Chapter 7
Chaos in Hamiltonian systems

Hamiltonian systems are a class of dynamical systems that occur in a wide variety of circumstances.[1] The special properties of Hamilton's equations endow these systems with attributes that differ qualitatively and fundamentally from other sytems. (For example, Hamilton's equations do not possess attractors.)

Examples of Hamiltonian dynamics include not only the well-known case of mechanical systems in the absence of friction, but also a variety of other problems such as the paths followed by magnetic field lines in a plasma, the mixing of fluids, and the ray equations describing the trajectories of propagating waves. In all of these situations chaos can be an important issue. Furthermore, chaos in Hamiltonian systems is at the heart of such fundamental questions as the foundations of statistical mechanics and the stability of the solar system. In addition, Hamiltonian mechanics and its structure are reflected in quantum mechanics. Thus, in Chapter 11 we shall treat the connection between chaos in Hamiltonian systems and related quantum phenomena. The present chapter will be devoted to a discussion of Hamiltonian dynamics and the role that chaos plays in these systems. We begin by presenting a summary of some basic concepts in Hamiltonian mechanics.[2,3]

7.1 Hamiltonian systems

The dynamics of a Hamiltonian system is completely specified by a single function, the Hamiltonian, $H(\mathbf{p}, \mathbf{q}, t)$. The state of the system is specified by its 'momentum' $\mathbf{p}$ and 'position' $\mathbf{q}$. Here the vectors $\mathbf{p}$ and $\mathbf{q}$ have the same dimensionality which we denote N. We call N the number of

degrees of freedom of the system. For example, Hamilton's equations for the motion of K point masses interacting in three-dimensional space via gravitational attraction has $N = 3K$ degrees of freedom, corresponding to the three spacial coordinates needed to specify the location of each mass. Hamilton's equations determine the trajectory $(\mathbf{p}(t), \mathbf{q}(t))$ that the system follows in the $2N$-dimensional phase space, and are given by

$$\mathrm{d}\mathbf{p}/\mathrm{d}t = -\partial H(\mathbf{p}, \mathbf{q}, t)/\partial \mathbf{q}, \tag{7.1a}$$

$$\mathrm{d}\mathbf{q}/\mathrm{d}t = \partial H(\mathbf{p}, \mathbf{q}, t)/\partial \mathbf{p}. \tag{7.1b}$$

In the special case that the Hamiltonian has no explicit time dependence, $H = H(\mathbf{p}, \mathbf{q})$, we can use Hamilton's equations to show that, as $\mathbf{p}$ and $\mathbf{q}$ vary with time, the value of $H(\mathbf{p}(t), \mathbf{q}(t))$ remains a constant:

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t}\cdot\frac{\partial H}{\partial \mathbf{q}} + \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}\cdot\frac{\partial H}{\partial \mathbf{p}} = \frac{\partial H}{\partial \mathbf{p}}\cdot\frac{\partial H}{\partial \mathbf{q}} - \frac{\partial H}{\partial \mathbf{q}}\cdot\frac{\partial H}{\partial \mathbf{p}} = 0.$$

Thus, identifying the value of the Hamiltonian with the energy E of the system, we see that the energy is conserved for time-independent systems, $E = H(\mathbf{p}, \mathbf{q}) = (\text{constant})$.

7.1.1 Symplectic structure

We can write Eqs. (7.1) in the form

$$\mathrm{d}\tilde{\mathbf{x}}/\mathrm{d}t = \mathbf{F}(\tilde{\mathbf{x}}, t), \tag{7.2}$$

by taking $\tilde{\mathbf{x}}$ to be the $2N$-dimensional vector

$$\tilde{\mathbf{x}} = \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix},$$

and by taking $\mathbf{F}(\tilde{\mathbf{x}})$ to be

$$\mathbf{F}(\tilde{\mathbf{x}}, t) = \mathbf{S}_N \cdot \partial H/\partial \tilde{\mathbf{x}}, \tag{7.3}$$

with

$$\mathbf{S}_N = \begin{bmatrix} \mathbf{O}_N & -\mathbf{I}_N \\ \mathbf{I}_N & \mathbf{O}_N \end{bmatrix} \tag{7.4}$$

where $\mathbf{I}_N$ is the N-dimensional identity matrix, $\mathbf{O}_N$ is the $N \times N$ matrix of zeros, and

$$\frac{\partial H}{\partial \tilde{\mathbf{x}}} = \begin{bmatrix} \partial H/\partial \mathbf{p} \\ \partial H/\partial \mathbf{q} \end{bmatrix}. \tag{7.5}$$

From this we see how restricted the class of Hamiltonian systems is. In particular, a general system of the form (7.2) requires the specification of all the components of the *vector* function $\mathbf{F}(\tilde{\mathbf{x}}, t)$, while by (7.3), if the

system is Hamiltonian, it is specified by a single scalar function of $\mathbf{p}$, $\mathbf{q}$ and t (the Hamiltonian).

One of the basic properties of Hamilton's equations is that they preserve $2N$-dimensional volumes in the phase space. This follows by taking the divergence of $\mathbf{F}(\tilde{\mathbf{x}})$ in Eq. (7.2), which gives

$$\frac{\partial}{\partial \tilde{\mathbf{x}}} \cdot \mathbf{F} = \frac{\partial}{\partial \mathbf{p}} \cdot \left(-\frac{\partial H}{\partial \mathbf{q}} \right) + \frac{\partial}{\partial \mathbf{q}} \cdot \left(\frac{\partial H}{\partial \mathbf{p}} \right) = 0. \tag{7.6}$$

Thus, if we consider an initial closed surface S_0 in the $2N$-dimensional phase space and evolve each point on the surface forward with time, we obtain at each instant of time t a new closed surface S_t which contains within it precisely the same $2N$-dimensional volume as does S_0. This follows from

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{S_t} d^{2N}\tilde{\mathbf{x}} = \oint_{S_t} \frac{\mathrm{d}\tilde{\mathbf{x}}}{\mathrm{d}t} \cdot \mathrm{d}\mathbf{S} = \oint_{S_t} \mathbf{F} \cdot \mathrm{d}\mathbf{S} = \int_{S_t} \frac{\partial}{\partial \tilde{\mathbf{x}}} \cdot \mathbf{F}\, \mathrm{d}^{2N}\tilde{\mathbf{x}} = 0,$$

where $\int_{S_t} \cdots$ denotes integration over the volume enclosed by S_t, $\oint_{S_t} \cdots$ denotes a surface integral over the closed surface S_t, and the third equality is from the divergence theorem (cf. Eq. (1.12)). As a consequence of this result, Hamiltonian systems do not have attractors in the usual sense. This incompressibility of phase space volumes for Hamiltonian systems is called Liouville's theorem.

Perhaps the most basic structural property of Hamilton's equations is that they are *symplectic*. That is, if we consider three orbits that are infinitesimally displaced from each other, $(\mathbf{p}(t), \mathbf{q}(t))$, $(\mathbf{p}(t) + \delta\mathbf{p}(t), \mathbf{q}(t) + \delta\mathbf{q}(t))$ and $(\mathbf{p}(t) + \delta\mathbf{p}'(t), \mathbf{q}(t) + \delta\mathbf{q}'(t))$, where $\delta\mathbf{p}$, $\delta\mathbf{q}$, $\delta\mathbf{p}'$ and $\delta\mathbf{q}'$ are infinitesimal N vectors, then the quantity,

$$\delta\mathbf{p} \cdot \delta\mathbf{q}' - \delta\mathbf{q} \cdot \delta\mathbf{p}',$$

which we call the differential symplectic area, is independent of time,

$$\frac{\mathrm{d}}{\mathrm{d}t}(\delta\mathbf{p} \cdot \delta\mathbf{q}' - \delta\mathbf{q} \cdot \delta\mathbf{p}') = 0. \tag{7.7}$$

The differential symplectic area can also be written as

$$\delta\mathbf{p} \cdot \delta\mathbf{q}' - \delta\mathbf{q} \cdot \delta\mathbf{p}' = \delta\tilde{\mathbf{x}}^{\dagger} \cdot \mathbf{S}_N \cdot \delta\tilde{\mathbf{x}}' \tag{7.8}$$

where $\dagger$ denotes transpose.

To derive (7.7) we differentiate (7.8) with respect to time and use Eqs. (7.2)–(7.5):

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\delta\tilde{\mathbf{x}}^{\dagger}\cdot\mathbf{S}_N\cdot\delta\tilde{\mathbf{x}}' &= \frac{\mathrm{d}\delta\tilde{\mathbf{x}}^{\dagger}}{\mathrm{d}t}\cdot\mathbf{S}_N\cdot\delta\tilde{\mathbf{x}}' + \delta\tilde{\mathbf{x}}^{\dagger}\cdot\mathbf{S}_N\cdot\frac{\mathrm{d}\delta\tilde{\mathbf{x}}'}{\mathrm{d}t} \\
&= \left(\frac{\partial\mathbf{F}}{\partial\tilde{\mathbf{x}}}\cdot\delta\tilde{\mathbf{x}}\right)^{\dagger}\cdot\mathbf{S}_N\cdot\delta\tilde{\mathbf{x}}' + \delta\tilde{\mathbf{x}}^{\dagger}\cdot\mathbf{S}_N\cdot\frac{\partial\mathbf{F}}{\partial\tilde{\mathbf{x}}}\cdot\delta\tilde{\mathbf{x}}' \\
&= \delta\tilde{\mathbf{x}}^{\dagger}\cdot\left[\left(\frac{\partial\mathbf{F}}{\partial\tilde{\mathbf{x}}}\right)^{\dagger}\cdot\mathbf{S}_N + \mathbf{S}_N\cdot\frac{\partial\mathbf{F}}{\partial\tilde{\mathbf{x}}}\right]\cdot\delta\tilde{\mathbf{x}}' \\
&= \delta\tilde{\mathbf{x}}^{\dagger}\cdot\left[\left(\mathbf{S}_N\cdot\frac{\partial^2 H}{\partial\tilde{\mathbf{x}}\partial\tilde{\mathbf{x}}}\right)^{\dagger}\cdot\mathbf{S}_N + \mathbf{S}_N\cdot\mathbf{S}_N\cdot\frac{\partial^2 H}{\partial\tilde{\mathbf{x}}\partial\tilde{\mathbf{x}}}\right]\cdot\delta\tilde{\mathbf{x}}' \\
&= \delta\tilde{\mathbf{x}}^{\dagger}\cdot\left[\left(\frac{\partial^2 H}{\partial\tilde{\mathbf{x}}\partial\tilde{\mathbf{x}}}\right)^{\dagger}\cdot\mathbf{S}_N^{\dagger}\cdot\mathbf{S}_N + \mathbf{S}_N\cdot\mathbf{S}_N\cdot\frac{\partial^2 H}{\partial\tilde{\mathbf{x}}\partial\tilde{\mathbf{x}}}\right]\cdot\delta\tilde{\mathbf{x}}' \\
&= 0,
\end{aligned}
$$

where we have used $\mathbf{S}_N\cdot\mathbf{S}_N = -\mathbf{I}_{2N}$ (where $\mathbf{I}_{2N}$ is the $2N$-dimensional identity matrix), $\mathbf{S}_N^{\dagger} = -\mathbf{S}_N$ and noted that $\partial^2 H/\partial\tilde{\mathbf{x}}\partial\tilde{\mathbf{x}}$ is a symmetric matrix. (In terms of the notation of Chapter 4, $\partial\mathbf{F}/\partial\tilde{\mathbf{x}} = \mathbf{DF}$.) For the case of one-degree-of-freedom systems ($N = 1$), Eq. (7.7) says that infinitesimal areas are preserved following the flow. (Figure 7.1 shows two infinitesimal vectors defining an infinitesimal parallelogram. The parallelogram area is $\delta p'\delta q - \delta q'\delta p$.) Since infinitesimal areas are preserved by a Hamiltonian flow with $N = 1$, so are noninfinitesimal areas. Thus for $N = 1$ Liouville's theorem and the symplectic condition are the same condition. For $N > 1$ the symplectic condition is not implied by Liouville's theorem. It can be shown,[2] however, that the symplectic condition implies volume conservation; so the symplectic condition is the more fundamental requirement for Hamiltonian mechanics. We interpret (7.7) as saying that the algebraic sum of the parallelogram areas formed

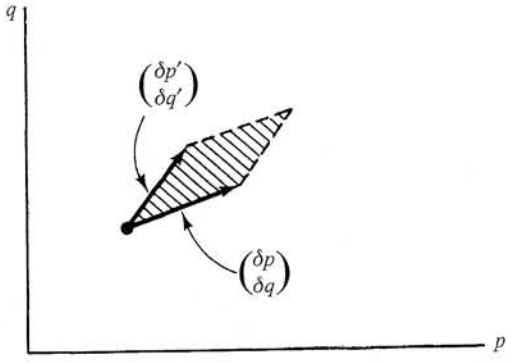


Figure 7.1 Infinitesimal area defined by the two infinitesimal vectors $(\delta p, \delta q)$ and $(\delta p', \delta q')$ for $N = 1$.

by projecting the vectors $\delta\mathbf{p}$, $\delta\mathbf{q}$, $\delta\mathbf{p}'$, $\delta\mathbf{q}'$ to the N coordinate planes (p_i, q_i) is conserved,

$$\delta\mathbf{p}\cdot\delta\mathbf{q}' - \delta\mathbf{q}\cdot\delta\mathbf{p}' = \sum_{i=1}^{N}(\delta p_i \delta q_i' - \delta q_i \delta p_i').$$

The quantity $\delta\mathbf{p}\cdot\delta\mathbf{q}' - \delta\mathbf{q}\cdot\delta\mathbf{p}'$ is the differential form of *Poincaré's integral invariant*,[2]

$$\oint_\gamma \mathbf{p}\cdot\mathrm{d}\mathbf{q} = \sum_{i=1}^{N}\oint_\gamma p_i\,\mathrm{d}q_i, \tag{7.9a}$$

where the integral is taken around a closed path γ in $(\mathbf{p}, \mathbf{q})$-space. We also refer to the quantity $\int_\gamma \mathbf{p}\cdot\mathrm{d}\mathbf{q}$ as the *symplectic area*. Poincaré's integral invariant is independent of time if the closed path γ is taken following the flow in phase space.[2] That is, $\gamma(t)$ is the path obtained from $\gamma(0)$ by evolving all the points on $\gamma(0)$ forward in time by the amount t via Hamilton's equations.

A useful generalization of the above statement of the invariance of $\oint\mathbf{p}\cdot\mathrm{d}\mathbf{q}$ following the flow is the *Poincaré–Cartan integral theorem*.[2] Consider the $(2N+1)$-dimensional extended phase space $(\mathbf{p}, \mathbf{q}, t)$. Let Γ_1 be a closed curve in this space and consider the tube of trajectories through points on Γ_1 as shown in Figure 7.2 for $N = 1$. The Poincare–Cartan integral theorem states that the 'action integral' around the path

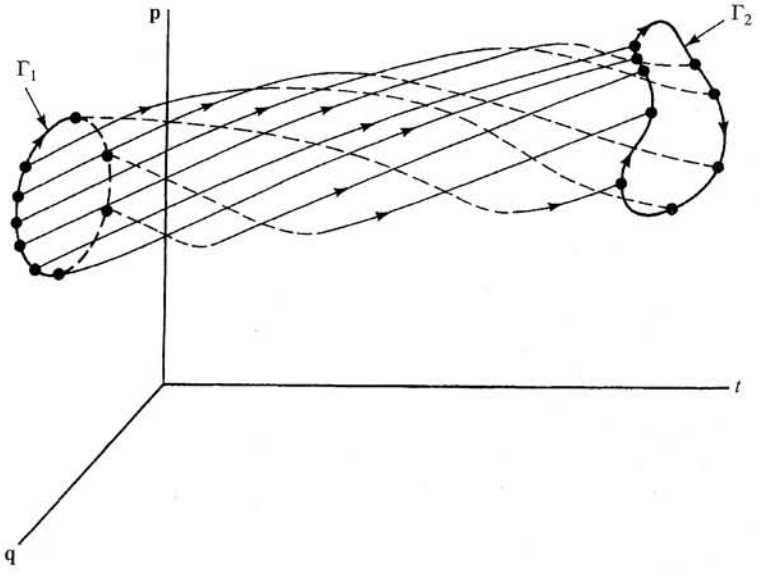


Figure 7.2 Trajectory tube through Γ_1.

Γ_1, $\oint_{\Gamma_1}(\mathbf{p}\cdot d\mathbf{q} - H\,dt)$, is the same value for any other path Γ_2 encircling the same tube of trajectories,

$$\oint_{\Gamma_1}(\mathbf{p}\cdot d\mathbf{q} - H\,dt) = \oint_{\Gamma_2}(\mathbf{p}\cdot d\mathbf{q} - H\,dt), \tag{7.9b}$$

where Γ_1 and Γ_2 are illustrated in Figure 7.2. If Γ_1 and Γ_2 are taken at two different (but constant) times, then $dt = 0$ in the above, and we recover the invariant (7.9a). Another important case is where H has no explicit t dependence, $H = H(\mathbf{p}, \mathbf{q})$. In this case the Hamiltonian is a constant of the motion, and, if we restrict the points on Γ_1 to all have the same value of H (i.e., Γ_1 lies on the $2N$-dimensional surface $H(\mathbf{p}, \mathbf{q}) = (\text{constant})$), then $\oint H\,dt = 0$ for any closed path on the $2N$-dimensional constant H surface in $(\mathbf{p}, \mathbf{q}, t)$ space. In this case we obtain

$$\oint_{\Gamma_1}\mathbf{p}\cdot d\mathbf{q} = \oint_{\Gamma_2}\mathbf{p}\cdot d\mathbf{q}. \tag{7.9c}$$

Note that, in contrast to (7.9a), the closed paths Γ_1 and Γ_2 in (7.9c) are in $(\mathbf{p}, \mathbf{q}, t)$ space and thus need not be taken at constant time.

