$

Certainly, $\theta = 0$ and $\theta = \pi$ solve (2.8.14) corresponding to the particle at the bottom or top of the hoop. If $\theta \neq 0$ or π, (2.8.14) becomes

$$R\omega^2 \cos\theta = g, \tag{2.8.15}$$

which has two solutions when $g/(R\omega^2) < 1$. The value

$$\omega_c = \sqrt{\frac{g}{R}} \tag{2.8.16}$$

is the ***critical rotation rate***. Notice that ω_c is the frequency of linearized oscillations for the simple pendulum, that is, for the equation

$$R\ddot{\theta} + g\theta = 0.$$

For $\omega < \omega_c$ there are only *two* solutions $\theta = 0,\ \pi$, while for $\omega > \omega_c$ there are *four solutions*,

$$\theta = 0,\ \pi,\ \pm\cos^{-1}\left(\frac{g}{R\omega^2}\right). \tag{2.8.17}$$

We say that a ***bifurcation*** (or a ***Hamiltonian pitchfork bifurcation***, to be accurate) has occurred as ω crosses ω_c. We can see this graphically in computer-generated solutions of (2.8.8). Set $x = \theta$, $y = \dot{\theta}$ and rewrite (2.8.8) as

$$\begin{aligned} \dot{x} &= y, \\ \dot{y} &= \frac{g}{R}(\alpha\cos x - 1)\sin x - \beta y, \end{aligned} \tag{2.8.18}$$

where

$$\alpha = R\omega^2/g \quad \text{and} \quad \beta = \nu/m.$$

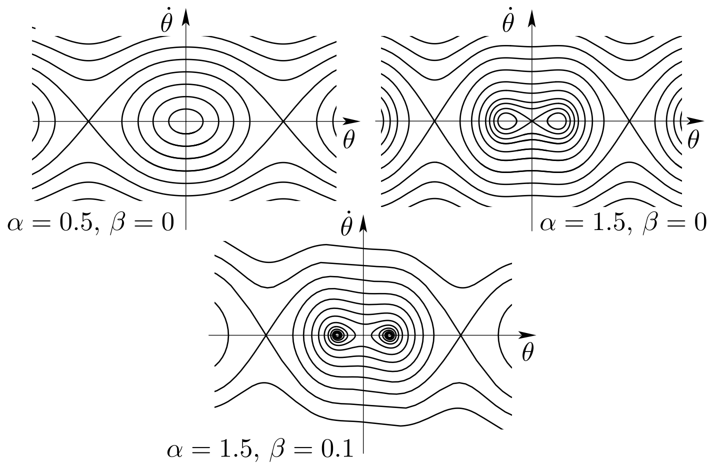

FIGURE 2.8.2. Phase portraits of the ball in the rotating hoop.

Taking $g = R$ for illustration, Figure 2.8.2 shows representative orbits in the phase portraits of (2.8.18) for various α, β.

This system with $\nu = 0$, that is, $\beta = 0$, is symmetric in the sense that the $\mathbb{Z}_2$-action given by

$$\theta \mapsto -\theta \quad \text{and} \quad \dot{\theta} \mapsto -\dot{\theta}$$

leaves the phase portrait invariant. If this $\mathbb{Z}_2$ symmetry is broken, by setting the rotation axis a little off center, for example, then one side gets preferred, as in Figure 2.8.3.

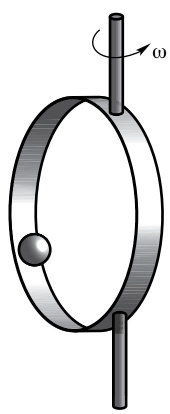

FIGURE 2.8.3. A ball in an off-center rotating hoop.

The evolution of the phase portrait for $\nu = 0$ is shown in Figure 2.8.4.

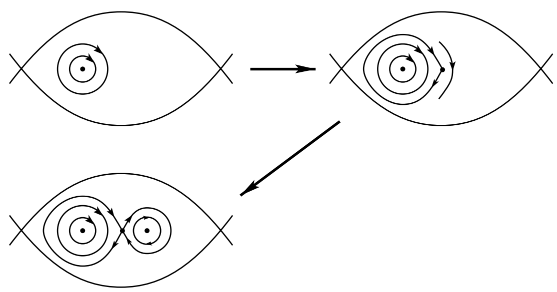

FIGURE 2.8.4. The phase portraits for the ball in the off-center hoop as the angular velocity increases.

Near $\theta = 0$, the potential function has changed from the symmetric bifurcation in Figure 2.8.5(a) to the unsymmetric one in Figure 2.8.5(b). This is what is known as the ***cusp catastrophe***; see Golubitsky and Schaeffer [1985] and Arnold [1968, 1984] for more information.