7.1.2 Canonical changes of variables

If we introduce some arbitrary change of variables, $\tilde{\mathbf{x}}' = \mathbf{g}(\tilde{\mathbf{x}})$, the Hamiltonian form of the equations may not be preserved. Changes of variables which preserve the Hamiltonian form of the equations are said to be *canonical*, and the momentum and position vectors in terms of which one has a system in the Hamiltonian form (7.1) are said to be *canonically conjugate*. Specifically, if $\mathbf{p}$ and $\mathbf{q}$ satisfy (7.1), then a canonical change of variables to a new set of canonically conjugate variables $\overline{\mathbf{p}}$ and $\overline{\mathbf{q}}$

$$\overline{\mathbf{p}} = \mathbf{g}_1(\mathbf{p}, \mathbf{q}, t),$$
$$\overline{\mathbf{q}} = \mathbf{g}_2(\mathbf{p}, \mathbf{q}, t),$$

leads to evolution equations for $\overline{\mathbf{p}}$ and $\overline{\mathbf{q}}$ of the form,

$$d\overline{\mathbf{p}}/dt = -\partial\overline{H}(\overline{\mathbf{p}}, \overline{\mathbf{q}}, t)/\partial\overline{\mathbf{q}},$$
$$d\overline{\mathbf{q}}/dt = \partial\overline{H}(\overline{\mathbf{p}}, \overline{\mathbf{q}}, t)/\partial\overline{\mathbf{p}},$$

where $\overline{H}$ is a new transformed Hamiltonian for the system. One way to specify a canonical change of variables is to introduce a *generating function*,[2] $S(\overline{\mathbf{p}}, \mathbf{q}, t)$ which is a function of the 'old' position coordinates $\mathbf{q}$ and the 'new' momentum coordinates $\overline{\mathbf{p}}$. In terms of $S(\overline{\mathbf{p}}, \mathbf{q}, t)$, the change of variables is specified by

$$\overline{\mathbf{q}} = \frac{\partial S(\overline{\mathbf{p}}, \mathbf{q}, t)}{\partial\overline{\mathbf{p}}}, \quad \mathbf{p} = \frac{\partial S(\overline{\mathbf{p}}, \mathbf{q}, t)}{\partial\mathbf{q}}. \tag{7.10}$$

Thus the change of variables is given implicitly: to obtain $\overline{\mathbf{p}}$ in terms of $\mathbf{p}$ and $\mathbf{q}$ solve $\mathbf{p} = \partial S/\partial \mathbf{q}$ for $\overline{\mathbf{p}}$; to obtain $\overline{\mathbf{q}}$ in terms of $\mathbf{p}$ and $\mathbf{q}$ substitute the solution for $\overline{\mathbf{p}}$ into $\overline{\mathbf{q}} = \partial S/\partial \mathbf{p}$. Note that the change of variables specified by Eq. (7.10) is guaranteed to be symplectic. That is

$$\delta\mathbf{p} \cdot \delta\mathbf{q}' - \delta\mathbf{q} \cdot \delta\mathbf{p}' = \delta\overline{\mathbf{p}} \cdot \delta\overline{\mathbf{q}}' - \delta\overline{\mathbf{q}} \cdot \delta\overline{\mathbf{p}}.$$

This can be checked by differentiating Eq. (7.10),

$$\delta\overline{\mathbf{q}} = \frac{\partial^2 S}{\partial\overline{\mathbf{p}}\partial\overline{\mathbf{p}}} \cdot \delta\overline{\mathbf{p}} + \frac{\partial^2 S}{\partial\overline{\mathbf{p}}\partial\mathbf{q}} \cdot \delta\mathbf{q},$$

$$\delta\mathbf{p} = \frac{\partial^2 S}{\partial\mathbf{q}\delta\overline{\mathbf{p}}} \cdot \partial\overline{\mathbf{p}} + \frac{\partial^2 S}{\partial\mathbf{q}\partial\mathbf{q}} \cdot \delta\mathbf{q},$$

and substituting into the symplectic condition given above. In terms of the generating function the new Hamiltonian is given by[2]

$$\overline{H}(\overline{\mathbf{p}}, \overline{\mathbf{q}}, t) = H(\mathbf{p}, \mathbf{q}, t) + \partial S/\partial t. \tag{7.11}$$

7.1.3 Hamiltonian maps

Say we consider a Hamiltonian system and define the 'time T map' $\mathscr{M}_T$ for the system as

$$\mathscr{M}_T(\tilde{\mathbf{x}}(t), t) = \tilde{\mathbf{x}}(t + T). \tag{7.12}$$

(The explicit dependence on t in the second argument of $\mathscr{M}_T$ is absent if the Hamiltonian is time-independent.)

Taking a differential variation of Eq. (7.12) with respect to $\tilde{\mathbf{x}}$, we have

$$\frac{\partial\mathscr{M}_T}{\partial\tilde{\mathbf{x}}} \cdot \delta\tilde{\mathbf{x}}(t) = \delta\tilde{\mathbf{x}}(t + T).$$

The symplectic condition for a Hamiltonian flow

$$\delta\tilde{\mathbf{x}}^\dagger(t + T) \cdot \mathbf{S}_N \cdot \delta\tilde{\mathbf{x}}'(t + T) = \delta\tilde{\mathbf{x}}(t) \cdot \mathbf{S}_N \cdot \delta\tilde{\mathbf{x}}'(t)$$

yields

$$\delta\tilde{\mathbf{x}}^\dagger(t) \cdot \mathbf{S}_N \cdot \delta\tilde{\mathbf{x}}'(t) = \left(\frac{\partial\mathscr{M}_T}{\partial\tilde{\mathbf{x}}} \cdot \delta\tilde{\mathbf{x}}(t)\right)^\dagger \cdot \mathbf{S}_N \cdot \left(\frac{\partial\mathscr{M}_T}{\partial\tilde{\mathbf{x}}} \cdot \delta\tilde{\mathbf{x}}'(t)\right)$$

which (since $\delta\tilde{\mathbf{x}}(t)$ and $\delta\tilde{\mathbf{x}}'(t)$ are arbitrary) implies that the matrix $\partial\mathscr{M}_T/\partial\tilde{\mathbf{x}}$ satisfies

$$\mathbf{S}_N = \left(\frac{\partial\mathscr{M}_T}{\partial\tilde{\mathbf{x}}}\right)^\dagger \cdot \mathbf{S}_N \cdot \left(\frac{\partial\mathscr{M}_T}{\partial\tilde{\mathbf{x}}}\right).$$

The matrix $\partial\mathscr{M}_T/\partial\tilde{\mathbf{x}}$ is said to be a symplectic matrix, and we define a general $2N \times 2N$ matrix $\mathbf{A}$ to be symplectic if it satisfies

$$\mathbf{S}_N = \mathbf{A}^\dagger \cdot \mathbf{S}_N \cdot \mathbf{A}. \tag{7.13}$$

The product of symplectic matrices is also symplectic. To see this, suppose that $\mathbf{A}$ and $\mathbf{B}$ are symplectic. Then

$$(\mathbf{AB})^\dagger \cdot \mathbf{S}_N(\mathbf{AB}) = \mathbf{B}^\dagger \cdot (\mathbf{A}^\dagger \cdot \mathbf{S}_N \cdot \mathbf{A}) \cdot \mathbf{B} = \mathbf{B}^\dagger \cdot \mathbf{S}_N \cdot \mathbf{B} = \mathbf{S}_N.$$

So $\mathbf{AB}$ is symplectic.

One consequence of the conservation of phase space volumes for Hamiltonian systems is the *Poincaré recurrence theorem*. Say we consider a time-independent Hamiltonian $H = H(\mathbf{p}, \mathbf{q})$ for the case where all orbits are bounded. (This occurs if the energy surface is bounded; i.e., there are no solutions of $E = H(\mathbf{p}, \mathbf{q})$ with $|\mathbf{p}| \to \infty$ or $|\mathbf{q}| \to \infty$.) Now pick *any* initial point in phase space, and surround it with a ball R_0 of small radius $\varepsilon > 0$. Poincaré's recurrence theorem states that, if there are points which leave the initial ball, there are always some of these which will return to it if we wait long enough, and this is true no matter how small we choose ε to be. In order to see that this is so consider the time T map, Eq. (7.12), which evolves points forward in time by an amount T. Say that under the time T map the initial ball R_0 is mapped to a region R_1 outside the initial ball ($R_1 \cap R_0$ is empty). Continue mapping so as to obtain regions R_2, R_3, ... By Liouville's theorem all these regions have the same volume, equal to the volume of the initial ball R_0. Since the orbits are bounded, they are confined to a finite volume region of phase space. Thus, as we repeatedly apply the time T map, we must eventually find that we produce a region R_r which overlaps a previously produced region R_s, $r > s$. (If this is not so, then we would eventually come to the impossible situation where the sum of the volumes of the nonoverlapping R_s would eventually be larger than the volume of the bounded region that they are confined to.) Now apply the inverse of the time T map to R_r and R_s. This inverse must produce intersecting regions (namely R_{r-1} and R_{s-1}). Applying the inverse map s times we conclude that R_{r-s} (recall that $r - s > 0$) intersects the original ball R_0. Thus, as originally claimed, there are points in R_0 which return to R_0 after some time $(r - s)T$.

As in the case of general non-Hamiltonian systems, the surface of section technique also provides an extremely useful tool for analysis in Hamiltonian systems. There are two cases that are of interest.

(*a*) The Hamiltonian depends periodically on time: $H = H(\mathbf{p}, \mathbf{q}, t) = H(\mathbf{p}, \mathbf{q}, t + \tau)$, where τ is the period.
(*b*) The Hamiltonian has no explicit dependence on time: $H = H(\mathbf{p}, \mathbf{q})$.

First, we consider the case of a time periodic Hamiltonian. In that case, we can consider the phase space as having dimension $2N + 1$ by replacing the argument t in H by a dependent variable ξ, taking the phase space

variables to be $(\mathbf{p}, \mathbf{q}, \xi)$, and supplementing Hamilton's equations by the addition of the equation $d\xi/dt = 1$. Since the Hamiltonian is periodic in ξ with period τ, we can consider ξ as an angle variable and replace its value in the Hamiltonian by

$$\bar{\xi} = \xi \text{ modulo } \tau.$$

We then use for our surface of section the surface, $\bar{\xi} = t_0$, where t_0 is a constant between zero and τ. (This is the same construction we used for the periodically driven damped pendulum equation in Chapter 1.) Since the Hamiltonian is time periodic, the time T map defined by (7.12) satisfies

$$\mathcal{M}_\tau(\tilde{\mathbf{x}}, t_0) = \mathcal{M}_\tau(\tilde{\mathbf{x}}, t_0 + n\tau),$$

where n is an integer and we have taken $T = \tau$. Hence the surface of section map, which we denote $\mathbf{M}(\tilde{\mathbf{x}})$, is

$$\mathbf{M}(\tilde{\mathbf{x}}) = \mathcal{M}_\tau(\tilde{\mathbf{x}}, t_0),$$

and $\mathbf{M}$ is endowed with the same symplectic properties as $\mathcal{M}_\tau$ (i.e., the matrix $\partial\mathbf{M}/\partial\tilde{\mathbf{x}}$ satisfies (7.13)).

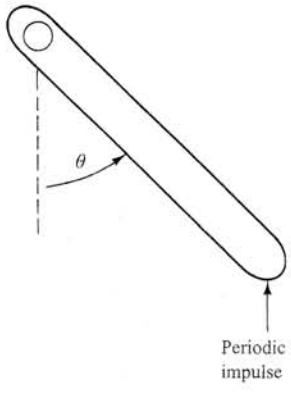
Example: Consider the 'kicked rotor' illustrated in Figure 7.3. There is a bar of moment of intertia $\bar{I}$ and length l, which is fastened at one end to a frictionless pivot. The other end is subjected to a vertical periodic impulsive force of impulse strength K/l applied at times $t = 0, \tau, 2\tau, \ldots$ (There is no gravity.) Using canonically conjugate variables p_θ (representing the angular momentum) and θ (the angular position of the rotor), we have that the Hamiltonian for this system and the corresponding equations of motion obtained from (7.1) are given by

$$H(p_\theta, \theta, t) = p_\theta^2/(2\bar{I}) + K\cos\theta \sum_n \delta(t - n\tau),$$

$$\frac{dp_\theta}{dt} = K\sin\theta \sum_n \delta(t - n\tau). \tag{7.14a}$$

$$\frac{d\theta}{dt} = p_\theta/\bar{I}, \tag{7.14b}$$

Figure 7.3 The kicked rotor. There is no gravity and no friction at the pivot.



where $\delta(\ldots)$ denotes the Dirac delta function. From Eqs. (7.14) we see that p_θ is constant between the kicks but changes discontinuously at each kick. The position variable θ varies linearly with t between kicks (because p_θ is constant) and is continuous at each kick. For our surface of section we examine the values of p_θ and θ just after each kick. Let p_n and θ_n denote the values of p_θ and θ at times $t = n\tau + 0^+$, where 0^+ denotes a positive infinitesimal. By integrating (7.14a) through the delta function at $t = (n+1)\tau$, we then obtain

$$p_{n+1} - p_n = K \sin \theta_{n+1},$$

and from (7.14b), $\theta_{n+1} - \theta_n = p_n \tau / \bar{I}$. Without loss of generality we can take $\tau/\bar{I} = 1$ to obtain the map

$$\theta_{n+1} = (\theta_n + p_n) \text{ modulo } 2\pi, \tag{7.15a}$$

$$p_{n+1} = p_n + K \sin \theta_{n+1}, \tag{7.15b}$$

where we have added a modulo 2π to Eq. (7.15a) since θ is an angle, and we wish to restrict its value to be between zero and 2π. The map given by Eqs. (7.15) is often called the 'standard map' and has proven to be a very convenient model for the study of the typical chaotic behavior of Hamiltonian systems that yield a two-dimensional map. It is a simple matter to check that Eqs. (7.15) preserve area in (p, θ)-space. To do this we calculate the determinant of the Jacobian of the map and verify that it is 1:

$$\det \begin{bmatrix} \partial\theta_{n+1}/\partial\theta_n & \partial\theta_{n+1}/\partial p_n \\ \partial p_{n+1}/\partial\theta_n & \partial p_{n+1}/\partial p_n \end{bmatrix} = \det \begin{bmatrix} 1 & 1 \\ K\cos\theta_{n+1} & 1 + K\cos\theta_{n+1} \end{bmatrix} = 1.$$

Since $N = 1$ this also implies that the map is symplectic, as required.

We now consider the second class of surface of sections that we have mentioned, namely the case where the Hamiltonian has no explicit time dependence. In this case, since the energy is conserved, the motion of the system is restricted to the $(2N-1)$-dimensional surface given by $E = H(\mathbf{p}, \mathbf{q})$. Taking a surface of section we would then obtain a $(2N-2)$-dimensional map. Say we choose for our surface of section the plane $q_1 = K_0$ (where K_0 is a constant), and say we give the values of the $2N-2$ quantities, $p_2, p_3, \ldots, p_N, q_2, q_3, \ldots, q_N$, on this plane. Let $\hat{\mathbf{x}}$ denote the vector specifying these coordinate values on the surface of section $\hat{\mathbf{x}} \equiv (p_2, p_3, \ldots, p_N, q_2, q_3, \ldots, q_N)$. Is there a map, $\hat{\mathbf{x}}_{n+1} = \mathbf{M}(\hat{\mathbf{x}}_n)$, evolving points forward on the surface of section; i.e., does a knowledge of $\hat{\mathbf{x}}_n$ uniquely determine the location of the next point on the surface of section? Given $\hat{\mathbf{x}}$ on the surface of section, the only unknown is p_1 (q_1 is known since we are on the surface of section $q_1 = K_0$). If we can determine p_1 then the phase space position $\tilde{\mathbf{x}}$ is known, and this uniquely determines the system's future evolution, and hence $\hat{\mathbf{x}}$ at the next piercing of the surface of section. To find p_1 we attempt to solve the equation $E = H(\mathbf{p}, \mathbf{q})$ for the single unknown p_1. The problem is that this equation will in general have multiple solutions for p_1. For example, for the commonly encountered case where the Hamiltonian is in the form of a kinetic energy $p^2/2m$, plus a potential energy,

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

for given $\hat{\mathbf{x}}$ there are two roots for p_1,

$$p_1 = \pm\{2m[E - \mathrm{V}(\mathbf{q})] - (p_2^2 + p_3^2 + \cdots + p_N^2)\}^{1/2}. \tag{7.16}$$

To make our determination of p_1 unique we adopt the following procedure. We specify $\hat{\mathbf{x}}_n$ to be the coordinates $(p_2, \ldots, p_N, q_2, \ldots, q_N)$ at the nth time at which $q_1(t) = K_0$ *and* $p_1 > 0$. That is, we only count surface of section piercings which cross $q_1 = K_0$ from $q_1 < K_0$ to $q_1 > K_0$ and not vice versa (for the Hamiltonian under consideration $\mathrm{d}q_1/\mathrm{d}t = p_1$). Hence, we *define* the surface of section so that we always take the positive root in (7.16) for p_1 (we could equally well have chosen $p_1 < 0$, rather than $p_1 > 0$, in our definition). With this definition, specification of $\hat{\mathbf{x}}_n$ uniquely determines a point in phase space. This point can be advanced by Hamilton's equations to the next time that $q_1(t) = K_0$ with $p_1 > 0$ thus determining $\hat{\mathbf{x}}_{n+1}$. In this way we determine a map,

$$\hat{\mathbf{x}}_{n+1} = \mathbf{M}(\hat{\mathbf{x}}_n).$$

This $(2N-2)$-dimensional map is symplectic in the remaining canonical variables $\hat{\mathbf{p}} = (p_2, \ldots, p_N)$ and $\hat{\mathbf{q}} = (q_2, \ldots, q_N)$. (This also implies that the map conserves $(2N-2)$-dimensional volumes.) To show that the map is symplectic, we need to demonstrate that the symplectic area,

$$\oint_\Gamma \hat{\mathbf{p}} \cdot \mathrm{d}\hat{\mathbf{q}},$$

is invariant when the closed path Γ around which the integral is taken is acted on by the map $\mathbf{M}$. This follows immediately from the Poincaré–Cartan theorem in the form of Eq. (7.9c). Writing $\mathbf{p} \cdot \mathrm{d}\mathbf{q} = p_1\,\mathrm{d}q_1 + \hat{\mathbf{p}} \cdot \mathrm{d}\hat{\mathbf{q}}$ and noting that $q_1 = K_0$ on the surface of section, we have $\mathrm{d}q_1 = 0$, and the desired result follows. (Note that use of the Poincaré–Cartan theorem (rather than the integral invariant (7.9a)) is necessary here because two different initial conditions starting in the surface of section take different amounts of time to return to it.)

Thus, we see that, in the cases of both a time-periodic Hamiltonian and a time-independent Hamiltonian, the resulting maps are symplectic. For this reason symplectic maps have played an important role, especially with respect to numerical experiments, in elucidating possible types of chaotic behavior in Hamiltonian systems.

One consequence of the symplectic nature of these maps is that the Lyapunov exponents occur in pairs $\pm h_1, \pm h_2, \pm h_3, \ldots$. Thus for each positive exponent there is a negative exponent of equal magnitude, and the number of zero exponents is even. To see why this is so we recall that the Lyapunov exponents are obtained from the product of the matrices $\mathbf{DM}(\tilde{\mathbf{x}}_n)\mathbf{DM}(\tilde{\mathbf{x}}_{n-1}) \ldots \mathbf{DM}(\tilde{\mathbf{x}}_0)$; see Section 4.4. In the Hamiltonian case the matrices $\mathbf{DM}(\tilde{\mathbf{x}}_j)$ are symplectic. Since the product of symplectic matrices is also symplectic, the overall matrix, $\mathbf{DM}(\tilde{\mathbf{x}}_n) \ldots \mathbf{DM}(\tilde{\mathbf{x}}_0)$, is symplectic. Now let us examine what the symplectic condition implies for

the eigenvalues of a matrix. The eigenvalues λ of a symplectic matrix $\mathbf{A}$ are the roots of its characteristic polynomial

$$D(\lambda) = \det[\mathbf{A} - \lambda \mathbf{I}].$$

Multiplying Eq. (7.13) on the left by $\mathbf{S}_N^{-1}(\mathbf{A}^\dagger)^{-1}$ we have

$$\mathbf{A} = \mathbf{S}_N^{-1}(\mathbf{A}^\dagger)^{-1}\mathbf{S}_N.$$

The characteristic polynomial then becomes

$$\begin{aligned} D(\lambda) &= \det\{\mathbf{S}_N^{-1}(\mathbf{A}^\dagger)^{-1}\mathbf{S}_N - \lambda \mathbf{I}] \\ &= \det[(\mathbf{A}^\dagger)^{-1} - \lambda \mathbf{I}] \\ &= \det[\mathbf{A}^{-1} - \lambda \mathbf{I}]. \end{aligned}$$

Thus, the eigenvalues of $\mathbf{A}$ and $\mathbf{A}^{-1}$ are the same. Since the eigenvalues of $\mathbf{A}^{-1}$ and $\mathbf{A}$ are also inverses of each other, we see that the eigenvalues must occur in pairs, (λ, λ^{-1}). Because the Lyapunov exponents are obtained from the logarithms of the magnitudes of the eigenvalues, $(h = \ln|\lambda|)$ we conclude that they occur in pairs $\pm h$.

As an example, we consider the stability of a periodic orbit of a symplectic two-dimensional map. If the period of the orbit is r, then the problem reduces to considering the stability of the fixed points of $\mathbf{M}^r(\bar{\mathbf{x}})$ which is also area preserving. Hence, it suffices to examine the stability of the fixed points of symplectic two-dimensional maps. Let $\mathbf{J} = \mathbf{D}\mathbf{M}^r$ denote the Jacobian matrix of the map at such a fixed point. Since the map is symplectic, we have $\det \mathbf{J} = 1$. The eigenvalues of $\mathbf{J}$ are given by

$$\det\begin{bmatrix} J_{11} - \lambda & J_{12} \\ J_{21} & J_{22} - \lambda \end{bmatrix} = \lambda^2 - \hat{T}\lambda + 1 = 0,$$

where $\hat{T} \equiv J_{11} + J_{22}$ is the trace of $\mathbf{J}$, and the last term in the quadratic is one by virtue of $\det \mathbf{J} = 1$. The solutions of the quadratic are

$$\lambda = [\hat{T} \pm (\hat{T}^2 - 4)^{1/2}]/2.$$

Since $\{[\hat{T} + (\hat{T}^2 - 4)^{1/2}/2\}\{\hat{T} - (\hat{T}^2 - 4)^{1/2}]/2\} = 1$, the roots are reciprocals of each other as required for the symplectic map. There are three cases:

(a) $\hat{T} > 2$; the roots are real and positive $(\lambda,\ 1/\lambda > 0)$;
(b) $2 > \hat{T} > -2$; the roots are complex and of magnitude one $(\lambda,\ 1/\lambda = \exp(\pm i\theta))$;
(c) $\hat{T} < -2$; the roots are real and negative $(\lambda,\ 1/\lambda < 0)$.

In case (a) we say that the periodic orbit is *hyperbolic*; in case (b) we say the periodic orbit is *elliptic*; and in case (c) we say that the periodic orbit is *hyperbolic with reflection*. Note that, in the linear approximation, cases

(a) and (c) lead typical nearby orbits to diverge exponentially from the periodic orbit (linear instability); while in case (b), in the linear approximation a nearby orbit remains nearby forever (linear stability). In the latter case, the nearby linearized orbit remains on an ellipse encircling the periodic orbit and circles around it at a rate $\theta/2\pi$ per iterate of $\mathbf{M}$. (Because the product of the two roots is one, in no case can the periodic orbit be an attractor, since that requires that the magnitude of *both* roots be less than one.)

7.1.4 Integrable systems

In the case where the Hamiltonian has no explicit time dependence, $H = H(\mathbf{p}, \mathbf{q})$, we have seen that Hamilton's equations imply that $\mathrm{d}H/\mathrm{d}t = 0$, and the energy $E = H(\mathbf{p}, \mathbf{q})$ is a conserved quantity. Thus, orbits with a given energy E are restricted to lie on the $(2N-1)$-dimensional energy surface $E = H(\mathbf{p}, \mathbf{q})$. A function $f(\mathbf{p}, \mathbf{q})$ is said to be a *constant of the motion* for a system with Hamiltonian H, if, as $\mathbf{p}(t)$ and $\mathbf{q}(t)$ evolve with time in accordance with Hamilton's equations, the value of the function f does not change, $f(\mathbf{p}, \mathbf{q}) = (\text{constant})$. For example, for time-independent Hamiltonians, H is a constant of the motion. More generally, differentiating $f(\mathbf{p}(t), \mathbf{q}(t))$ with respect to time, and assuming that there is no explicit time dependence of the Hamiltonian, we have

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}\cdot\frac{\partial f}{\partial \mathbf{p}} + \frac{\mathrm{d}q}{\mathrm{d}t}\cdot\frac{\partial f}{\partial \mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}\cdot\frac{\partial f}{\partial \mathbf{q}} - \frac{\partial H}{\partial \mathbf{q}}\cdot\frac{\partial f}{\partial \mathbf{p}}.$$

We call the expression appearing on the right-hand side of the second equality the *Poisson bracket* of f and H, and we abbreviate it as $[f, H]$, where

$$[g_1, g_2] \equiv \frac{\partial g_1}{\partial \mathbf{q}}\cdot\frac{\partial g_2}{\partial \mathbf{p}} - \frac{\partial g_1}{\partial \mathbf{p}}\cdot\frac{\partial g_2}{\partial \mathbf{q}}. \tag{7.17}$$

Note that $[g_1, g_2] = -[g_2, g_1]$. Thus the condition that f be a constant of the motion for a time-independent Hamiltonian is that its Poisson bracket with H be zero,

$$[f, H] = 0. \tag{7.18}$$

(The Hamiltonian is a constant of the motion since $[H, H] = 0$.)

A time-independent Hamiltonian system is said to be *integrable* if it has N *independent* global constants of the motion $f_i(\mathbf{p}, \mathbf{q})$, $i = 1, 2, \ldots, N$ (one of these is the Hamiltonian itself; we choose this to be the $i = 1$ constant, $f_1(\mathbf{p}, \mathbf{q}) \equiv H(\mathbf{p}, \mathbf{q})$), and, furthermore, if

$$[f_i, f_j] = 0, \tag{7.19}$$

for all i and j.

We already know that the Poisson bracket of f_i with f_1 is zero for all $i = 1, 2, \ldots, N$, since the f_i are constants of the motion (see Eq. (7.18)). If the condition (7.19) holds for all i and j, then we say that the N constants of the motion f_i are *in involution*. The constants of the motion f_i are 'independent' if no one of them can be expressed as a function of the $(N-1)$ other constants. The requirement that an integrable system has N independent constants of the motion implies that the trajectory of the system in the phase space is restricted to lie on the N-dimensional surface

$$f_i(\mathbf{p}, \mathbf{q}) = k_i, \tag{7.20}$$

$i = 1, 2, \ldots, N$, where k_i are N constants. The requirement that the N independent constants f_i be in involution (Eq. (7.19)) restricts the topology of the surface, Eq. (7.20), to be of a certain type: it must be an N-dimensional torus (as defined in Section 6.3). This is demonstrated in standard texts[2] and will not be shown here. For the case $N = 2$, an orbit on the torus is as shown in Figure 7.4(a).

Given an integrable system it is possible to introduce a canonical change of variables $(\mathbf{p}, \mathbf{q}) \to (\overline{\mathbf{p}}, \overline{\mathbf{q}})$ such that the new Hamiltonian $\overline{H}$ depends only on $\overline{\mathbf{p}}$ and not on $\overline{\mathbf{q}}$. One possibility is to choose the constants

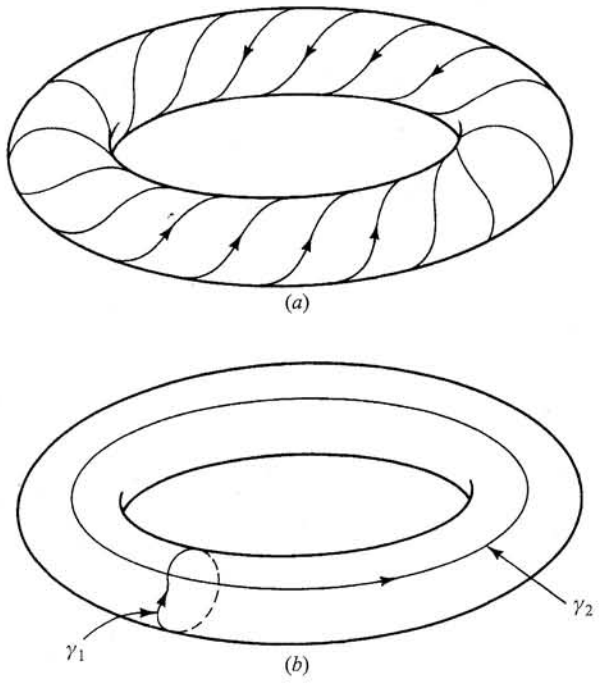


Figure 7.4 (a) Orbit on a 2-torus. (b) Two irreducible paths on a 2-torus.

of the motion themselves as the N components of $\overline{\mathbf{p}}$, $\bar{p}_i = f_i(\mathbf{p}, \mathbf{q})$. Since the f_i are constants, $d\overline{\mathbf{p}}/dt = \partial\overline{H}/\partial\overline{\mathbf{q}} = 0$ and hence $\overline{H} = \overline{H}(\overline{\mathbf{p}})$. In fact, we can construct many equivalent sets of constants of the motion by noting that any N independent functions of the N constants f_i could be used for the components of $\overline{\mathbf{p}}$ with the same result (namely, $d\overline{\mathbf{p}}/dt = \partial\overline{H}/\partial\overline{\mathbf{q}} = 0$). Of all these choices, one particular choice is especially convenient. This choice is the *action-angle variable* which we denote

$$(\overline{\mathbf{p}}, \overline{\mathbf{q}}) = (\mathbf{I}, \boldsymbol{\theta}),$$

where $\mathbf{I}$ is defined by

$$I_i = \frac{1}{2\pi}\oint_{\gamma_i} \mathbf{p} \cdot d\mathbf{q}, \tag{7.21}$$

$i = 1, 2, \ldots, N$. (In (7.21) the γ_i denote N irreducible paths on the N-torus, each of which wrap around the torus in N angle directions that can be used to parameterize points on the torus (see Figure 7.4(b)). Note that deformations of the paths γ_i on the torus do not change the values of the integrals in (7.21) by virtue of the Poincaré–Cartan theorem, Eq. (7.9c).

The position coordinate $\boldsymbol{\theta}$, canonically conjugate to the momentum coordinate $\mathbf{I}$, is angle-like because, on one circuit following one of the irreducible paths γ_i around the torus, the variable θ_i increases by 2π, while the other variables θ_j with $j \neq i$ return to their original values. In order to see how this result is obtained, we first write the change of variables $(\mathbf{p}, \mathbf{q}) \rightarrow (\mathbf{I}, \boldsymbol{\theta})$ in terms of the generating function (Eq. (7.10)).

$$\boldsymbol{\theta} = \partial S(\mathbf{I}, \mathbf{q})/\partial\mathbf{I}, \tag{7.22a}$$

$$\mathbf{p} = \partial S(\mathbf{I}, \mathbf{q})/\partial\mathbf{q}. \tag{7.22b}$$

Let $\Delta_i\boldsymbol{\theta}$ denote the change of $\boldsymbol{\theta}$ on one circuit around the irreducible path γ_i, and $\Delta_i S$ denote the corresponding change of the generating function on one circuit. From (7.22b)

$$\Delta_i S = \oint_{\gamma_i} \mathbf{p} \cdot d\mathbf{q} = 2\pi \boldsymbol{I}_i.$$

From (7.22a)

$$\Delta_i\boldsymbol{\theta} = \frac{\partial}{\partial\mathbf{I}}\Delta_i S = 2\pi\frac{\partial}{\partial\mathbf{I}} I_i$$

or

$$\Delta_i\theta_j = 2\pi\delta_{ij} \tag{7.23}$$

which is the desired result (here $\delta_{ij} \equiv 1$ if $i = j$ and $\delta_{ij} \equiv 0$ if $i \neq j$).

The new Hamiltonian in action-angle coordinates is, by construction, independent of $\boldsymbol{\theta}$, and hence Hamilton's equations reduce to

$$d\mathbf{I}/dt = 0,$$
$$d\boldsymbol{\theta}/dt = \partial \overline{H}(\mathbf{I})/\partial \mathbf{I} \equiv \boldsymbol{\omega}(\mathbf{I}).$$

The solution of these equations is $\mathbf{I}(t) = \mathbf{I}(0)$ and

$$\boldsymbol{\theta}(t) = \boldsymbol{\theta}(0) + \boldsymbol{\omega}(\mathbf{I})t. \tag{7.24}$$

Thus we can interpret $\boldsymbol{\omega}(\mathbf{I}) = \partial \overline{H}(\mathbf{I})/\partial \mathbf{I}$ as an angular velocity vector specifying trajectories on the N-torus. As in our discussion in Chapter 6, trajectories on a torus are N frequency quasiperiodic if there is no vector of integers $\mathbf{m} = (m_1, m_2, \ldots, m_N)$ such that

$$\mathbf{m} \cdot \boldsymbol{\omega} = 0, \tag{7.25}$$

except when $\mathbf{m}$ is the trivial vector all of whose components are zero. Assuming a typical smooth variation of $\overline{H}$ with $\mathbf{I}$ the condition $\mathbf{m} \cdot \boldsymbol{\omega} = 0$ with nonzero $\mathbf{m}$ is only satisfied for a *countable* set of $\mathbf{I}$. Thus, if one picks a point randomly with uniform probability in phase space, the probability is 1 that the point chosen will be on a torus for which the orbits are N frequency quasiperiodic and *fill up the torus*. Thus, for integrable systems, we can view the phase space as being completely occupied by N-tori almost all of which are in turn filled by N frequency quasiperiodic orbits. In contrast with the case of N frequency quasiperiodicity is the case of periodic motion, where orbits on the N-torus close on themselves (Figure 6.5). In this case

$$\boldsymbol{\omega} = \mathbf{m}\omega_0, \tag{7.26}$$

where $\mathbf{m}$ is again a vector of integers and ω_0 is a scalar. The orbit in this case closes on itself after m_1 circuits in θ_1, m_2 circuits in $\theta_2, \ldots$. (Alternatively to (7.26), the condition for a periodic orbit can also be stated as requiring that $(N-1)$-independent relations of the form (7.25) hold.[4]) Again assuming typical smooth variation of $\overline{H}$ with $\mathbf{I}$, we have that for integrable systems the set of tori which satisfy (7.26) and hence have periodic orbits, while having zero Lebesgue measure (i.e., zero phase space volume), is *dense in the phase space*. Thus, arbitrarily near any torus on which there is N frequency quasiperiodicity there are tori on which the orbits are periodic.

We now give an example of the procedure used for the reduction of an integrable system to action-angle variables. This procedure is based on the *Hamilton–Jacobi equation* obtained by combining (7.10) and (7.11),

$$H\left(\frac{\partial S(\mathbf{I}, \mathbf{q})}{\partial \mathbf{q}}, \mathbf{q}\right) = \overline{H}(\mathbf{I}). \tag{7.27}$$

This equation may be regarded as a first-order partial differential equation for the generating function $S(\mathbf{I}, \mathbf{q})$.

Example: We consider a one-degree-of-freedom Hamiltonian,

$$H(p, q) = p^2/(2m) + V(q),$$

where $V(q)$ is of the form shown in Figure 7.5. From Eq. (7.21) we have

$$I = \frac{1}{\pi}\int_{q_1}^{q_2} \{2m[E - V(q)]\}^{1/2}\, \mathrm{d}q. \tag{7.28}$$

In the case of a harmonic oscillator, $V(q) = \frac{1}{2}m\Omega^2 q^2$, we have $q_2 = -q_1 = [2E/(m\Omega^2)]^{1/2}$, and the integral for I yields $I = E/\Omega$. Thus we have

$$\overline{H}(I) = \Omega I.$$

For this case $\omega(I) = \mathrm{d}\overline{H}/\mathrm{d}I = \Omega$ is independent of I, and (7.24) becomes

$$\theta(t) = \theta(0) + \Omega t.$$

From (7.27) we have for the harmonic oscillator

$$\partial S/\partial q = [2m(\Omega I - \tfrac{1}{2}m\Omega^2 q^2)]^{1/2},$$

which on integration and application of (7.22) gives

$$q = (2I/m\Omega)^{1/2} \cos\theta,$$

$$p = -(2mI\Omega)^{1/2} \sin\theta.$$

The trajectory in p, q phase space is an ellipse on which the orbit circulates one time every period of oscillation $2\pi/\Omega$ (Figure 7.6(a)). (Since $N = 1$ we have a 'one-dimensional torus', namely a closed curve. For $N > 1$ we typically have N frequencies and quasiperiodic motion.) The harmonic oscillator is exceptional in that $\omega(I)$ is independent of I. As

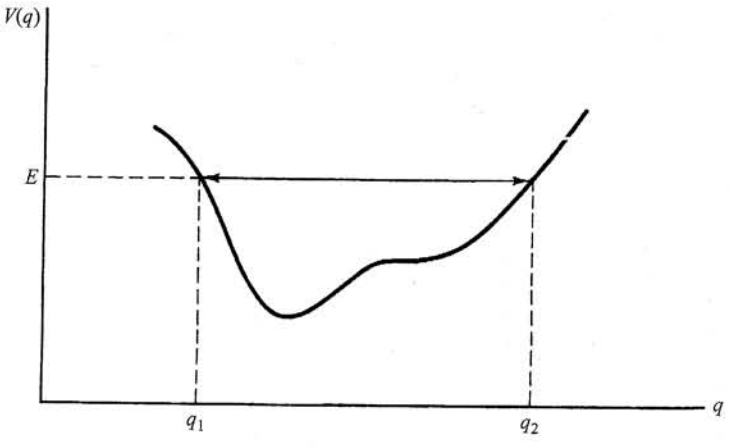


Figure 7.5 Particle of energy E in a potential well $V(q)$.

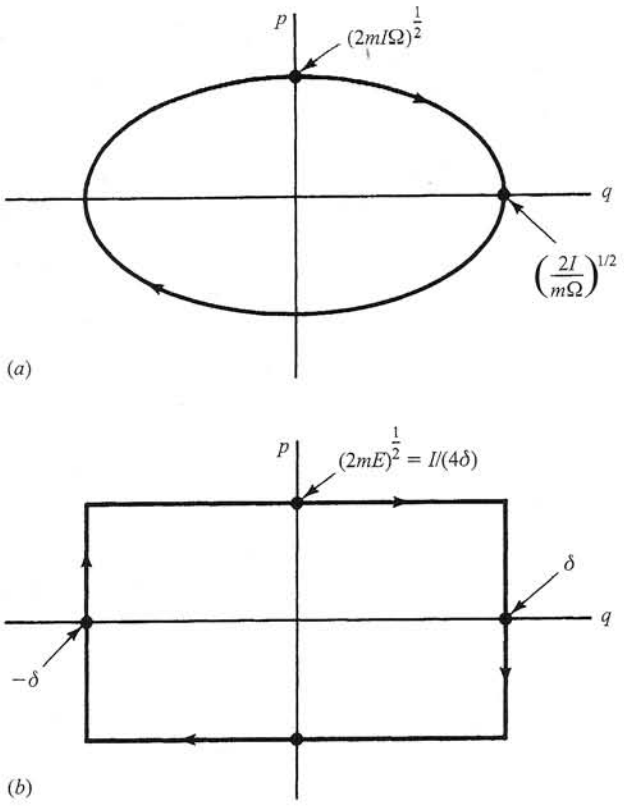


Figure 7.6 Phase space paths followed by (*a*) the harmonic oscillator and (*b*) a particle bouncing between hard walls at $q = \pm\delta$.

an example of the more typical situation where $\omega(I)$ depends on I consider the case of a hard-wall potential: $V(q) = 0$ for $|q| < \delta$ and $V(q) = +\infty$ for $|q| > \delta$. In this case the trajectory in phase space is as shown in Figure 7.6(*b*). The integral for I, Eq. (7.28), is just $(2\pi)^{-1}$ times the phase space area in the rectangle; $I = 4(2\pi)^{-1}(2mE\delta^2)^{1/2}$. Thus $\overline{H}(\mathbf{I}) = (\pi I)^2/8m\delta^2$ and $\omega(I) = \pi^2 I/4m\delta^2$ which increases linearly with I.