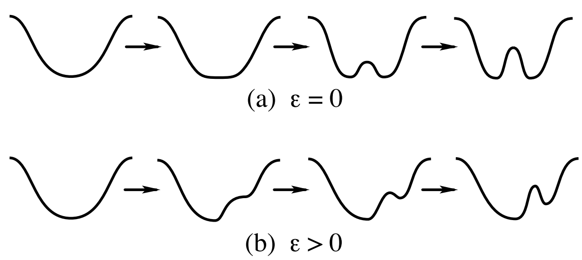

FIGURE 2.8.5. The evolution of the potential for the ball in the (a) centered and (b) off-center hoop as the angular velocity increases.

In (2.8.8), imagine that the hoop is subject to small periodic pulses, say $\omega = \omega_0 + \rho\cos(\eta t)$. Using the Melnikov method described in the introduction and in the following section, it is presumably true (but a messy calculation to prove) that the resulting time-periodic system has horseshoe chaos if ϵ and ν are small (where ϵ measures how off-center the hoop is) but ρ/ν exceeds a critical value. See Exercise 2.8-3 and §2.8.

Exercises

◇ **2.8-1.** Derive the equations of motion for a particle in a hoop spinning about a line a distance ϵ off center. What can you say about the equilibria as functions of ϵ and ω?

◇ **2.8-2.** Derive the formula of Exercise 1.9-1 for the homoclinic orbit (the orbit tending to the saddle point as $t \to \pm\infty$) of a pendulum $\ddot{\psi} + \sin\psi = 0$.

Do this using conservation of energy, determining the value of the energy on the homoclinic orbit, solving for $\dot{\psi}$, and then integrating.

◇ **2.8-3.** Using the method of the preceding exercise, derive an integral formula for the homoclinic orbit of the frictionless particle in a rotating hoop.

◇ **2.8-4.** Determine all equilibria of Duffing's equation

$$\ddot{x} - \beta x + \alpha x^3 = 0,$$

where α and β are positive constants, and study their stability. Derive a formula for the two homoclinic orbits.

◇ **2.8-5.** Determine the equations of motion and bifurcations for a ball in a light rotating hoop, but this time the hoop is not forced to rotate with constant *angular velocity*, but rather is free to rotate so that its *angular momentum* μ is conserved.

◇ **2.8-6.** Consider the pendulum shown in Figure 2.8.6. It is a planar pendulum whose suspension point is being whirled in a circle with angular velocity ω by means of a vertical shaft, as shown. The plane of the pendulum is orthogonal to the radial arm of length R. Ignore frictional effects.

(i) Using the notation in the figure, find the equations of motion of the pendulum.

(ii) Regarding ω as a parameter, show that a supercritical pitchfork bifurcation of equilibria occurs as the angular velocity of the shaft is increased.

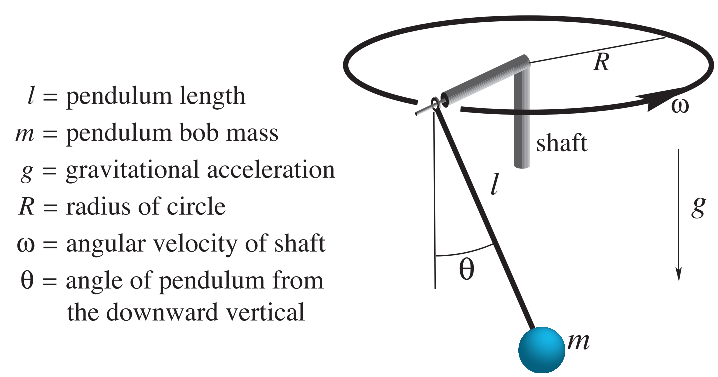

FIGURE 2.8.6. A whirling pendulum.

2.9 The Poincaré–Melnikov Method

Recall from the introduction that in the simplest version of the Poincaré–Melnikov method we are concerned with dynamical equations that perturb a planar Hamiltonian system

$$\dot{z} = X_0(z) \tag{2.9.1}$$

to one of the form

$$\dot{z} = X_0(z) + \epsilon X_1(z, t), \tag{2.9.2}$$

where ϵ is a small parameter, $z \in \mathbb{R}^2$, X_0 is a Hamiltonian vector field with energy H_0, X_1 is periodic with period T and is Hamiltonian with energy a T-periodic function H_1. We assume that X_0 has a homoclinic orbit $\overline{z}(t)$, that is, an orbit such that $\overline{z}(t) \to z_0$, a hyperbolic saddle point, as $t \to \pm\infty$. Define the ***Poincaré–Melnikov function*** by

$$M(t_0) = \int_{-\infty}^{\infty} \{H_0, H_1\}(\overline{z}(t - t_0), t)\, dt, \tag{2.9.3}$$

where $\{\,,\}$ denotes the Poisson bracket.

There are two convenient ways of visualizing the dynamics of (2.9.2). Introduce the ***Poincaré map*** $P^s_\epsilon : \mathbb{R}^2 \to \mathbb{R}^2$, which is the time T map for (2.9.2) starting at time s. For $\epsilon = 0$, the point z_0 and the homoclinic orbit are invariant under P^s_0, which is independent of s. The hyperbolic saddle z_0 persists as a nearby family of saddles z_ϵ for $\epsilon > 0$, small, and we are interested in whether or not the stable and unstable manifolds of the point z_ϵ for the map P^s_ϵ intersect transversally (if this holds for one s, it holds for all s). If so, we say that (2.9.2) has ***horseshoes*** *for* $\epsilon > 0$.

The second way to study (2.9.2) is to look directly at the suspended system on $\mathbb{R}^2 \times S^1$, where S^1 is the circle; (2.9.2) becomes the autonomous ***suspended system***

$$\begin{aligned} \dot{z} &= X_0(z) + \epsilon X_1(z, \theta), \\ \dot{\theta} &= 1. \end{aligned} \tag{2.9.4}$$

From this point of view, θ gets identified with time, and the curve

$$\gamma_0(t) = (z_0, t)$$

is a periodic orbit for (2.9.4). This orbit has ***stable manifolds*** and ***unstable manifolds***, denoted by $W^s_0(\gamma_0)$ and $W^u_0(\gamma_0)$ defined as the sets of points tending exponentially to γ_0 as $t \to \infty$ and $t \to -\infty$, respectively. (See Abraham, Marsden, and Ratiu [1988], Guckenheimer and Holmes [1983], or Wiggins [1988, 1990, 1992] for more details.) In this example, they coincide:

$$W^s_0(\gamma_0) = W^u_0(\gamma_0).$$

For $\epsilon > 0$ the (hyperbolic) closed orbit γ_0 perturbs to a nearby (hyperbolic) closed orbit that has stable and unstable manifolds $W^s_\epsilon(\gamma_\epsilon)$ and $W^u_\epsilon(\gamma_\epsilon)$. If $W^s_\epsilon(\gamma_\epsilon)$ and $W^u_\epsilon(\gamma_\epsilon)$ intersect transversally, we again say that (2.9.2) has ***horseshoes***. These two definitions of admitting horseshoes are readily seen to be equivalent.

Theorem 2.9.1 (Poincaré–Melnikov Theorem). *Let the Poincaré–Melnikov function be defined by* (2.9.3). *Assume that* $M(t_0)$ *has simple zeros as a* T*-periodic function of* t_0. *Then for sufficiently small* ϵ, *equation* (2.9.2) *has horseshoes, that is, homoclinic chaos in the sense of transversal intersecting separatrices.*

Idea of the Proof. In the suspended picture, we use the energy function H_0 to measure the first-order movement of $W^s_\epsilon(\gamma_\epsilon)$ at $\overline{z}(0)$ at time t_0 as ϵ is varied. Note that points of $\overline{z}(t)$ are regular points for H_0, since H_0 is constant on $\overline{z}(t)$, and $\overline{z}(0)$ is not a fixed point. That is, the differential of H_0 does not vanish at $\overline{z}(0)$. Thus, the values of H_0 give an accurate measure of the distance from the homoclinic orbit. If $(z^s_\epsilon(t, t_0), t)$ is the curve on $W^s_\epsilon(\gamma_\epsilon)$ that is an integral curve of the suspended system and has an condition $z^s_\epsilon(t_0, t_0)$ that is the perturbation of

$$W^s_0(\gamma_0) \cap \{ \text{the plane } t = t_0 \}$$

in the normal direction to the homoclinic orbit, then $H_0(z^s_\epsilon(t_0, t_0))$ measures the normal distance. But

$$\begin{aligned} H_0(z^s_\epsilon(\tau_+, t_0)) - H_0(z^s_\epsilon(t_0, t_0)) &= \int_{t_0}^{\tau_+} \frac{d}{dt} H_0(z^s_\epsilon(t, t_0))\, dt \\ &= \int_{t_0}^{\tau_+} \{H_0, H_0 + \epsilon H_1\}(z^s_\epsilon(t, t_0), t)\, dt. \end{aligned} \tag{2.9.5}$$