7.2 Perturbation of integrable systems

7.2.1 The KAM theorem

We next address a very fundamental question concerning Hamiltonian systems; namely, how prevalent is integrability? One extreme conjecture is that integrability generally applies, and whatever difficulty we might encounter in obtaining the solution to any given probem only arises because we are not clever enough to determine the N independent

constants of the motion which must surely exist. Another conjecture, which is essentially the opposite of this, given any integrable Hamiltonian $H_0(\mathbf{p}, \mathbf{q})$, if we alter it slightly by the addition of a perturbation

$$H(\mathbf{p}, \mathbf{q}) = H_0(\mathbf{p}, \mathbf{q}) + \varepsilon H_1(\mathbf{p}, \mathbf{q}), \tag{7.29}$$

then we should expect that for a typical form of the perturbation, $H_1(\mathbf{p}, \mathbf{q})$, all the constants of the motion for the integrable system, $H_0(\mathbf{p}, \mathbf{q})$, except for the energy constant, $E = H(\mathbf{p}, \mathbf{q})$, are immediately destroyed as soon as $\varepsilon \neq 0$. Presumably, if this second conjecture were to hold, then, for small ε, orbits would initially approximate the orbits of the integrable system, staying close to the unperturbed N-tori that exist for $\varepsilon = 0$ for some time. Eventually, however, the orbit, if followed for a long enough time, could ergodically wander anywhere on the energy surface. These two opposing views both have some support in experimental observation. On the one hand, the solar system appears to have been fairly stable. In particular, ever since its formation the Earth has been in a position relative to the position of the Sun such that its climate has been conducive to life. Thus, in spite of the perturbation caused by the gravitational pull of other planets (particularly that of Jupiter), the Earth's orbit has behaved as we would have expected had we neglected all the other planets. (In that case the system is integrable, and we obtain the elliptical Kepler orbit of the Earth around the Sun.) On the other hand, in support of the second conjecture, we have the amazing success of the predictions of statistical mechanics. In statistical mechanics one considers a Hamiltonian system with a large number of degrees of freedom ($N \gg 1$), and then makes the fundamental ansatz that at any given time the system is equally likely to be located at any point on the energy surface (the motion is ergodic on the energy surface). This would not be possible if there were additional constants of the motion constraining the orbit of the system. The success of statistical mechanics in virtually every case to which it may reasonably be applied can be interpreted as evidence supporting the validity of its fundamental ansatz in a wide variety of systems with $N \gg 1$.

Given the discussion above, it should not be too surprising to find out that the true situation lies somewhere between the two extremes that we have discussed. The resolution of the basic question of how prevalent integrability is has come only with the rigorous mathematical work of Kolmogorov, Arnold and Moser (KAM) and with the subsequent extensive computer studies of chaos and integrability in Hamilitonian systems. The basic question considered by Kolmogorov, Arnold and Moser was what happens when an integrable Hamiltonian is perturbed, Eq. (7.29). The research was initiated by Kolmogorov (1954) who conjectured what would happen with the addition of the perturbation. He also outlined an inge-

nious method which he felt could be used to prove his conjecture. The actual carrying out of this program, accomplished by Arnold and Moser (see Arnold (1963) and Moser (1973)), was quite difficult. The result they obtained is called the KAM theorem. We shall only briefly indicate some of the sources of the difficulty, and then state the main result.

We express (7.29) in the action-angle variables $(\mathbf{I}, \boldsymbol{\theta})$ of the unperturbed Hamiltonian H_0,

$$H(\mathbf{I}, \boldsymbol{\theta}) = H_0(\mathbf{I}) + \varepsilon H_1(\mathbf{I}, \boldsymbol{\theta}). \tag{7.30}$$

We are interested in determining whether this perturbed Hamiltonian has N-dimensional tori to which its orbits are restricted. If there are tori, there is a new set of action-angle variables $(\mathbf{I}', \boldsymbol{\theta}')$ such that

$$H(\mathbf{I}, \boldsymbol{\theta}) = H'(\mathbf{I}'),$$

where, in terms of the generating function S, we have using (7.10)

$$\mathbf{I} = \frac{\partial S(\mathbf{I}', \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \quad \boldsymbol{\theta}' = \frac{\partial S(\mathbf{I}', \boldsymbol{\theta})}{\partial \mathbf{I}'}. \tag{7.31}$$

The Hamilton–Jacobi equation for S is

$$H\left(\frac{\partial S}{\partial \boldsymbol{\theta}}, \boldsymbol{\theta}\right) = H'(\mathbf{I}'). \tag{7.32}$$

One approach to solving (7.32) for S might be to look for a solution in the form of a power series in ε,

$$S = S_0 + \varepsilon S_1 + \varepsilon^2 S_2 + \cdots. \tag{7.33}$$

For S_0 we use $S_0 = \mathbf{I}' \cdot \boldsymbol{\theta}$ which when substituted in (7.31) gives $\mathbf{I} = \mathbf{I}'$, $\boldsymbol{\theta} = \boldsymbol{\theta}'$, corresponding to the original action-angle variables applicable for $\varepsilon = 0$. Substituting the series (7.33) for S in (7.32) gives,

$$H_0(\mathbf{I}' + \varepsilon \partial S_1/\partial \boldsymbol{\theta} + \varepsilon^2 \partial S_2/\partial \boldsymbol{\theta} + \cdots) + \varepsilon H_1(\mathbf{I}' + \varepsilon \partial S_1/\partial \boldsymbol{\theta} + \cdots, \boldsymbol{\theta}) = H'(\mathbf{I}'). \tag{7.34}$$

Expanding (7.34) for small ε and only retaining first-order terms, we have

$$H_0(\mathbf{I}') + \varepsilon \frac{\partial H_0}{\partial \mathbf{I}'} \cdot \frac{\partial S_1}{\partial \boldsymbol{\theta}} + \varepsilon H_1(\mathbf{I}', \boldsymbol{\theta}) = H'(\mathbf{I}'). \tag{7.35}$$

We next express $H_1(\mathbf{I}', \boldsymbol{\theta})$ and $S_1(\mathbf{I}', \theta)$ as Fourier series in the angle vector $\boldsymbol{\theta}$,

$$H_1 = \sum_{\mathbf{m}} H_{1,\mathbf{m}}(\mathbf{I}') \exp(\mathrm{i}\mathbf{m} \cdot \boldsymbol{\theta}),$$

$$S_1 = \sum_{\mathbf{m}} S_{1,\mathbf{m}}(\mathbf{I}') \exp(\mathrm{i}\mathbf{m} \cdot \boldsymbol{\theta}),$$

where $\mathbf{m}$ is an N-component vector of integers. Substituting these Fourier series in (7.35), we obtain

$$S_1 = \mathrm{i} \sum_{\mathbf{m}} \frac{H_{1,\mathbf{m}}(\mathbf{I}')}{\mathbf{m} \cdot \boldsymbol{\omega}_0(\mathbf{I}')} \exp(\mathrm{i}\mathbf{m} \cdot \boldsymbol{\theta}) \tag{7.36}$$

where $\boldsymbol{\omega}_0(\mathbf{I}) \equiv \partial H_0(\mathbf{I})/\partial \mathbf{I}$ is the unperturbed N-dimensional frequency vector for the torus corresponding to action $\mathbf{I}$. One question is that of whether the infinite sum (7.36) converges. This same question also arises in taking (7.34) to higher order in ε to determine successively the other terms, S_2, S_3, etc., appearing in the series (7.33).

This problem is precisely the 'problem of small denominators' encountered in Section 6.2 where we treated frequency locking of quasi-periodic orbits for dissipative systems. In particular, clearly (7.36) does not work for values of $\mathbf{I}$ for which $\mathbf{m} \cdot \boldsymbol{\omega}_0(\mathbf{I}) = 0$ for some value of $\mathbf{m}$. These $\mathbf{I}$ define *resonant tori* of the unperturbed system. (These resonant tori are typically destroyed by the perturbation for any small $\varepsilon > 0$.) We emphasize that the resonant tori are dense in the phase space of the unperturbed Hamiltonian. On the other hand, there is still a large set of 'very nonresonant' tori. These are tori for which $\boldsymbol{\omega}$ satisfies the condition

$$|\mathbf{m} \cdot \boldsymbol{\omega}| > K(\boldsymbol{\omega})|\mathbf{m}|^{-(N+1)}, \tag{7.37}$$

for *all* integer vectors $\mathbf{m}$ (except the zero vector). Here $|\mathbf{m}| \equiv |m_1| + |m_2| + \cdots + |m_N|$, and $K(\boldsymbol{\omega}) > 0$ is a number independent of $\mathbf{m}$. The set of N-dimensional vectors $\boldsymbol{\omega}$ which do not satisfy (7.37) has zero Lebesgue measure in $\boldsymbol{\omega}$-space, and thus the 'very nonresonant' tori are, in this sense, very common. For $\boldsymbol{\omega}$ satisfying (7.37), the series (7.36), and others of similar form giving $S_2, S_3, \ldots$, converges. This follows if we assume that H_1 is analytic in $\boldsymbol{\theta}$ which implies that $H_{1,\mathbf{m}}$ decreases exponentially with m; i.e., $|H_{1,\mathbf{m}}| < (\text{constant}) \exp(-\alpha|\mathbf{m}|)$ for some constant $\alpha > 0$. (Refer to the discussion in Section 6.2.)

Even given that all the terms $S_1, S_2, \ldots$ exist and can be found, we would still be faced with the problem of whether there is convergence of the successive approximations to S obtained by taking more and more terms in the series (7.33). Actually, the scheme we have outlined (wherein S is expanded in a straightforward series in ε, Eq. (7.33)) is too crude, and the proof of the KAM theorem relies on a more sophisticated method of successive approximations which has much faster convergence properties. We shall not pursue this discussion further. Suffice it to say that the KAM theorem essentially states that under very general conditions for small ε 'most' (in the sense of the Lebesgue measure of the phase space) of the tori of the unperturbed integrable Hamiltonian survive. We say that a torus of the unperturbed system with frequency vector $\boldsymbol{\omega}_0$ 'survives' perturbation if there exists a torus of the perturbed ($\varepsilon \neq 0$) system which has a

frequency vector $\boldsymbol{\omega}(\varepsilon) = k(\varepsilon)\boldsymbol{\omega}_0$, where $k(\varepsilon)$ goes continuously to 1 as $\varepsilon \to 0$, and such that the perturbed toroidal surface with frequency $\boldsymbol{\omega}(\varepsilon)$ goes continuously to the unperturbed torus as $\varepsilon \to 0$. Thus, writing $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_N)$, the unperturbed and perturbed frequency vectors $\boldsymbol{\omega}_0$ and $\boldsymbol{\omega}(\varepsilon)$ have the same frequency ratios of their components, $\omega_{0j}/\omega_{01} = \omega_j(\varepsilon)/\omega_1(\varepsilon)$ for $j = 2, 3, \ldots, N$. According to the KAM theorem, for small ε, the perturbed system's phase space volume (Lebesgue measure) not occupied by surviving tori is small and approaches zero as ε approaches zero.

Note, however, that, since the resonant tori on which $\mathbf{m} \cdot \boldsymbol{\omega}_0(\mathbf{I}) = 0$ are dense, we expect that, arbitrarily near surviving tori of the perturbed system, there are regions of phase space where the orbits are not on surviving tori. We shall, in fact, see that these regions are occupied by chaotic orbits as well as new tori and elliptic and hyperbolic periodic orbits all created by the perturbation. In the language of Section 3.10, the set in the phase space occupied by surviving perturbed tori is a fat fractal. That is, it is the same kind of set on which values of the parameter r yielding chaos for the logistic map (2.10) exist and on which values of the parameter w in the circle map (6.11) yield two frequency quasiperiodic orbits (for $k < 1$). The Poincaré–Birkhoff theorem discussed in the next subsection sheds light on the exceedingly complex and intricate situation which arises in the vicinity of resonant tori when an integrable system is perturbed.

7.2.2 The fate of resonant tori

We have seen that most tori survive small perturbation. The resonant tori, however, do not. What happens to them? To simplify the discussion of this question we consider the case of a Hamiltonian system described by a two-dimensional area preserving map. We can view this map as arising from a surface of section for a time-independent Hamiltonian with $N = 2$, as illustrated for the integrable case in Figure 7.7. The tori of the integrable system intersect the surface of section in a family of nested closed curves. Without ioss of generality we can take these curves to be concentric circles represented by polar coordinates (r, ϕ). In this case we obtain a map $(r_{n+1}, \phi_{n+1}) = \mathbf{M}_0(r_n, \phi_n)$,

$$\left.\begin{aligned} r_{n+1} &= r_n, \\ \phi_{n+1} &= [\phi_n + 2\pi R(r_n)] \text{ modulo } 2\pi. \end{aligned}\right\} \tag{7.38}$$

Here $R(r)$ is the ratio of the frequencies ω_1/ω_2 where $\boldsymbol{\omega}_0 = (\omega_1, \omega_2) = (\partial H_0/\partial I_1, \partial H_0/\partial I_2)$ for the torus which intersects the surface of section in a circle of radius r, and we have taken the surface of section to be $\theta_2 = (\text{const.})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2)$ are the angle variables

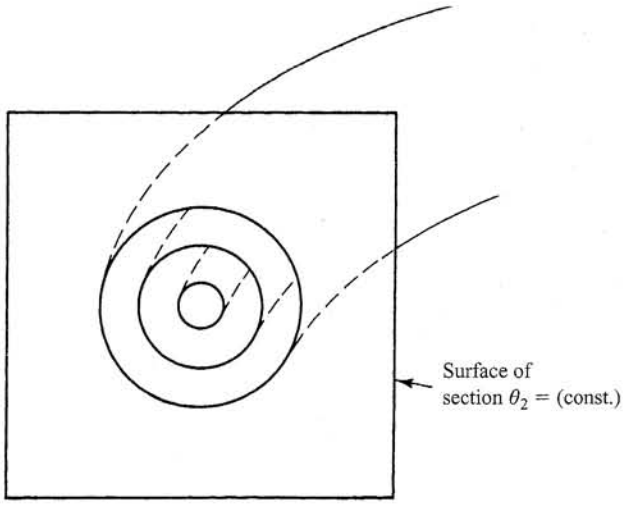


Figure 7.7 Surface of section for an integrable system.

conjugate to the actions $\mathbf{I} = (I_1, I_2)$. The quantity ϕ_n is the value of θ_1 at the nth piercing of the surface of section by the orbit. On a resonant torus the rotation number $R(r)$ is rational:

$$R = \omega_1/\omega_2 = \tilde{p}/\tilde{q};\ \tilde{q}\omega_1 - \tilde{p}\omega_2 = 0,$$

where $\tilde{p}$ and $\tilde{q}$ are integers which do not have a common factor. At the radius $r = \hat{r}(\tilde{p}/\tilde{q})$ corresponding to $R(\hat{r}) = \tilde{p}/\tilde{q}$ we have that application of the map (7.38) $\tilde{q}$ times returns every point on the circle to its original position,

$$\mathbf{M}_0^{\tilde{q}}(r, \phi) = [r, (\phi + 2\pi\tilde{p}) \text{ modulo } 2\pi] = (r, \phi).$$

Now we consider a perturbation of the integrable Hamiltonian H_0, Eq. (7.30). This will perturb the map $\mathbf{M}_0$ to a new area preserving map $\mathbf{M}_\varepsilon$ which differs slightly from $\mathbf{M}_0$,

$$\begin{aligned} r_{n+1} &= r_n + \varepsilon g(r_n, \phi_n), \\ \phi_{n+1} &= [\phi_n + 2\pi R(r_n) + \varepsilon h(r_n, \phi_n)] \text{ modulo } 2\pi. \end{aligned} \tag{7.39}$$

We have seen that, on the intersection $r = \hat{r}(\tilde{p}/\tilde{q})$ of the resonant torus with the surface of section, every point is a fixed point of $\mathbf{M}_0^{\tilde{q}}$. We now inquire, what happens to this circle when we add the terms proportional to ε in (7.39)? Assume that $R(r)$ is a smoothly increasing function of r in the vicinity of $r = \hat{r} = (\tilde{p}/\tilde{q})$. (Equation (7.38) is called a 'twist map' if $R(r)$ increases with r.) Then for the unperturbed map we can choose a circle at $r = r_+ > \hat{r}(\tilde{p}/\tilde{q})$ which is rotated by $\mathbf{M}_0^{\tilde{q}}$ in the direction of increasing ϕ (i.e., counterclockwise) and a circle at $r = r_- < \hat{r}(\tilde{p}/\tilde{q})$ which is rotated

by $\mathbf{M}_0^{\tilde{q}}$ in the direction of decreasing ϕ (i.e., clockwise). The circle $r = \hat{r}(\tilde{p}/\tilde{q})$ is not rotated at all. See Figure 7.8(a). If ε is sufficiently small, then $\mathbf{M}_0^{\tilde{q}}$ still maps all the points initially on the circle $r = r_-$ to new positions whose ϕ coordinate is clockwise displaced from its initial position (the radial coordinate, after application of the perturbed map, will in general differ from r_-). Similarly, for small enough ε all points on r_+

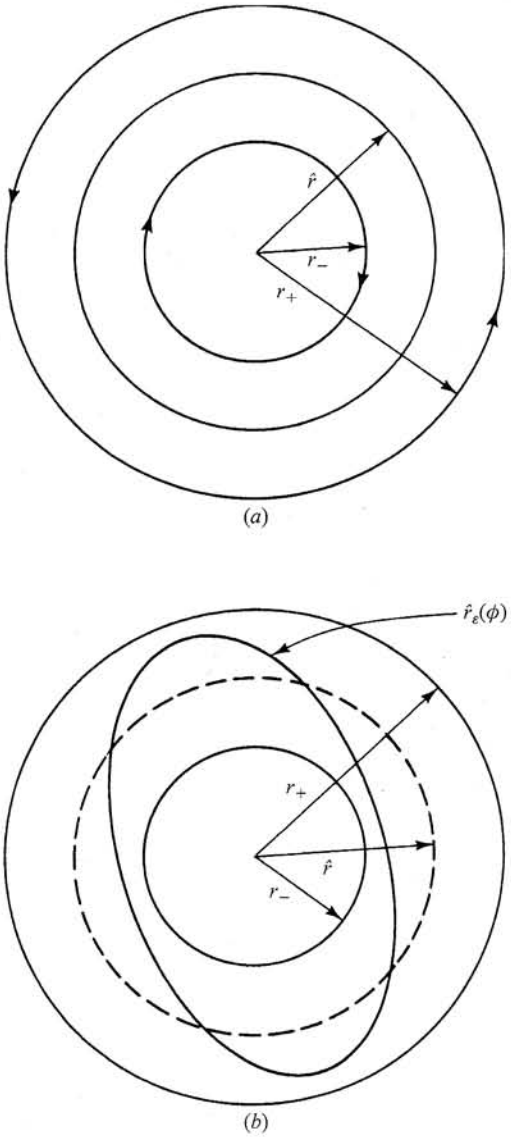


Figure 7.8 (a) Three invariant circles of the unperturbed map. (b) The curve $r = \hat{r}_\varepsilon(\phi)$.

will be counterclockwise displaced. Given this situation, we have that for any given fixed value of ϕ, as r increases from r_- to r_+, the value of the angle that the point (r, ϕ) maps to increases from below ϕ to above ϕ. Hence, there is a value of r between r_+ and r_- for which the angle is not changed. We conclude that, for the perturbed map, there is a closed curve, $r = \hat{r}_\varepsilon(\phi)$, lying between $r_+ \geqslant r \geqslant r_-$ and close to $r = \hat{r}(\tilde{p}, \tilde{q})$, on which points are mapped by $\mathbf{M}_0^{\tilde{q}}$ purely in the radial direction. This is illustrated in Figure 7.8(b). We now apply the map $\mathbf{M}_0^{\tilde{q}}$ to this curve obtaining a new curve $r = \hat{r}'_\varepsilon(\phi)$. The result is shown schematically in Figure 7.9. Since $\mathbf{M}_\varepsilon$ is area preserving the areas enclosed by the curve $\hat{r}_\varepsilon(\phi)$ and by the curve $\hat{r}'_\varepsilon(\phi)$ are equal. Hence, these curves must intersect. Generically these curves intersect at an even number of distinct points. (Here by use of the word generic we mean to rule out cases where the two curves are tangent or else (as in the integrable case) coincide. These nongeneric cases can be destroyed by small changes in ε or in the form of the perturbing functions g and h in Eq. (7.39).) The intersections of $\hat{r}_\varepsilon$ and $\hat{r}'_\varepsilon$ correspond to fixed points of $\mathbf{M}_\varepsilon^{\tilde{q}}$. Thus, the circle of fixed points $r = \hat{r}(\tilde{p}/\tilde{q})$ for the unperturbed map $\mathbf{M}_0^{\tilde{q}}$ is replaced by a finite number of fixed points when the map is perturbed.