From invariant manifold theory one learns that $z^s_\epsilon(t, t_0)$ converges exponentially to $\gamma_\epsilon(t)$, a periodic orbit for the perturbed system as $t \to +\infty$. Notice from the right-hand side of the first equality above that if $z^s_\epsilon(t, t_0)$ were replaced by the periodic orbit $\gamma_\epsilon(t)$, the result would be zero. Since the convergence is exponential, one concludes that the integral is of order ϵ for an interval from some large time to infinity. To handle the finite portion of the integral, we use the fact that $z^s_\epsilon(t, t_0)$ is ϵ-close to $\overline{z}(t - t_0)$ (uniformly as $t \to +\infty$) and that $\{H_0, H_0\} = 0$. Therefore, we see that

$$\{H_0, H_0 + \epsilon H_1\}(z^s_\epsilon(t, t_0), t) = \epsilon \{H_0, H_1\}(\overline{z}(t - t_0), t) + O(\epsilon^2).$$

Using this over a large but finite interval $[t_0, t_1]$ and the exponential closeness over the remaining interval $[t_1, \infty)$, we see that (2.9.5) becomes

$$\begin{aligned} &H_0(z^s_\epsilon(\tau_+, t_0)) - H_0(z^s_\epsilon(t_0, t_0)) \\ &\qquad = \epsilon \int_{t_0}^{\tau_+} \{H_0, H_1\}(\overline{z}(t - t_0), t)\, dt + O(\epsilon^2), \end{aligned} \tag{2.9.6}$$

where the error is uniformly small as $\tau_+ \to \infty$. Similarly,

$$H_0(z_\epsilon^u(t_0, t_0)) - H_0(z_\epsilon^u(\tau_-, t_0)) = \epsilon \int_{\tau_-}^{t_0} \{H_0, H_1\}(\overline{z}(t - t_0), t)\, dt + O(\epsilon^2). \quad (2.9.7)$$

Again we use the fact that $z_\epsilon^s(\tau_+, t_0) \to \gamma_\epsilon(\tau_+)$ exponentially fast, a periodic orbit for the perturbed system as $\tau_+ \to +\infty$. Notice that since the orbit is *homoclinic*, the *same* periodic orbit can be used for negative times as well. Using this observation, we can choose τ_+ and τ_- such that

$$H_0(z_\epsilon^s(\tau_+, t_0)) - H_0(z_\epsilon^u(\tau_-, t_0)) \to 0$$

as $\tau_+ \to \infty$ and $\tau_- \to -\infty$. Adding (2.9.6) and (2.9.7), letting $\tau_+ \to \infty$ and $\tau_- \to -\infty$, we get

$$H_0(z_\epsilon^u(t_0, t_0)) - H_0(z_\epsilon^s(t_0, t_0)) = \epsilon \int_{-\infty}^{\infty} \{H_0, H_1\}(\overline{z}(t - t_0), t)\, dt + O(\epsilon^2). \quad (2.9.8)$$

The integral in this expression is convergent because the curve $\overline{z}(t - t_0)$ tends exponentially to the saddle point as $t \to \pm\infty$ and because the differential of H_0 vanishes at this point. Thus, the integrand tends to zero exponentially fast as t tends to plus and minus infinity.

Since the energy is a "good" measure of the distance between the points $z_\epsilon^u(t_0, t_0))$ and $z_\epsilon^s(t_0, t_0))$, it follows that if $M(t_0)$ has a simple zero at time t_0, then $z_\epsilon^u(t_0, t_0)$ and $z_\epsilon^s(t_0, t_0)$ intersect transversally near the point $\overline{z}(0)$ at time t_0. ■

If in (2.9.2) only X_0 is Hamiltonian, the same conclusion holds if (2.9.3) is replaced by

$$M(t_0) = \int_{-\infty}^{\infty} (X_0 \times X_1)(\overline{z}(t - t_0), t)\, dt, \quad (2.9.9)$$

where $X_0 \times X_1$ is the (scalar) cross product for planar vector fields. In fact, X_0 need not even be Hamiltonian if an area expansion factor is inserted.

Example A. Equation (2.9.9) applies to the forced damped Duffing equation

$$\ddot{u} - \beta u + \alpha u^3 = \epsilon(\gamma \cos \omega t - \delta \dot{u}). \quad (2.9.10)$$