What is the character of these fixed points of $\mathbf{M}_\varepsilon^{\tilde{q}}$? Recall that for $r > \hat{r}_\varepsilon$ points are rotated counterclockwise by $\mathbf{M}_\varepsilon^{\tilde{q}}$. Also recall that $\mathbf{M}_\varepsilon^{\tilde{q}}$ maps $\hat{r}_\varepsilon$ to $\hat{r}'_\varepsilon$. Thus, we have the picture shown in Figure 7.10, where the arrows indicate the displacements experienced by points as a result of applying the map $\mathbf{M}_\varepsilon^{\tilde{q}}$. We see that elliptic and hyperbolic fixed points of

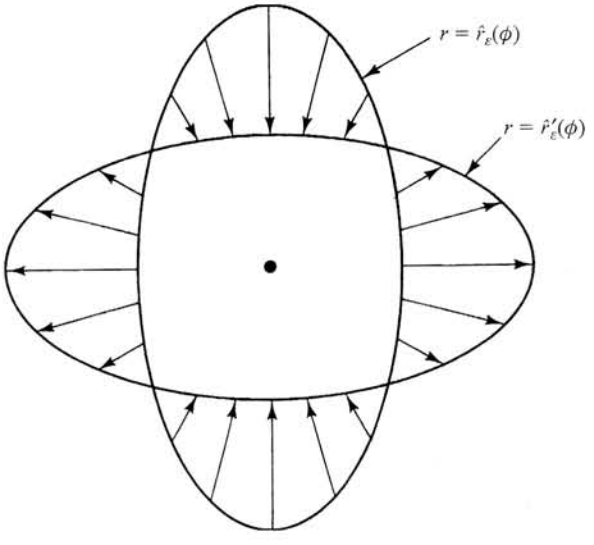


Figure 7.9 Points on the curve $\hat{r}_\varepsilon(\phi)$ map under $\mathbf{M}_\varepsilon^{\tilde{q}}$ purely radially to the curve $\hat{r}'_\varepsilon(\phi)$.

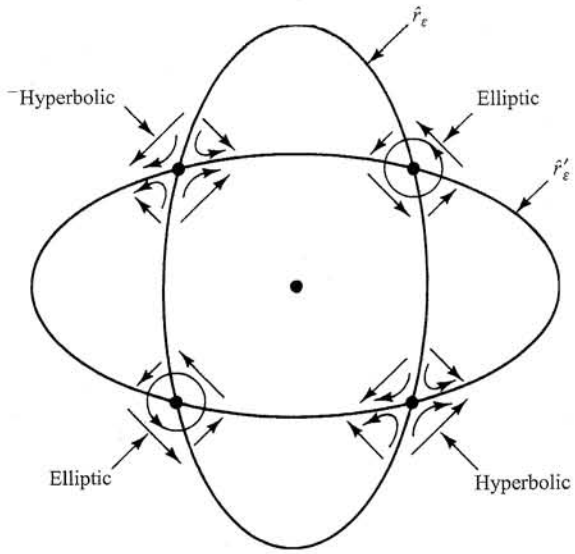


Figure 7.10 Illustration of the Poincaré–Birkhoff theorem. Recall that, in the case of an elliptic fixed point of $\mathbf{M}_\varepsilon^{\tilde{q}}$ the eigenvalues of the linearized map are of the form $\exp(\mathrm{i}\theta)$ and nearby points remain near the elliptic fixed point, lying on an approximate ellipse around it, and rotating around the fixed point on average θ radians per iterate of $\mathbf{M}_\varepsilon^{\tilde{q}}$.

$\mathbf{M}_\varepsilon^{\tilde{q}}$ alternate. Hence, perturbation of the resonant torus with rational rotation number $\tilde{p}/\tilde{q}$ results in an equal number of elliptic and hyperbolic fixed points of $\mathbf{M}_\varepsilon^{\tilde{q}}$. This result is known as the Poincaré–Birkhoff theorem (Birkhoff, 1927). Since fixed points of $\mathbf{M}_\varepsilon^{\tilde{q}}$ necessarily are on period $\tilde{q}$ orbits of $\mathbf{M}_\varepsilon$, we see that there are $\tilde{q}$ (or a multiple of $\tilde{q}$) elliptic fixed points of $\mathbf{M}_\varepsilon^{\tilde{q}}$ and the same number of hyperbolic fixed points. Thus, for example, the two elliptic fixed points of $\mathbf{M}_\varepsilon^{\tilde{q}}$ shown in Figure 7.10 might be a single periodic orbit of $\mathbf{M}_\varepsilon$ of period two (and similarly for the two hyperbolic fixed points in the figure). Thus, in this case, we have $\tilde{q} = 2$. Near each resonant torus of the unperturbed map we can expect a structure of elliptic and hyperbolic orbits to appear as illustrated schematically in Figures 7.11(a) and (b) where we only include the $\tilde{q} = 3$ and the $\tilde{q} = 4$ resonances.

Points near the elliptic fixed points rotate around them as shown by the linear theory (cf. Section 7.1.3). *Very* near an elliptic fixed point the linear approximation is quite good, and in such a small neighborhood the map can again be put in the form of Eq. (7.39). Thus, if we examine the small region around one of the elliptic points of a periodic orbit, such as the area indicated by the dashed box in Figure 7.11(b), then what we will see is qualitatively similar to what we see in Figure 7.11(b) itself. Thus, surrounding an elliptic point there are encircling KAM curves between which are destroyed resonant KAM curves that have been replaced by elliptic and hyperbolic periodic orbits. Furthermore, this repeats *ad infinitum*, since any elliptic point has surrounding elliptic points of destroyed

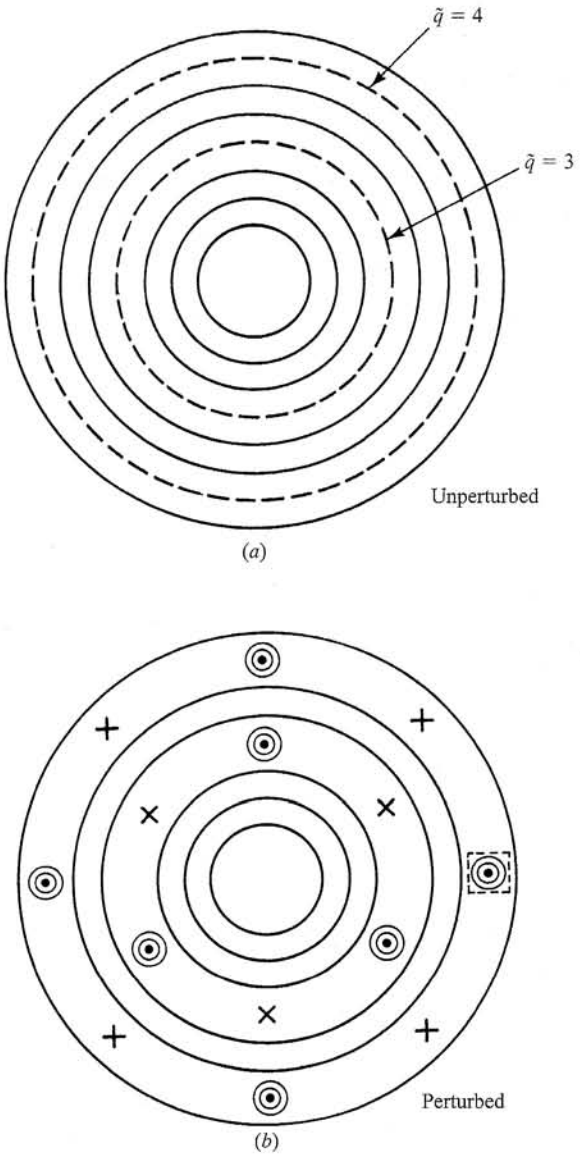


Figure 7.11 Perturbation of $\tilde{q} = 3$ and $\tilde{q} = 4$ resonant tori.

resonances which themselves have elliptic points of destroyed resonances, and so on.

What influence do the hyperbolic orbits created from the destroyed resonant tori have on the dynamics? If we follow the stable and unstable manifolds emanating from the hyperbolic points, they typically result in heteroclinic intersections, as shown in Figure 7.12. As we have seen in

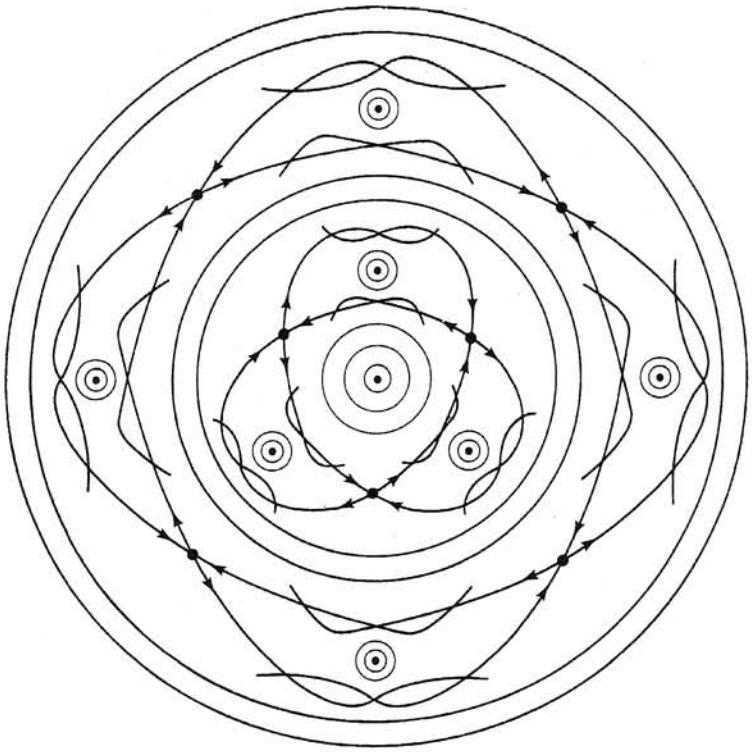


Figure 7.12 Stable and unstable manifolds of hyperbolic periodic orbits.

Chapter 4 (see Figure 4.10(d)), one such heteroclinic intersection between the stable and unstable manifolds of two hyperbolic points implies an infinite number of intersections between them.[5] Furthermore (as we have discussed for the homoclinic case, Figure 4.11), this also implies the presence of horseshoe type dynamics and hence chaos. Thus, not only do we have a dense set of destroyed resonance regions containing elliptic and hyperbolic orbits, but now we find that these regions of destroyed resonances also have embedded within them chaotic orbits. Furthermore, this repeats on all scales as we successively magnify regions around elliptic points. A very fascinating and intricate picture indeed!

7.3 Chaos and KAM tori in systems describable by two-dimensional Hamiltonian maps

Numerical examples clearly show the general phenomenology described for small perturbations of integrable systems in the previous section. In addition, numerical examples give information concerning what occurs when the perturbations are not small. Such information in turn points the

way for theories applicable in the far-from-integrable regime. The clearest and easiest numerical experiments are those that result in a two-dimensional map (a two-dimensional Poincaré surface of section).

7.3.1 The standard map

As an example, we consider the standard map, Eq. (7.15), which results from periodic impulsive kicking of the rotor in Figure 7.3. Setting the kicking strength to zero, $K = 0$, the standard map becomes

$$\theta_{n+1} = (\theta_n + p_n) \text{ modulo } 2\pi, \tag{7.40a}$$

$$p_{n+1} = p_n. \tag{7.40b}$$

This represents an integrable case. The intersections of the tori in the (θ, p) surface of section are just the lines of constant p (according to (7.40b) p is a constant of the motion). On each such line the orbit is given by $\theta_n = (\theta_0 + np_0)$ modulo 2π, and, if $p_0/2\pi$ is an irrational number, a single orbit densely fills the line $p = p_0$. If $p_0/2\pi$ is a rational number, then orbits on the line return to themselves after a finite number of iterates (the unperturbed orbit is periodic), and we have a resonant torus.

Increasing K slightly from zero introduces a small nonintegrable perturbation to the integrable case (7.40). Figure 7.13 shows plots of $\bar{p} \equiv p$ modulo 2π versus θ modulo 2π resulting from iterating a number of different initial conditions for a long time and for various values of K. If the initial condition is on an invariant torus it traces out the closed curve corresponding to the torus. If the initial condition yields a chaotic orbit, then it will wander throughout an area densely filling that area. We see that, for the relatively small perturbation, $K = 0.5$, Figure 7.13(a), there are many KAM tori running roughly horizontally from $\theta = 0$ to $\theta = 2\pi$. These tori are those that originate from the nonresonant tori of the unperturbed system ($p = p_0$, $p_0/2\pi$ irrational) and have survived the perturbation. Also, clearly seen in Figure 7.13(a) are tori, created by the perturbations nested around elliptic periodic orbits originating from resonant tori. In particular, the period one elliptic orbits, $(\theta, p) = (\pi, 0)$ and $(\theta, p) = (\pi, 2\pi)$, and the period two elliptic orbit, $(0, \pi) \rightleftarrows (\pi, \pi)$, are clearly visible. We call the structure surrounding a period $\tilde{q}$ elliptic periodic orbit a *period $\tilde{q}$ island chain*.

An important property of two-dimensional smooth area preserving maps is that the area bounded by two invariant KAM curves is itself invariant. This is illustrated in Figure 7.14 where we show two invariant curves (tori) bounding a shaded annular shaped region. Since the two curves are invariant and areas are preserved, the shaded region must map into itself. Thus, while there may be chaotic orbits sandwiched between

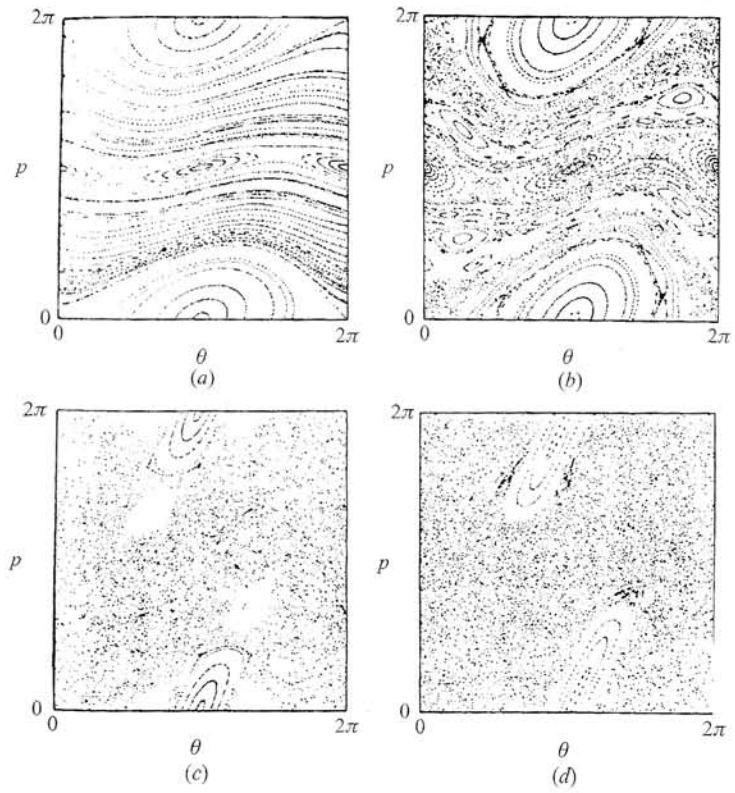


Figure 7.13 Plots of p modulo 2π for four values of K: (*a*) $K = 0.5$; (*b*) $K = 1.0$; (*c*) $K = 2.5$; (*d*) $K = 4.0$. (This figure courtesy of Y. Du.)

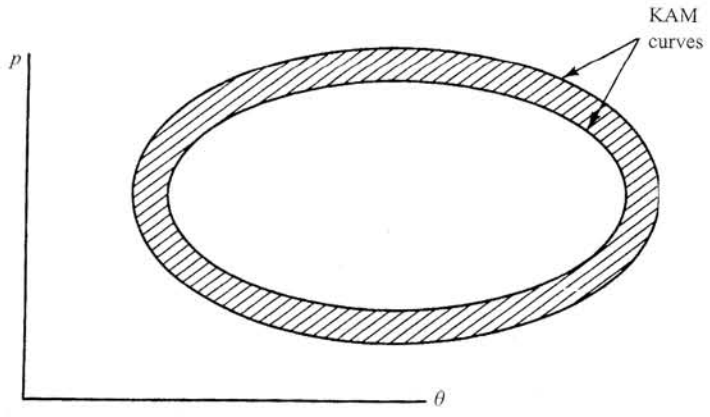


Figure 7.14 Two KAM curves bounding an annular region.

KAM curves (as, for example, in the island structures surrounding elliptic orbits), these chaotic orbits are necessarily restricted to lie between the bounding KAM curves. (As we shall discuss later, this picutre is fundamentally different for systems of higher dimensionality.)