Here the homoclinic orbits are given by (see Exercise 2.8-4)

$$u(t) = \pm\sqrt{\frac{2\beta}{\alpha}}\, \operatorname{sech}(\sqrt{\beta} t), \quad (2.9.11)$$

and (2.9.9) becomes, after a residue calculation,

$$M(t_0) = \gamma\pi\omega\sqrt{\frac{2}{\alpha}}\,\operatorname{sech}\left(\frac{\pi\omega}{2\sqrt{\beta}}\right)\sin(\omega t_0) - \frac{4\delta\beta^{3/2}}{3\alpha}, \tag{2.9.12}$$

so one has simple zeros and hence chaos of the horseshoe type if

$$\frac{\gamma}{\delta} > \frac{2\sqrt{2}\beta^{3/2}}{3\omega\sqrt{\alpha}}\cosh\left(\frac{\pi\omega}{2\sqrt{\beta}}\right) \tag{2.9.13}$$

and ϵ is small. ♦

Example B. Another interesting example, due to Montgomery [1985], concerns the equations for superfluid ^{3}He. These are the Leggett equations, and we shall confine ourselves to what is called the A phase for simplicity (see Montgomery's paper for additional results). The equations are

$$\dot{s} = -\frac{1}{2}\left(\frac{\chi\Omega^2}{\gamma^2}\right)\sin 2\theta$$

and

$$\dot{\theta} = \left(\frac{\gamma^2}{\chi}\right)s - \epsilon\left(\gamma B\sin\omega t + \frac{1}{2}\Gamma\sin 2\theta\right). \tag{2.9.14}$$

Here s is the spin, θ an angle (describing the "order parameter"), and $\gamma, \chi, \ldots$ are physical constants. The homoclinic orbits for $\epsilon = 0$ are given by

$$\overline{\theta}_\pm = 2\tan^{-1}(e^{\pm\Omega t}) - \frac{\pi}{2} \quad \text{and} \quad \overline{s}_\pm = \pm 2\frac{\Omega e^{\pm 2\Omega t}}{1 + e^{\pm 2\Omega t}}. \tag{2.9.15}$$

One calculates the Poincaré–Melnikov function to be

$$M_\pm(t_0) = \mp\frac{\pi\chi\omega B}{8\gamma}\operatorname{sech}\left(\frac{\omega\pi}{2\Omega}\right)\cos\omega t - \frac{2}{3}\frac{\chi}{\gamma^2}\Omega\Gamma, \tag{2.9.16}$$

so that (2.9.14) has chaos in the sense of horseshoes if

$$\frac{\gamma B}{\Gamma} > \frac{16}{3\pi}\frac{\Omega}{\omega}\cosh\left(\frac{\pi\omega}{2\Omega}\right) \tag{2.9.17}$$

and if ϵ is small. ♦

For references and information on higher-dimensional versions of the method and applications, see Wiggins [1988]. We shall comment on some aspects of this shortly. There is even a version of the Poincaré–Melnikov method applicable to PDEs (due to Holmes and Marsden [1981]). One basically still uses formula (2.9.9) where $X_0 \times X_1$ is replaced by the symplectic pairing between X_0 and X_1. However, there are two new difficulties in addition to standard technical analytic problems that arise with PDEs. The first is that there is a serious problem with resonances. This can be dealt with using the aid of damping. Second, the problem seems to be *not* reducible to two dimensions: The horseshoe involves all the modes. Indeed, the higher modes do seem to be involved in the physical buckling processes for the beam model discussed next.

Example C. A PDE model for a buckled forced beam is

$$\ddot{w} + w'''' + \Gamma w' - \kappa \left(\int_0^1 [w']^2 \, dz \right) w'' = \epsilon (f \cos \omega t - \delta \dot{w}), \tag{2.9.18}$$

where $w(z,t)$, $0 \leq z \leq 1$, describes the deflection of the beam,

$$\dot{} = \partial/\partial t, \quad {}' = \partial/\partial z,$$

and $\Gamma, \kappa, \ldots$ are physical constants. For this case, one finds that if

(i) $\pi^2 < \Gamma < 4\rho^3$ (first mode is buckled),

(ii) $j^2\pi^2(j^2\pi^2 - \Gamma) \neq \omega^2$, $j = 2, 3, \ldots$ (resonance condition),

(iii) $\dfrac{f}{\delta} > \dfrac{\pi(\Gamma - \pi^2)}{2\omega\sqrt{\kappa}} \cosh \left(\dfrac{\omega}{2\sqrt{\Gamma - \omega^2}} \right)$ (transversal zeros for $M(t_0)$),

(iv) $\delta > 0$,

and ϵ is small, then (2.9.18) has horseshoes. Experiments (see Moon [1987]) showing chaos in a forced buckled beam provided the motivation that led to the study of (2.9.18). ♦

This kind of result can also be used for a study of chaos in a van der Waals fluid (Slemrod and Marsden [1985]) and for soliton equations (see Birnir [1986], Ercolani, Forest, and McLaughlin [1990], and Birnir and Grauer [1994]). For example, in the damped, forced sine–Gordon equation one has chaotic transitions between breathers and kink-antikink pairs, and in the Benjamin–Ono equation one can have chaotic transitions between solutions with different numbers of poles.