As K is increased, more of the deformed survivors originating from the unperturbed tori are destroyed. At $K = 1$ (Figure 7.13(b)) we see that there are none left; that is, there are no tori running as continuous curves from $\theta = 0$ to $\theta = 2\pi$. In their place we see chaotic regions with interspersed island chains. As K is increased further (Figures 7.13(c) and (d)) many of the KAM surfaces associated with the island chains disappear, and the chaotic region enlarges. At $K = 4.0$, for example, we see (Figure 7.13(d)) that the only discernable islands are those associated with the period one orbits at $(\theta, p) = (\pi, 0), (\pi, 2\pi)$. Increasing K, Chirikov (1979) numerically found values of K (e.g., $K \simeq 8\frac{8}{9}$) for which there are no discernable tori, and the entire square $0 \leq (\theta, p) \leq 2\pi$ is, to within the available numerical resolution, ergodically covered densely by a single orbit. Thus, if any island chains are present, they are very small.

7.3.2 The destruction of KAM surfaces and island chains

Considering the standard map, the absence of a period $\tilde{q}$ island chain at some value $K = K'$, implies that the period $\tilde{q}$ elliptic periodic orbit has become unstable as K increases from $K = 0$ to $K = K'$. How does this occur? The answer is that as K increases, the eigenvalues of the $\tilde{q}$th iterate of the linearized map $\mathbf{DM}^{\tilde{q}}$ evaluated on the period $\tilde{q}$ orbit eventually change from complex and of magnitude 1 (i.e., $\exp(\pm i\theta)$) to real and negative with one eigenvalue with magnitude larger than 1 and one with magnitude less than 1 (i.e., λ and $1/\lambda$ with $|\lambda| > 1$). That is, the periodic orbit of period $\tilde{q}$ changes from elliptic to hyperbolic with reflection as K passes through some value $K = K_{\tilde{q}}$. When a periodic orbit becomes hyperbolic with reflection its eigenvalues in the elliptic range, $\exp(\pm i\theta)$, both approach -1 by having θ approach π as K approaches $K_{\tilde{q}}$. The migration of the eigenvalues in the complex plane as K passes through $K_{\tilde{q}}$ is illustrated in Figure 7.15. This leads to a period doubling bifurcation and is typically followed by an infinite period doubling cascade (Bountis, 1981; Greene *et al.*, 1981). In such a cascade, the period $\tilde{q}$ elliptic orbit destabilizes (becomes hyperbolic) simultaneously with the appearance of a period $2\tilde{q}$ elliptic orbit, which then period doubles to produce a period $2^2\tilde{q}$ elliptic orbit, and so on. Eventually, at some finite amount past $K_{\tilde{q}}$, all orbits of period $2^n\tilde{q}$ have been stably created and then rendered unstable (hyperbolic) as they period double. This is a Hamiltonian version of the period doubling cascade phenomena we have discussed for one-dimensional maps in Chapter 2. As in that situation, there are universal numbers that describe the scaling properties of such cascades (cf. Chapter 8), although these numbers differ in the Hamiltonian case from those given in Chapter 2. Note that, in this period doubling cascade, whenever K is in the range where there is an elliptic period $2^n\tilde{q}$ periodic orbit, there is a nested

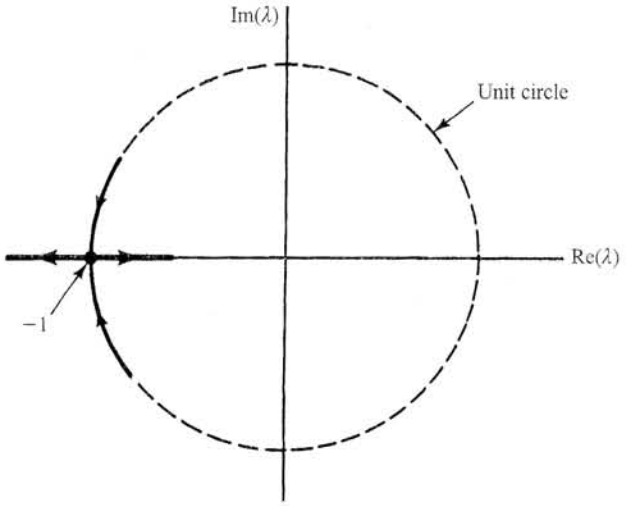


Figure 7.15 Path followed by the eigenvalues as an elliptic orbit changes to hyperbolic.

set of invariant tori surrounding that orbit (i.e., there is a period $2^n\tilde{q}$ island chain).

When $K = 0$ the standard map is integrable. As K is increased, chaotic regions occupy increasingly large areas, and the original KAM tori of the integrable system are successively destroyed. Say we identify a particular nonresonant KAM torus by its rotation number

$$R = \lim_{m\to\infty} \frac{1}{2\pi m} \sum_{n=1}^{m} p_n$$

(p_n is the amount by which θ increases on each iterate; see Eq. (7.15a)). As we increase K, the torus deforms from the straight horizontal line, $p = 2\pi R$, that it occupied for $K = 0$. Past some critical value $K > K_{\text{crit}}(R)$ the torus no longer exists. How does one numerically calculate $K_{\text{crit}}(R)$? To answer this question we note the result from number theory that the irrational number R can be represented as an infinite continued fraction,

$$R = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \cdots}}}$$

where the a_i are integers. As a shorthand we write $R = [a_1, a_2, a_3, \ldots]$. If one cuts off the continued fraction at a_n,

$$R_n = [a_1, a_2, \ldots, a_n, 0, 0, 0, \ldots],$$

then one obtains a rational approximation to R which converges to R as $n \to \infty$,

$$R = \lim_{n\to\infty} R_n.$$

If we examine the $K > 0$ island chain with rotation number R_n, we find that the elliptic Poincaré–Birkhoff periodic orbits (Figures 7.10–7.12) for the island chain approach the nonresonant torus of irrational rotation number R as n increases. This leads one to investigate the stability of these periodic orbits. As illustrated in Figure 7.15, the complex eigenvalues $\exp(i\theta)$ of the Jacobian matrix corresponding to such a periodic orbit change to real negative eigenvalues λ and $1/\lambda$ at some critical K-value (which depends on the particular periodic orbit). It is found numerically that the critical K-values of these Poincaré–Birkhoff periodic orbits of rational rotation number R_n rapidly approach the value $K_{\text{crit}}(R)$ as n increases. Since efficient numerical procedures exist for finding such orbits, this provides an efficient way of accurately determining $K_{\text{crit}}(R)$. Schmidt and Bialek (1982) have used this procedure to investigate the pattern accompanying torus destruction of arbitrary irrational tori. Greene (1979) conjectured that, since the golden mean $R_g = (\sqrt{5} - 1)/2$ is the 'most irrational' number in the sense that it is most slowly approached by cutoffs of its continued fraction expansion,

$$R_g = [1, 1, 1, 1, \ldots] = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}}}$$

the torus with $R = R_g$ will be the last surviving torus as K is increased (i.e., $K_{\text{crit}}(R)$ is largest for $R = R_g$). Using the periodic orbit technique described above, Greene finds that $K_{\text{crit}}(R_g) = 0.97\ldots$ Figure 7.16 shows the standard map for $K = 0.97$ (Greene, 1979) with the $R = R_g$ tori and some chaotic orbits plotted (there are two such tori in $0 \leq p \leq 2\pi$). An important result concerning the $R = R_g$ torus is that the phase space structure in its vicinity exhibits intricate scaling properties at and near $K = K_{\text{crit}}(R_g)$, and this phenomenon has been investigated by the renormalization group technique (Kadanoff, 1981; Escande, 1982; MacKay, 1983).

In constructing the figure plotted in Figure 7.13 we have made use of the fact that the standard map is invariant to translation in momentum by an integer multiple of 2π, $p \to p + 2\pi k$. This periodicity of the map

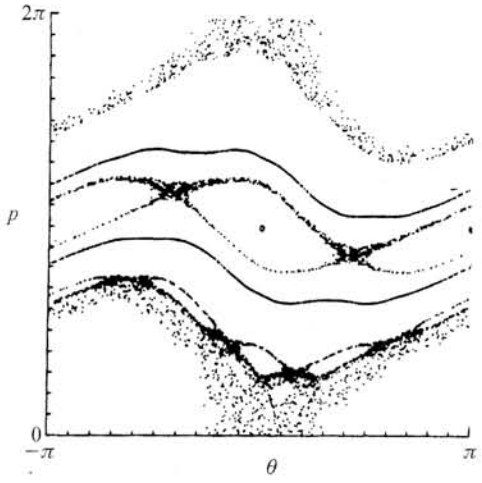


Figure 7.16 Many iterates for $K = K_c$ resulting from four different initial conditions, two of which are on the last surviving KAM surfaces encircling the (θ, p) cylinder (Greene, 1979).

allowed us to use p modulo 2π for the vertical coordinate in Figure 7.13 rather than p. Thus, if we were to ask for the structure of the solutions for all p, our answer would be given by pasting together an infinite string of pictures obtained by successively translating the basic unit (as in Figure 7.13) by 2π. For example, for the case of $K = 4.0$ (Figure 7.13(c)), we see that a single chaotic component connects regions of the line $p = 0$ with regions of the line $p = 2\pi$. By the periodicity in p, this implies that this chaotic region actually runs from $p = -\infty$ to $p = +\infty$. Thus, in terms of the rotor model (Figure 8.3), if we start an initial condition in this chaotic component, it can wander with time to arbitrarily large rotor energies, $p^2/2$ (here we have taken the rotor's moment of intertia to be 1). On the other hand, if we were to start an initial condition for $K = 4.0$ inside the period one island surrounding one of the period one fixed points, it would remain there forever; its energy would thus be bounded for all time. Note that if we plot the actual momentum, $+\infty > p > -\infty$, versus θ, then we are treating the phase space of the two-dimensional standard map as a cylinder. On the other hand, our plot where we utilized p modulo 2π reduced the phase space to the surface of a torus. While the toroidal surface representation is convenient for displaying the structure of intermixed chaotic and KAM regions, we emphasize that p and $p + 2\pi k$ (k = an integer) are not physically equivalent, since they generally represent different kinetic energies of the rotor.

The case shown in Figure 7.16 corresponds to the largest value of K for which there are KAM curves running completely around the (θ, p) cylinder in the θ-direction. The presence of a KAM curve running around the (θ, p) cylinder implies an infinite number of such curves by translation

of p by multiples of 2π. Furthermore, any orbit lying between two such curves cannot cross them (Figure 7.14) and so is restricted to lie between them forever. Thus, the energy of the rotor cannot increase without bound. When K increases past the critical value $K_c \simeq 0.97$ the last invariant tori encircling the cylinder are destroyed, and a chaotic area connecting $p = -\infty$ and $p = +\infty$ exists. This means that for $K > K_c$ the rotor energy can increase without bound if the initial condition lies in the chaotic component connecting $p = -\infty$ and $p = +\infty$.

7.3.3 Diffusion in momentum

Let us now consider the case of large K such that there are no discernible KAM surfaces present, and the entire region of a plot of p modulo 2π versus θ appears to be densely covered by a single chaotic orbit. Referring to Eq. (7.15b), we see that the change in momentum (not taken modulo 2π), $\Delta p_n \equiv p_{n+1} - p_n = K \sin\theta_{n+1}$, is typically large (i.e., of the order of K). If we assume $K \gg 2\pi$, then p will also typically be large compared to 2π. Thus, by Eq. (7.15a), we expect θ (which is taken modulo 2π) to vary very wildly in $[0, 2\pi]$. We, therefore, treat θ_n as effectively random, uniformly distributed, and uncorrelated for different times (i.e., different n). With these assumptions, the motion in p becomes a random walk with step size $\Delta p_n = K \sin\theta_{n+1}$. Thus, over momentum scales larger than K, the momentum evolves according to a diffusion process with diffusion coefficient,

$$\frac{\langle(\Delta p_n)^2\rangle}{2} = \frac{K^2}{2}\langle\sin^2\theta_{n+1}\rangle, \tag{7.41}$$

where the angle brackets denote a time average, and by virtue of the randomness assumption for the θ_n we have $\langle\sin^2\theta_{n+1}\rangle = \frac{1}{2}$. Inserting the latter in (7.41) gives the so-called *quasilinear* approximation to the diffusion coefficient,

$$D \cong D_{\mathrm{QL}} = K^2/4. \tag{7.42}$$

If we imagine that we spread a cloud of initial conditions uniformly in θ and p in the cell $-\pi \leq p \leq \pi$, then the momentum distribution function $f(p, n)$ at time n, coarse grained over intervals in p greater than 2π, is

$$f(p, n) \simeq \frac{1}{(2\pi nD)^{1/2}} \exp\left(-\frac{p^2}{2nD}\right). \tag{7.43}$$

That is, the distribution is a spreading Gaussian. This result follows from the fact that the process is diffusive. Taking the second moment of the distribution function, $\int p^2 f \, dp$, we see that the average rotor energy increases linearly with time,[6]

$$\langle p^2/2 \rangle \simeq Dn. \tag{7.44}$$

The quasilinear result (7.42) is valid for very large K. For moderately large, but not very large, values of K, neglected correlation effects can significantly alter the diffusion coefficient from the quasilinear value. These effects have been analytically calculated by Rechester and White (1980) (see also Rechester *et al.* (1981), Karney *et al.* (1981) and Carey *et al.* (1981)). Figure 7.17 shows a plot of the diffusion coefficient D normalized to D_{QL} as a function of K from the paper of Rechester and White. The solid curve is their theory, and the dots are obtained by numerically calculating the spreading of a cloud of points and obtaining D from Eq. (7.44). Note the decaying oscillations about the quasilinear value as K increases.[7]

7.3.4 Other examples

So far in this section we have dealt exclusively with the standard map. We now discuss some other examples, also reducible to two-dimensional maps, where similar phenomena are observed.

We first consider a time-independent two-degree-of-freedom system investigated by Schmidt and Chen (1994). This system, depicted in Figure 7.18, consists of two masses, a large mass M connected to a linearly behaving spring of spring constant k_s and a small mass m which elastically

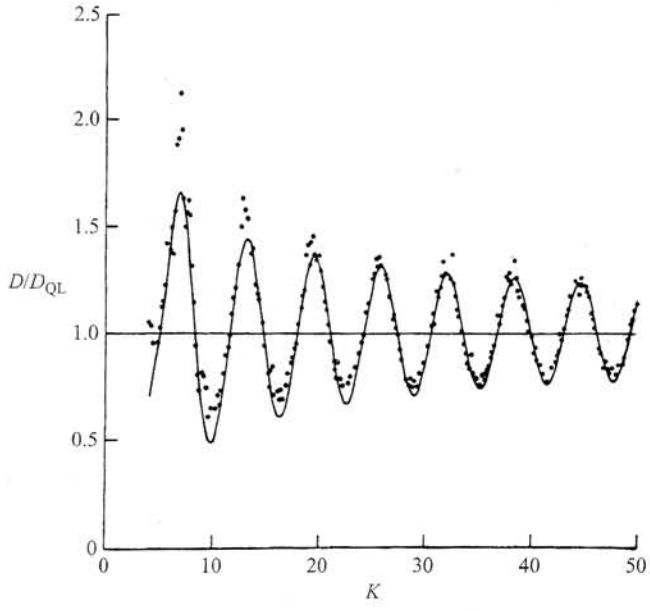


Figure 7.17 D/D_{QL} versus K for the standard map (Rechester and White, 1980).

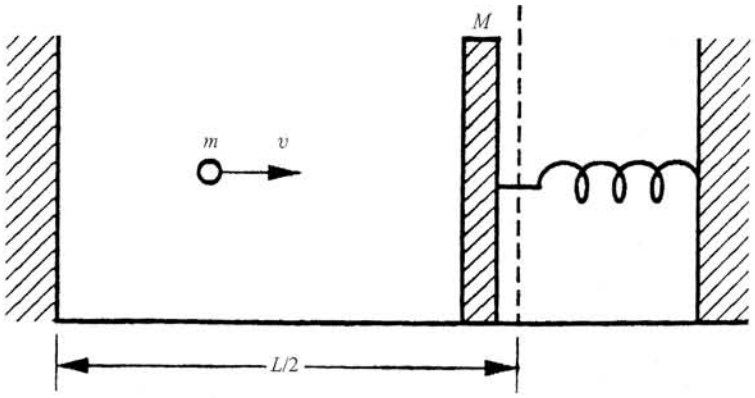


Figure 7.18 The system considered by Schmidt and Chen. $L/2$ represents the distance between the left-hand wall and the right-hand surface of mass M when the spring is in its equilibrium position. (Courtesy of Q. Chen and G. Schmidt.)

bounces between a fixed wall on the left and the oscillating large mass on the right. The motion in space is purely one-dimensional. This represents a time-independent Hamiltonian system which Schmidt and Chen call the 'autonomous Fermi system'. Since the Hamiltonian is time-independent, the total energy of the system, consisting of the sum of the kinetic energy of the two masses and the potential energy in the spring, is conserved. (This is unlike the rotor system (Figure 7.3) which has external kicking that enables the energy to increase without bound for $K > K_c$.) Schmidt and Chen numerically calculate a Poincaré surface of section and plot the state of the system at the instants of time just after the masses m and M collide. Plots corresponding to three cases are shown in Figure 7.19. In this figure v is the velocity of the small mass and ϕ is the phase of the large mass in the sinusoidal oscillation that it experiences between bounces. The maximum value of v is 1 (for the normalization used) and is attained if all the system energy is in the small mass. The three cases shown correspond to successively larger values of $\omega_0 \bar{T}$, where $\omega_0 = (k_s/M)^{1/2}$ is the natural oscillation frequency of the large mass and $\bar{T}$ is the mean time between bounces. We note that at low $\omega_0 \bar{T}$ (Figure 7.19(a)) we see many KAM surfaces as well as island chains and chaos for lower $v(v \gtrsim 0.25)$. At higher $\omega_0 \bar{T}$ (Figure 7.19(b)), the chaotic region enlarges substantially, while at the highest value plotted (Figure 7.19(c)) a single orbit appears to cover the available area of the surface of section ergodically. This latter situation corresponds to ergodic wandering of the orbit over the energy surface in the full four-dimensional phase space. Accordingly, for the case of Figure 7.19(c) Schmidt and Chen numerically confirm that there is a time average energy equipartition between the energies of the two masses and the energy of the spring, each having on average very close to one third of the total energy of the system. The equipartition of time averaged kinetic energy is a familiar result