More Degrees of Freedom. For Hamiltonian systems with two-degrees-of-freedom, Holmes and Marsden [1982a] show how the Melnikov method may be used to prove the existence of horseshoes on energy surfaces in nearly integrable systems. The class of systems studied has a Hamiltonian of the form

$$H(q, p, \theta, I) = F(q, p) + G(I) + \epsilon H_1(q, p, \theta, I) + O(\epsilon^2), \tag{2.9.19}$$

where (θ, I) are action-angle coordinates for the oscillator G; we assume that $G(0) = 0$, $G' > 0$. It is also assumed that F has a homoclinic orbit

$$\overline{x}(t) = (\overline{q}(t), \overline{p}(t))$$

and that

$$M(t_0) = \int_{-\infty}^{\infty} \{F, H_1\}\, dt; \tag{2.9.20}$$

the integral taken along $(\overline{x}(t - t_0), \Omega t, I)$ has simple zeros. Then (2.9.19) has horseshoes on energy surfaces near the surface corresponding to the homoclinic orbit and small I; the horseshoes are taken relative to a Poincaré map strobed to the oscillator G. The paper by Holmes and Marsden [1982a] also studies the effect of positive and negative damping. These results are related to those for forced one-degree-of-freedom systems, since one can often reduce a two-degrees-of-freedom Hamiltonian system to a one-degree-of-freedom forced system.

For some systems in which the variables do not split as in (2.9.19), such as a nearly symmetric heavy top, one needs to exploit a symmetry of the system, and this complicates the situation to some extent. The general theory for this is given in Holmes and Marsden [1983] and was applied to show the existence of horseshoes in the nearly symmetric heavy top; see also some closely related results of Ziglin [1980a].

This theory has been used by Ziglin [1980b] and Koiller [1985] in vortex dynamics, for example, to give a proof of the nonintegrability of the restricted four-vortex problem. Koiller, Soares, and Melo Neto [1985] give applications to the dynamics of general relativity showing the existence of horseshoes in Bianchi IX models. See Oh, Sreenath, Krishnaprasad, and Marsden [1989] for applications to the dynamics of coupled rigid bodies.

Arnold [1964] extended the Poincaré–Melnikov theory to systems with several degrees of freedom. In this case the transverse homoclinic manifolds are based on KAM tori and allow the possibility of chaotic drift from one torus to another. This drift, sometimes known as ***Arnold diffusion***, is a much studied topic in Hamiltonian systems, but its theoretical foundations are still the subject of much study.

Instead of a single Melnikov function, in the multidimensional case one has a ***Melnikov vector*** given schematically by

$$\mathbf{M} = \begin{pmatrix} \int_{-\infty}^{\infty} \{H_0, H_1\}\, dt \\ \int_{-\infty}^{\infty} \{I_1, H_1\}\, dt \\ \cdots \\ \int_{-\infty}^{\infty} \{I_n, H_1\}\, dt \end{pmatrix}, \tag{2.9.21}$$

where $I_1, \ldots, I_n$ are integrals for the unperturbed (completely integrable) system and where $\mathbf{M}$ depends on t_0 and on angles conjugate to $I_1, \ldots, I_n$.

One requires $\mathbf{M}$ to have transversal zeros in the vector sense. This result was given by Arnold for forced systems and was extended to the autonomous case by Holmes and Marsden [1982b, 1983]; see also Robinson [1988]. These results apply to systems such as a pendulum coupled to several oscillators and the many-vortex problems. It has also been used in power systems by Salam, Marsden, and Varaiya [1983], building on the horseshoe case treated by Kopell and Washburn [1982]. See also Salam and Sastry [1985]. There have been a number of other directions of research on these techniques. For example, Gruendler [1985] developed a multidimensional version applicable to the spherical pendulum, and Greenspan and Holmes [1983] showed how the Melnikov method can be used to study subharmonic bifurcations. See Wiggins [1988] for more information.

Poincaré and Exponentially Small Terms. In his celebrated memoir on the three-body problem, Poincaré [1890] introduced the mechanism of transversal intersection of separatrices that obstructs the integrability of the system of equations for the three-body problem as well as preventing the convergence of associated series expansions for the solutions. This idea has been developed by Birkhoff and Smale using the horseshoe construction to describe the resulting chaotic dynamics. However, in the region of phase space studied by Poincaré, it has never been proved (except in some generic sense that is not easy to interpret in specific cases) that the equations really are nonintegrable. In fact, Poincaré himself traced the difficulty to the presence of terms in the separatrix splitting that are exponentially small. A crucial component of the measure of the splitting is given by the following formula of Poincaré [1890, p. 223]:

$$J = \frac{-8\pi i}{\exp\left(\frac{\pi}{\sqrt{2\mu}}\right) + \exp\left(-\frac{\pi}{\sqrt{2\mu}}\right)},$$

which is exponentially small (also said to be beyond all orders) in μ. Poincaré was aware of the difficulties that this exponentially small behavior causes; on page 224 of his article, he states, "En d'autres termes, si on regarde μ comme un infiniment petit du premier ordre, la distance BB', sans être nulle, est un infiniment petit d'ordre infini. C'est ainsi que la fonction $e^{-1/\mu}$ est un infiniment petit d'ordre infini sans être nulle ... Dans l'example particulier que nous avons traité plus haut, la distance BB' est du mème ordre de grandeur que l'integral J, c'est à dire que $\exp(-\pi/\sqrt{2\mu})$."

This is a serious difficulty that arises when one uses the Melnikov method near an elliptic fixed point in a Hamiltonian system or in bifurcation problems giving birth to homoclinic orbits. The difficulty is related to those described by Poincaré. Near elliptic points, one sees homoclinic orbits in normal forms, and after a temporal rescaling this leads to a rapidly oscillatory perturbation that is modeled by the following variation of the pendulum equation:

$$\ddot{\phi} + \sin\phi = \epsilon \cos\left(\frac{\omega t}{\epsilon}\right). \tag{2.9.22}$$

If one-formally computes $M(t_0)$, one obtains

$$M(t_0, \epsilon) = \pm 2\pi \operatorname{sech}\left(\frac{\pi\omega}{2\epsilon}\right) \cos\left(\frac{\omega t_0}{\epsilon}\right). \tag{2.9.23}$$

While this has simple zeros, the proof of the Poincaré–Melnikov theorem is no longer valid, since $M(t_0, \epsilon)$ is now of order $\exp(-\pi/(2\epsilon))$, and the error analysis in the proof gives errors only of order ϵ^2. In fact, no expansion in powers of ϵ can detect exponentially small terms like $\exp(-\pi/(2\epsilon))$.

Holmes, Marsden, and Scheurle [1988] and Delshams and Seara [1991] show that (2.9.22) has chaos that is, in a suitable sense, *exponentially small* in ϵ. The idea is to expand expressions for the stable and unstable manifolds in a Perron type series whose terms are of order $\epsilon^k \exp(-\pi/(2\epsilon))$. To do so, the extension of the system to complex time plays a crucial role. One can hope that since such results for (2.9.22) can be proved, it may be possible to return to Poincaré's 1890 work and complete the arguments he left unfinished. In fact, the existence of these exponentially small phenomena is one reason that the problem of Arnold diffusion is both hard and delicate.

To illustrate how exponentially small phenomena enter bifurcation problems, consider the problem of a Hamiltonian saddle node bifurcation

$$\ddot{x} + \mu x + x^2 = 0 \tag{2.9.24}$$

with the addition of higher-order terms and forcing:

$$\ddot{x} + \mu x + x^2 + \text{h.o.t.} = \delta f(t). \tag{2.9.25}$$

The phase portrait of (2.9.24) is shown in Figure 2.9.1.

The system (2.9.24) is Hamiltonian with

$$H(x, \dot{x}) = \frac{1}{2}\dot{x}^2 + \frac{1}{2}\mu x^2 + \frac{1}{3}x^3. \tag{2.9.26}$$

Let us first consider the system without higher-order terms:

$$\ddot{x} + \mu x + x^2 = \delta f(t). \tag{2.9.27}$$

To study it, we rescale to blow up the singularity; let

$$x(t) = \lambda \xi(\tau), \tag{2.9.28}$$

where $\lambda = |\mu|$ and $\tau = t\sqrt{\lambda}$. Letting $' = d/d\tau$, we get

$$\begin{aligned} \xi'' - \xi + \xi^2 &= \frac{\delta}{\mu^2} f\left(\frac{\tau}{\sqrt{-\mu}}\right), \quad \mu < 0, \\ \xi'' + \xi + \xi^2 &= \frac{\delta}{\mu^2} f\left(\frac{\tau}{\sqrt{\mu}}\right), \qquad \mu > 0. \end{aligned} \tag{2.9.29}$$