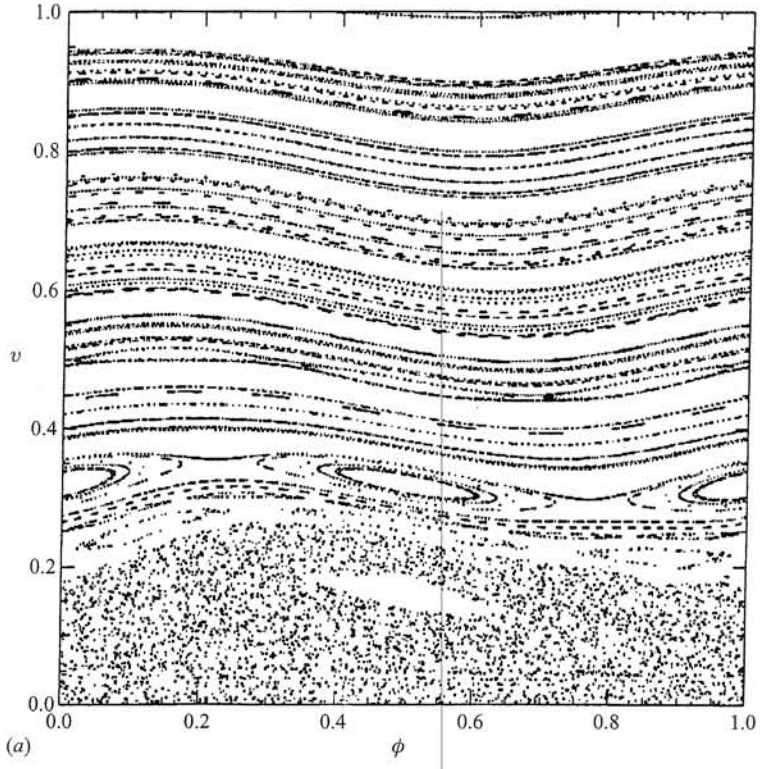


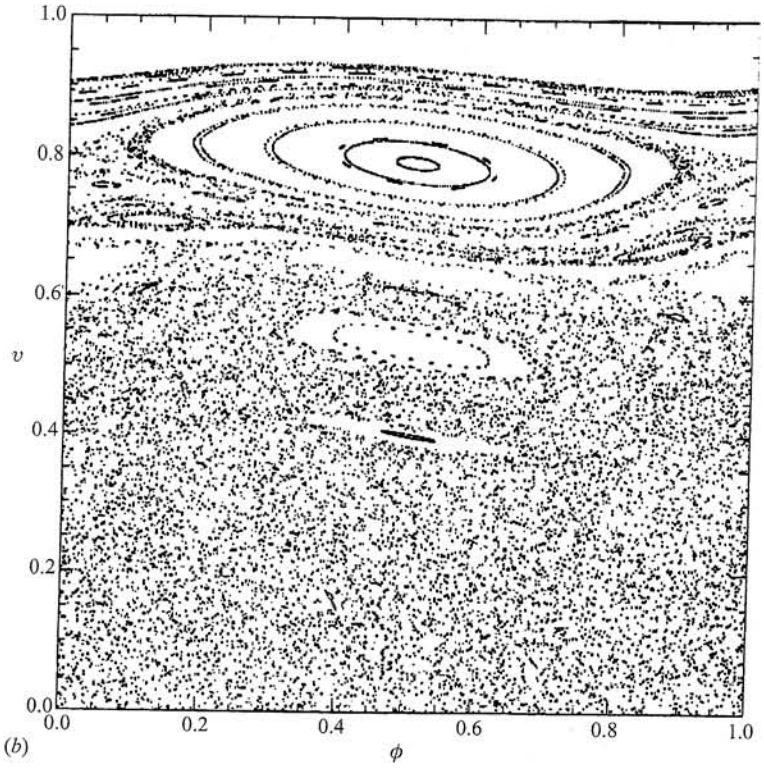
Figure 7.19 Poincaré sections of v versus ϕ. (*a*), (*b*) and (*c*) correspond to three successively larger values of $\omega_0 \overline{T}$. (Courtesy of Q. Chen and G. Schmidt.)

in the statistical mechanics of many-degree-of-freedom systems ($\langle p_1^2\rangle/2m_1 = \langle p_2^2\rangle/2m_2 = \cdots = \langle p_N^2\rangle/2m_N$). Here equipartition of kinetic energy for a system of only two degrees of freedom holds because the system is essentially ergodic on the energy surface. (Indeed it is the most important fundamental assumption of statistical mechanics that typical many-degree-of-freedom systems are ergodic on their energy surface. The justification of this assumption, however, is far from obvious, and remains an open problem.)

We could go on to cite many other examples of mechanical systems displaying the type of behavior seen in the two examples of the kicked rotor (Figure 7.3) and the autonomous Fermi system (Figure 7.19). It is, perhaps, somewhat more surprising that these same phenomena apply to situations in which one is not dealing with straightforward problems of mechanics. The point is that these problems are also described by Hamilton's equations. Three examples of this type are the following.

(1) Nonturbulent mixing in fluids.
(2) The trajectories of magnetic field lines in plasmas.

Figure 7.19 (cont.)



(3) The ray equations for the propagation of short wavelength waves in inhomogeneous media.

In the case of mixing in fluids we restrict ourselves to the situation of a two-dimensional incompressible flow: $\mathbf{v}(\mathbf{x}, t) = v_x(x, y, t)\mathbf{x}_0 + v_y(x, y, t)\mathbf{y}_0$ with $\partial v_x/\partial x + \partial v_y/\partial y = \nabla \cdot \mathbf{v} = 0$. The incompressibility condition, $\nabla \cdot \mathbf{v} = 0$, means that we can express $\mathbf{v}$ in terms of a stream function ψ,

$$\mathbf{v} = \mathbf{z}_0 \times \nabla\psi(x, y, t)$$

or

$$v_x = -\partial\psi/\partial y,\ v_y = \partial\psi/\partial x.$$

Now consider the motion of an impurity particle convected with the fluid. The location of this particle is given by $\mathrm{d}\mathbf{x}/\mathrm{d}t = \mathbf{v}(\mathbf{x}, t)$, or, using the stream function,

$$\mathrm{d}x/\mathrm{d}t = -\partial\psi/\partial y, \tag{7.45a}$$

$$\mathrm{d}y/\mathrm{d}t = \partial\psi/\partial x. \tag{7.45b}$$

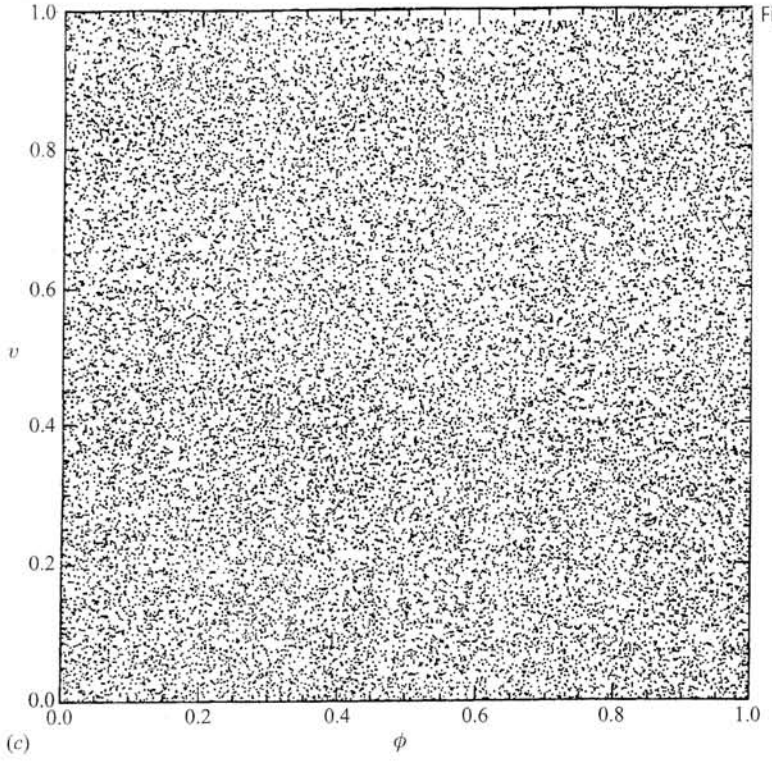


Figure 7.19 (cont.)

Comparing Eqs. (7.45) with Eqs. (7.1), we see that (7.45) are in the form of a one-degree-of-freedom ($N = 1$) time-dependent Hamiltonian system, if we identify the stream function ψ with the Hamiltonian H, x with the momentum p, and y with the 'position' q:

$$\psi(x, y, t) \leftrightarrow H(p, q, t),$$

$$x \leftrightarrow p,$$

$$y \leftrightarrow q.$$

Thus, in our fluid problem the canonically conjugate variables are x and y.

As an example, we consider the 'blinking vortex' flow of Aref (1984). In this flow there are two vortices of equal strength, one located at $(x, y) = (a, 0)$ and the other located at $(x, y) = (-a, 0)$. The vortices (which may be thought of as thin rotating stirring rods) are taken to 'blink' on and off with time with period $2T$. That is, for $2kT \leqslant t \leqslant (2k + 1)T (k = 0, 1, 2, 3, \ldots)$, the vortex at $(a, 0)$ is on, while the vortex at $(-a, 0)$ is turned off, and, for $(2k + 1)T \leqslant t \leqslant 2(k + 1)T$, the vortex at $(a, 0)$ is off while the vortex at $(-a, 0)$ is on. The flow induced by a single

vortex of strength Γ can be expressed in (ρ, θ) polar coordinates centered at the vortex as

$$v_\theta = \Gamma/2\pi\rho \text{ and } v_\rho = 0.$$

Thus, the blinking vortex has the effect of alternatively rotating points in concentric circles first about one vortex center and then about the other vortex center, each time by an angle $\Delta\theta = \Gamma T/2\pi\rho^2$. Sampling the position of a particle at times $t = 2kT$ defines a two-dimensional area preserving map which depends on the strength parameter $\mu = \Gamma T/2\pi a^2$. Figure 7.20 from Doherty and Ottino (1988) shows results from iterating several different initial conditions for successively larger values of the strength parameter μ. For very small μ (Figure 7.20(a), $\mu = 0.1$) the result is very close to the completely integrable case where both vortices act simultaneously and steadily in time. As μ is increased we see that the area occupied by chaotic motion increases.

The practical effect of this type of result for fluid mixing can be seen by considering a small dollop of dye in such a flow. For example, say the dye is initially placed in the location indicated by the shaded circle in Figure 7.20(a). In the near integrable case Figure 7.20(a), as time goes on, this dye would always necessarily be located between the two KAM curves that initially bound it. Due to the different rotation rates on the different KAM surfaces, the dye will mix throughout the annular region bounded by these KAM curves, but (in the absence of molecular diffusion) it can never mix with the fluid outside this annular region. In contrast, for the case $\mu = 0.4$ (Figure 7.20(e)), we see that there is a large single connected chaotic region, and an initial dollop of dye in the same location as before would thus mix uniformly throughout this much larger region. Thus, we see that, for the purposes of achieving the most uniform mixing in fluids, chaos is a desirable attribute of the flow that one should strive to maximize.

Several representative references on chaotic mixing in fluids are Aref and Balachandar (1986), Chaiken *et al.* (1986), Dombre *et al.* (1986), Feingold *et al.* (1988), Ott and Antonsen (1989), Rom-Kedar *et al.* (1990), and the comprehensive book on the subject by Ottino (1989).

We now discuss the second of the three applications mentioned above, namely, the trajectory of magnetic field lines in plasmas. Let $\mathbf{B}(\mathbf{x})$ denote the magnetic field vector. The field line trajectory equation gives a parametric function $\mathbf{x}(s)$ for the curve on which a magnetic field line lies, where s is a parameter which we can think of as a (distorted) measure of distance along the field line. The equation for $\mathbf{x}(s)$ is

$$\mathrm{d}\mathbf{x}(s)/\mathrm{d}s = \mathbf{B}(\mathbf{x}). \tag{7.46}$$

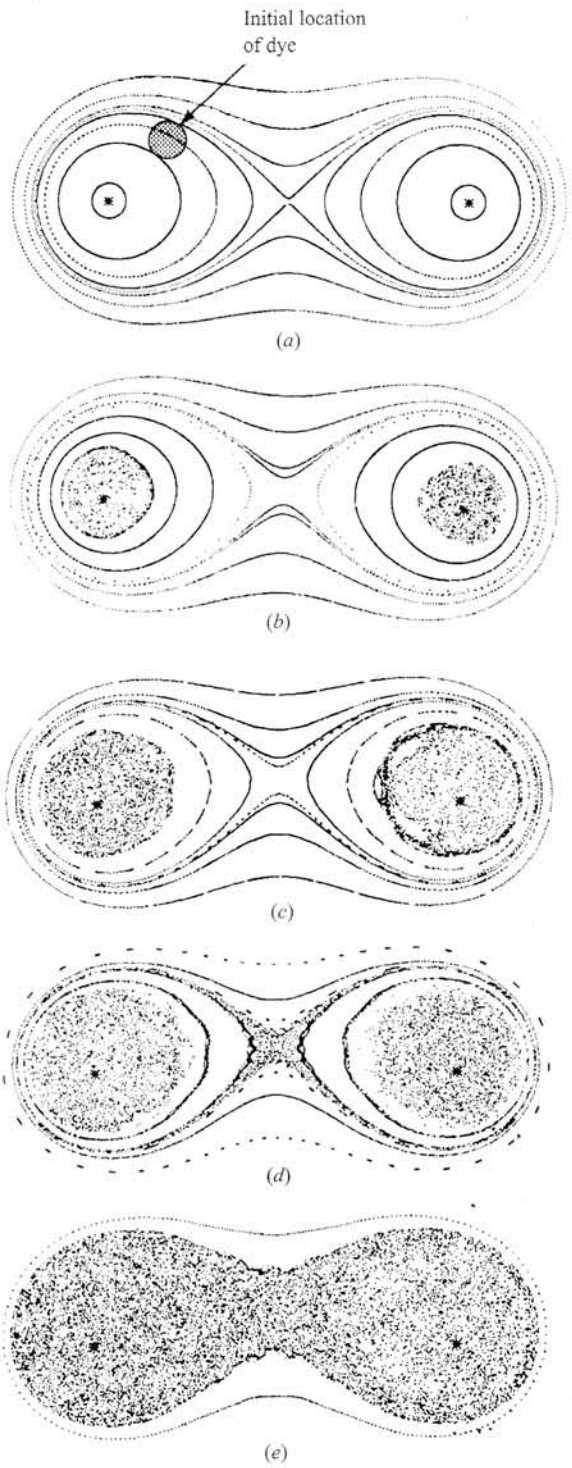


Figure 7.20 Blinking vortex orbits for (*a*) $\mu = 0.01$, (*b*) $\mu = 0.15$, (*c*) $\mu = 0.25$, (*d*) $\mu = 0.3$ and (*e*) $\mu = 0.4$ (Doherty and Ottino, 1988).

(Alternatively, we can multiply the right-hand side of (7.46) by any positive scalar function of $\mathbf{x}$.) Since $\nabla \cdot \mathbf{B} = 0$, Eq. (7.46) represents a conservative flow, if we make an analogy between s and time. Thus, the magnetic field lines in physical space are mathematically analogous to the trajectory of a dynamical system in its phase space. In Problem 3 you are asked to establish for a simple example that (7.46) can be put in Hamiltonian form.

The Hamiltonian nature of 'magnetic field line flow' means that under many circumstances we can expect that some magnetic field line trajectories fill up toroidal surfaces, while other field lines wander chaotically over a volume which may be bounded by tori. In other words, the situation can be precisely as depicted in Figure 7.12.

These considerations are of great importance in plasma physics and controlled nuclear fusion research. In the latter, the fundamental problem is to confine a hot plasma (gas of electrons and ions) for a long enough time that sufficient energy-releasing nuclear fusion reactions take place. If the magnetic field is strong, then, to a first approximation, the motion of the charged particles constituting the plasma is constrained to follow the magnetic field lines. (This approximation is better for the lighter mass electrons than for the ions.) In this view, the problem of confining the plasma becomes that of creating a magnetic field line configuration such that the magnetic field lines are confined. That is, the magnetic field lines do not connect the plasma interior to the walls of the device. The most simple example of such a configuration is provided by the tokamak device, originally invented in the Soviet Union. (This device is currently the one on which most of the attention of the nuclear fusion community is focused.) Figure 7.21(a) illustrates the basic idea of the tokamak. An external current system (in the figure the wire with current I_0) creates a magnetic field B_T running the long way (called the 'toroidal direction') around a toroid of plasma. At the same time, another current is induced to flow in the plasma in the direction running the long way around the torus. (This toroidal plasma current is typically created by transformer action wherein the plasma loop serves as the secondary coil of a transformer.) The toroidal plasma current then creates a magnetic field component B_p which circles the short way around the torus (the 'poloidal direction'). Assuming that the configuration is perfectly symmetric with respect to rotations around the axis of the system, the superposition of the toroidal and poloidal magnetic fields leads to field lines that typically circle on a toroidal surface, simultaneously in both the toroidal and poloidal directions, filling the surface ergodically. Thus, the field lines are restricted to lie on a nested set of tori and never intersect bounding walls of the device. This is precisely analogous to the case of an integrable Hamiltonian system.

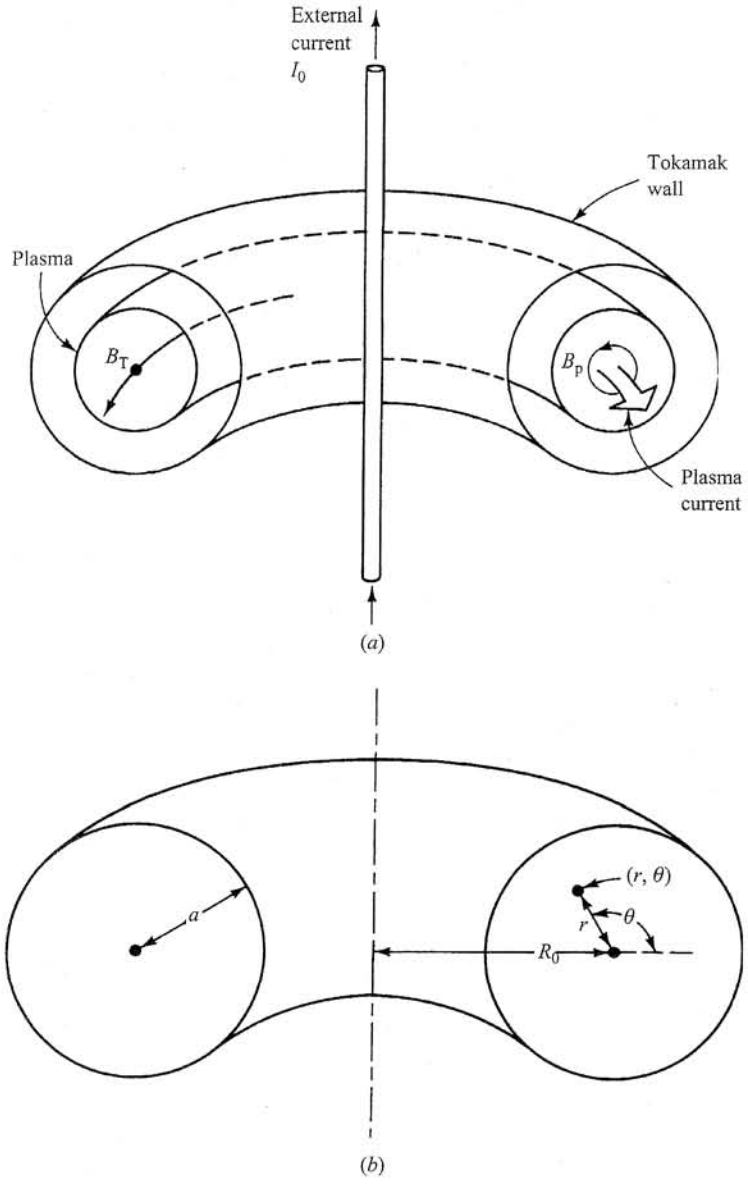


Figure 7.21 (*a*) Schematic illustration of a tokamak. (*b*) Toroidal coordinates.

This is the situation if there is perfect toroidal symmetry. Unfortunately, symmetry can be destroyed by errors in the external field coils, by necessary asymmetries in the walls, and, most importantly, by toroidal dependences of the current flowing the plasma. (The latter can arise due to collective motions of the plasma as a result of a variety of instabilities that have been very extensively investigated.) Such symmetry-breaking

magnetic field perturbations play a role analogous to nonintegrable perturbations of an integrable Hamiltonian system. Thus, they can destroy some of the nested set of toroidal magnetic surfaces that exists in the symmetric case. If the perturbation is too strong, chaotic field lines can wander from the interior of the plasma to the wall. This leads to rapid heat and particle loss of the plasma. (Refer to Figure 7.13 and think of K as the strength of the asymmetric field perturbation.)