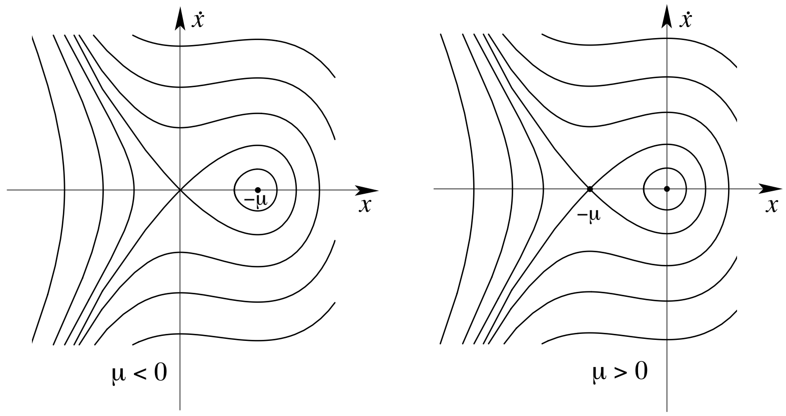

FIGURE 2.9.1. Phase portraits of $\ddot{x} + \mu x + x^2 = 0$.

The exponentially small estimates of Holmes, Marsden, and Scheurle [1988] apply to (2.9.29). One gets exponentially small upper and lower estimates in certain algebraic sectors of the (δ, μ) plane that depend on the nature of f. The estimates for the splitting have the form $C(\delta/\mu^2)\exp(-\pi/\sqrt{|\mu|})$. Now consider

$$\ddot{x} + \mu x + x^2 + x^3 = \delta f(t). \tag{2.9.30}$$

With $\delta = 0$, there are equilibria at the three points with $\dot{x} = 0$ and

$$x = 0, \; -r, \text{ and } -\frac{\mu}{r}, \tag{2.9.31}$$

where

$$r = \frac{1 + \sqrt{1 - 4\mu}}{2}, \tag{2.9.32}$$

which is approximately 1 when $\mu \approx 0$. The phase portrait of (2.9.30) with $\delta = 0$ and $\mu = -1/2$ is shown in Figure 2.9.2. As μ passes through 0, the small lobe in Figure 2.9.2 undergoes the same bifurcation as in Figure 2.9.1, with the large lobe changing only slightly.

Again we rescale, to give

$$\begin{aligned} \ddot{\xi} - \xi + \xi^2 - \mu\xi^3 &= \frac{\delta}{\mu^2} f\left(\frac{\tau}{\sqrt{-\mu}}\right), \quad \mu < 0, \\ \ddot{\xi} + \xi + \xi^2 + \mu\xi^3 &= \frac{\delta}{\mu^2} f\left(\frac{\tau}{\sqrt{\mu}}\right), \quad \mu > 0. \end{aligned} \tag{2.9.33}$$

Notice that for $\delta = 0$, the phase portrait is μ-dependent. The homoclinic orbit surrounding the small lobe for $\mu < 0$ is given explicitly in terms of ξ

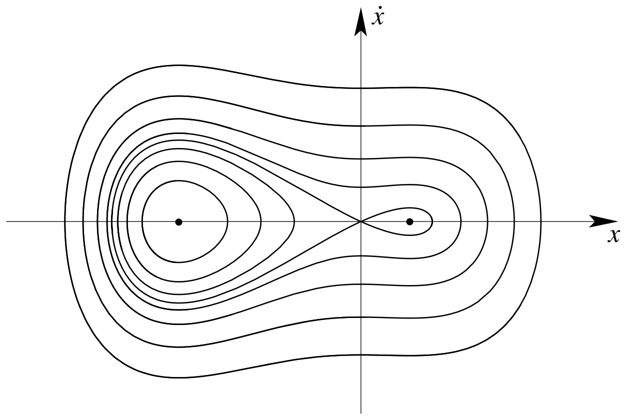

FIGURE 2.9.2. The phase portrait of $\ddot{x} - \frac{1}{2}x + x^2 + x^3 = 0$.

by

$$\xi(\tau) = \frac{4e^\tau}{\left(e^\tau + \frac{2}{3}\right)^2 - 2\mu}, \tag{2.9.34}$$

which is μ-dependent. An interesting technicality is that without the cubic term, we get μ-independent *double* poles at $t = \pm i\pi + \log 2 - \log 3$ in the complex τ-plane, while (2.9.34) has a pair of simple poles that splits these double poles to the pairs of simple poles at

$$\tau = \pm i\pi + \log\left(\frac{2}{3} \pm i\sqrt{2\lambda}\right), \tag{2.9.35}$$

where again $\lambda = |\mu|$. (There is no particular significance to the real part, such as $\log 2 - \log 3$ in the case of no cubic term; this can always be gotten rid of by a shift in the base point $\xi(0)$.)

If a quartic term x^4 is added, these pairs of simple poles will split into quartets of branch points, and so on. Thus, while the analysis of higher-order terms has this interesting μ-dependence, it seems that the basic exponential part of the estimates, namely

$$\exp\left(-\frac{\pi}{\sqrt{|\mu|}}\right), \tag{2.9.36}$$

remains intact.