Some representative papers which discuss chaotic magnetic field line trajectories in plasmas and their physical effects are Rosenbluth *et al.* (1966), Sinclair *et al.* (1970), Finn (1975), Rechester and Rosenbluth (1978), Cary and Littlejohn (1983), Hanson and Cary (1984) and Lau and Finn (1991).

As our final example, we consider the ray equations describing the propagation of short wavelength waves in a time-independent spatially inhomogeneous medium. In the absence of inhomogeneity, we assume that the partial differential equations governing the evolution of small amplitude perturbations of the dependent quantities admit plane wave solutions in which the perturbations vary as $\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{x} - \mathrm{i}\omega t)$, where ω and $\mathbf{k}$ are the frequency and wavenumber of the wave. The quantities ω and $\mathbf{k}$ are constrained by the governing equations (e.g., Maxwell's equations if we are dealing with electomagnetic waves) to satisfy a dispersion relation

$$D(\mathbf{k}, \omega) = 0.$$

Now assume that the medium is inhomogeneous with variations occurring on a scale size L which is much longer than the typical wavelength of the wave, $|\mathbf{k}|L \gg 1$. For propagation distances small compared to L, waves behave approximately as if the medium were homogeneous. For propagation distances of the order of L or longer, the spatial part of the homogeneous medium expression for the phase, namely $\mathbf{k}\cdot\mathbf{x}$, is distorted. We, therefore, assume that the perturbations have a rapid (compared to L) spatial variation of the form

$$\exp[\mathrm{i}\tilde{S}(\mathbf{x}) - \mathrm{i}\omega t], \tag{7.47}$$

where the function $\tilde{S}(\mathbf{x})$ is called the *eikonal* and replaces the homogeneous medium phase term $\mathbf{k}\cdot\mathbf{x}$. The *local* wavenumber $\mathbf{k}$ is given by

$$\mathbf{k} = \nabla\tilde{S}(\mathbf{x}). \tag{7.48}$$

We wish to find an equation for the propagation of a wave along some path (called the 'ray path'). Along this path we seek parametric equations for $\mathbf{x}$ and $\mathbf{k}$. That is, we seek $(\mathbf{x}(s), \mathbf{k}(s))$, where s measures the distance along the ray. In terms of these ray path functions, we can determine the function $\tilde{S}(\mathbf{x})$ using Eq. (7.48),

$$\tilde{S}(\mathbf{x}) = \tilde{S}(\mathbf{x}_0) + \int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{k}(s) \cdot \mathrm{d}\mathbf{x}(s).$$

(In addition, by a systematic expansion in $|\mathbf{k}|L$, one can also determine the variation of the wave amplitudes along the rays as well as higher order corrections. See, for example, Courant and Hilbert (1966) and references therein.) Using the local wavenumber $\mathbf{k}$, we can write down a local dispersion relation,

$$D(\mathbf{k}, \omega, \mathbf{x}) = 0, \tag{7.49}$$

which is identical with the homogeneous medium dispersion relation if we use the local parameters of the inhomogeneous medium at the point $\mathbf{x}$. Since the medium is assumed time-independent, the frequency ω is constant. Differentiating (7.49) with respect to $\mathbf{x}$, we obtain

$$\frac{\partial D}{\partial \mathbf{x}} + \nabla \mathbf{k} \cdot \frac{\partial D}{\partial \mathbf{k}} = 0. \tag{7.50}$$

The x-component of this equation is

$$\frac{\partial D}{\partial x} + \frac{\partial D}{\partial k_x}\frac{\partial k_x}{\partial x} + \frac{\partial D}{\partial k_y}\frac{\partial k_y}{\partial x} + \frac{\partial D}{\partial k_z}\frac{\partial k_z}{\partial x} = 0.$$

However, from (7.48) we have $\nabla \times \mathbf{k} = 0$, so that $\partial k_y/\partial x = \partial k_x/\partial y$ and $\partial k_z/\partial x = \partial k_x/\partial z$. Thus, the above equation becomes

$$\frac{\partial D}{\partial x} + \frac{\partial D}{\partial \mathbf{k}} \cdot \frac{\partial k_x}{\partial \mathbf{x}} = 0.$$

Since similar equations apply for the y- and z-components, (7.50) becomes

$$(\partial D/\partial \mathbf{k}) \cdot \nabla \mathbf{k} = -\partial D/\partial \mathbf{x}. \tag{7.51}$$

If we regard $\partial D/\partial \mathbf{k}$ as being like a velocity, the $(\partial D/\partial \mathbf{k}) \cdot \nabla$ is like the time derivative following a point with the 'velocity' $\partial D/\partial \mathbf{k}$. Thus,

$$\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}s} = -\frac{\partial D}{\partial \mathbf{x}}, \tag{7.52a}$$

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s} = \frac{\partial D}{\partial \mathbf{k}}, \tag{7.52b}$$

which are the ray equations, the second of which comes from our interpretation of $\partial D/\partial \mathbf{k}$ as being like a velocity with the parametric variable s playing the role of the fictitious time variable. More formally, Eqs. (7.52) represent the solution of the first-order partial differential equation, Eq. (7.51), by the method of characteristics (Courant and Hilbert, 1966). In the special case where the dispersion relation is written in the form, $D(\mathbf{k}, \omega, \mathbf{x}) = \hat{\omega}(\mathbf{k}, \mathbf{x}) - \omega$, we have that $\partial D/\partial \mathbf{k} = \partial \hat{\omega}/\partial \mathbf{k}$

which is just the group velocity of a wavepacket. Thus, letting τ denote time, we have $d\mathbf{x}/d\tau = \partial\hat{\omega}/\partial\mathbf{k}$, and hence $s = \tau$ in this case. This yields

$$d\mathbf{k}/d\tau = -\partial\hat{\omega}/\partial\mathbf{x}, \tag{7.53a}$$

$$d\mathbf{x}/d\tau = \partial\hat{\omega}/\partial\mathbf{k}, \tag{7.53b}$$

which can be interpreted as giving the temporal evolution of the position $\mathbf{x}$ and wavenumber $\mathbf{k}$ of a wavepacket. Both (7.52) and (7.53) are Hamiltonian with D and $\hat{\omega}$ respectively playing the role of the Hamiltonian, and $(\mathbf{k}, \mathbf{x})$ being the canonically conjugate momentum $(\mathbf{k})$ and position $(\mathbf{x})$ variables.

We now discuss a particular example where chaotic solutions of the ray equations play a key role (Bonoli and Ott, 1982; Wersinger *et al.*, 1978). One of the central problems in creating a controlled thermonuclear reactor lies in raising the temperature of the confined plasma sufficiently to permit fusion reactions to take place. One way of doing this is by launching waves from outside the plasma that then propagate to the plasma interior, where they dissipate their energy to heat (similar, in principle, to a kitchen microwave oven). Clearly, conditions must be such that the wave is able to reach the plasma interior. In this case, in the terminology of the field, the wave is said to be 'accessible.' Of the various types of waves that can be used for plasma heating, the so-called 'lower hybrid' wave is one of the most attractive from a technological point of view. The accessibility problem for this wave was originally considered by Stix (1965) for the case in which there are two symmetry directions. For example, in a straight circular cylinder, k_z and $m = k_\theta r$ are constants of the ray equations due to the translational symmetry along the axis of the cylinder (the z-axis) and to the rotational symmetry around the cylinder (in θ). The accessibility situation for this case is illustrated in Figure 7.22 for a cylinder with an applied magnetic field $\mathbf{B} = B_0\mathbf{z}_0 + B_\theta\boldsymbol{\theta}_0$, and a wave launched from vacuum with $m = 0$. Let $n_\parallel = k_z c/\omega$ (where c is the speed of light) and $n_\perp = k_r c/\omega$. We assume that the plasma density N_0 increases with decreasing radius r from $N_0 = 0$ at the plasma edge ($r = a$) to its maximum value in the center of the cylinder ($r = 0$). Figure 7.22 shows plots of $n_\perp^2$ (obtained from the dispersion relation) as a function of N_0. For $n_\parallel < n_a$, $n_\parallel = n_a$, and $n_\parallel > n_a$, Figures 7.22(a), (b) and (c) apply, respectively, where n_a is a certain critical value (Stix, 1965). Between $N_0 = 0$ and $N_0 = N_S$, there is a narrow cutoff region through which a slow wave (i.e., lower hybrid wave), launched from the vacuum region, typically has little trouble in tunneling. Figure 7.22(a) shows that for $n_\parallel < n_a$ an additional, effectively much wider, cutoff region between $N_0 = N_{T1}$ and $N_0 = N_{T2}$ exists. This cutoff region presents a barrier for propagation to the plasma center and prevents

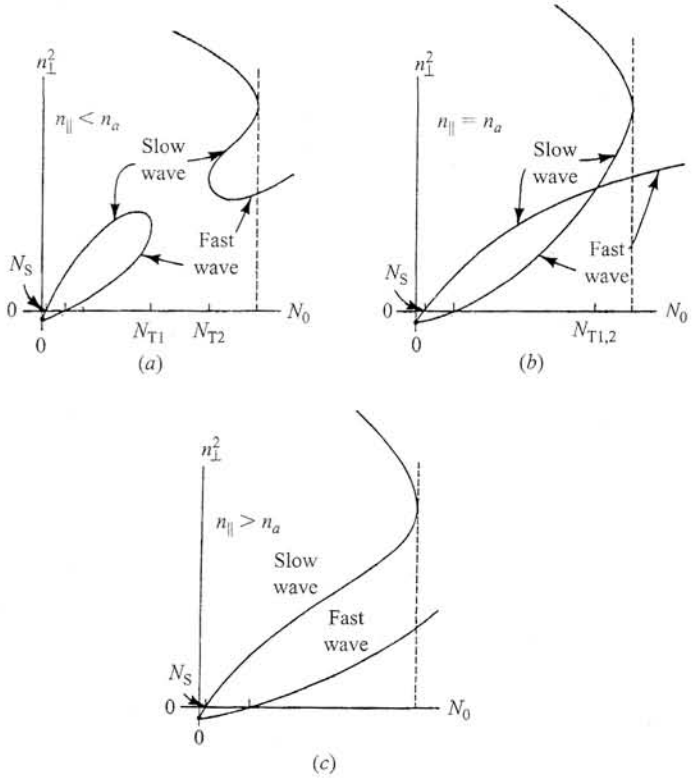


Figure 7.22 Plots of $n_\perp^2$ versus density N_0 for (*a*) $n_\parallel < n_a$, (*b*) $n_\parallel = n_a$ and (*c*) $n_\parallel > n_a$.

accessibility. Figure 7.22(*c*) shows that for $n_\parallel > n_a$ this barrier is absent, and the lower hybrid (slow) wave becomes accessible.

Now we consider heating a circular cross section toroidally symmetric plasma (a tokamak). We use toroidal coordinates wherein r, θ are circular polar coordinates centered in their circular cross section of the tokamak plasma such that the distance of a point from the major axis of the torus is $R = R_0 + r\cos\theta$, where R_0 is the distance from the major axis of the torus to the center of the plasma cross section (Figure 7.21(*b*)). We refer to θ as the poloidal angle, and we denote by ϕ the toroidal angle (i.e., the angle running the long way around the plasma torus). Let $\varepsilon = a/R_0$, where $r = a$ denotes the plasma boundary. As $\varepsilon \to 0$ with a fixed, the straight cylinder limit is approached. However, for finite ε the plasma equilibrium depends on θ. Thus, it is no longer expected that $m = rk_\theta$ is a constant of the motion, although the toroidal symmetry still guarantees that a constant of the motion analogous to k_z in the cylinder still exists; namely, $n = Rk_\phi$ is a constant. The questions that now arise are what happens to the constant m, and how is the accessibility condition for lower

hybrid waves affected? For finite ε there may still be some other constant $\tilde{m} = \tilde{m}(r, \theta, k_r, m)$, which takes the place of m. For small ε, regions where $\tilde{m}$ exist (KAM tori) occupy most of the phase space. As ε increases the regions occupied by chaotic trajectories increase, until almost all regions where KAM tori exist are gone. A ray in the region with no tori may eventually approach the plasma interior and be absorbed even if $n_\parallel$ at launch does not satisfy the straight cylinder accessibility condition. Thus we need to know at what value of ε most of the tori are gone.

Figure 7.23 shows numerical results (Bonoli and Ott, 1982) testing for the existence of tori by the surface of section method with $\theta = 0 \pmod{2\pi}$ as the surface of section. Figure 7.23(a) shows that for $\varepsilon = 0.10$, most tori are not destroyed, and initially inaccessible rays (i.e., $n_\parallel < n_a$ at launch) do not reach the plasma interior. Figures 7.23(b) and (c) show a case for $\varepsilon = 0.15$, illustrating the coexistence of chaotic and integrable orbits including (Figure 7.23(c)) higher-order island structures. For $\varepsilon = 0.25$ all appreciable KAM surfaces are numerically found to be completely destroyed, and even waves launched with $n_\parallel$ substantially below n_a are absorbed in the plasma interior after a few piercings of the surface of section.

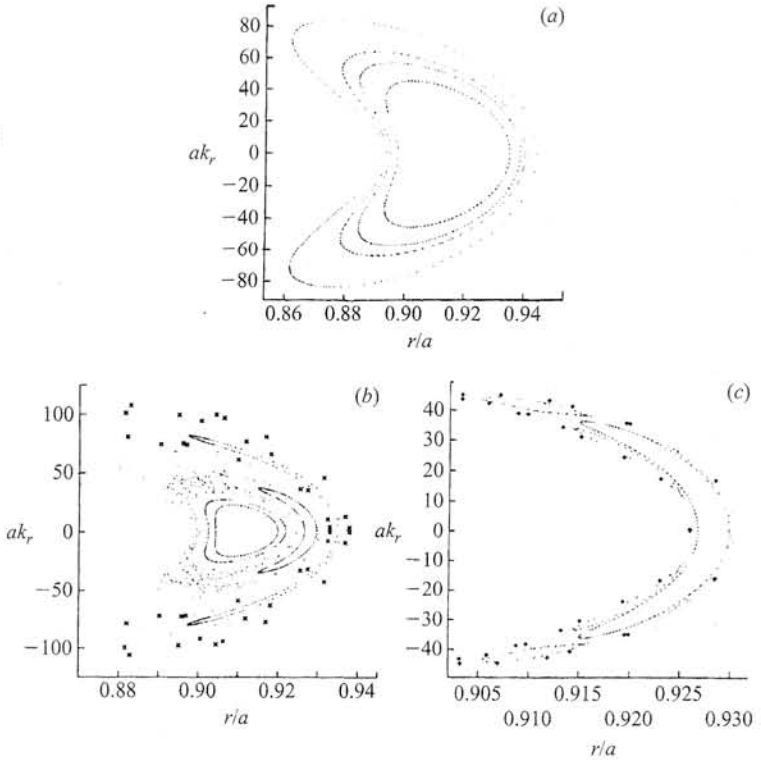


Figure 7.23 Surface of section ($\theta = 0$) plots for several different initial conditions. $n_a = 2.0$. (a) $1.3 \leq n_\parallel \leq 1.4$, $a/R_0 = 0.10$; (b) $1.25 \leq n_\parallel \leq 1.4$; $a/R_0 = 0.15$; (c) same as (b) but with a different initial condition (Bonoli and Ott. 1981).

7.4 Higher-dimensional systems

There is a very basic topological distinction to be made between the case of time-independent Hamiltonians with $N = 2$ degrees of freedom, on the one hand, and $N \geqslant 3$ degrees of freedom, on the other. (For a time periodic Hamiltonian the same distinction applies for the cases $N = 1$ and $N \geqslant 2$.) In particular, since the energy is a constant of the motion for a time-independent system, the motion is restricted to the $(2N - 1)$-dimensional energy surface $H(\mathbf{p}, \mathbf{q}) = E$. Thus, we can regard the dynamics as taking place in an effectively $(2N - 1)$-dimensional space. In general, in order for a closed surface to divide a $(2N - 1)$-dimensional space into two distinct parts, one inside the closed surface and another outside, the closed surface must have a dimension one less than the dimension of the space; i.e., its dimension is $2N - 2$. Thus, KAM surfaces which are N-dimensional tori only satisfy this condition for $N = 2$. In particular, for $N = 2$, the energy surface has a dimension 3, and a two-dimensional toroidal surface in a three dimension space has an inside and an outside. As an example of a toroidal 'surface' which does not divide the space in which it lies, consider a circle (which can be regarded as a 'one-dimensional torus') in a three-dimensional Cartesian space. For $N > 2$ the situation for KAM tori in the energy surface is similar (e.g., for $N = 3$ we have $2N - 1 = 5$ and $2N - 2 = 4 > 3 = N$).

Now consider the situation where an integrable system is perturbed. In this case tori begin to break up and are replaced by chaotic orbits. For the case $N = 2$ these chaotic regions are necessarily sandwiched between surviving KAM tori (Figure 7.15). In particular, if such an orbit is outside (inside) a particular torus it remains outside (inside) that torus forever. Because of this sandwiching effect, the chaotic orbit of a slightly perturbed integrable two-degree-of-freedom system must lie close to the orbit on a torus of the unperturbed integrable system for all time. Hence, two-degree-of-freedom integrable systems are relatively stable to perturbations. The situation for $N \geqslant 3$ is different because chaotic orbits are not enclosed by tori, and hence their motions are not restricted as in the case $N = 2$. In fact, it is natural to assume that all the chaos created by destroyed tori can form a single connected ergodic chaotic region which is dense in the phase space. Under this assumption a chaotic orbit can, in principle, come arbitrarily close to *any* point in phase space, if we wait long enough. This phenomenon was first demonstrated for a particular example by Arnold (1964) and is known as 'Arnold diffusion.' For further discussion of Arnold diffusion and other aspects of chaos in Hamiltonian systems with more than two degrees of freedom we refer the reader to the exposition of these topics in the book by Lichtenberg and Lieberman (1983).

7.5 Strongly chaotic systems

We have seen in Sections 7.2 and 7.3 that when elliptic periodic orbits are present there is typically an exceedingly intricate mixture of chaotic regions and KAM tori: surrounding each elliptic orbit are KAM tori, between which are chaotic regions and other elliptic orbits, which are themselves similarly surrounded, and so on *ad infinitum*. It would seem that the situation would be much simpler if there were no elliptic periodic orbits (i.e., all were hyperbolic). In such a case one would expect that the whole phase space would be chaotic and no KAM tori would be present at all.

As a model of such a situation, one can consider two-dimensional area preserving maps which are hyperbolic. One example is the cat map, Eq. (4.29), discussed in Chapter 4. In that case we saw that the map took the picture of the cat and stretched it out (chaos) and reassembled it in the square (Figure 4.13). (Recall that the square is an unwrapping of a two-dimensional toroidal surface.) More iterations mix the striations of the unfortunate cat more and more finely within the square. Given any small fixed region $\mathscr{R}$ within the square, as we iterate more and more times, the fraction of the area of the region $\mathscr{R}$ occupied by black striations that were originally part of the face of the cat approaches the fraction of area of the entire square that was originally occupied by the cat's face (the black region in Figure 4.13). We say that the cat map is *mixing* on the unit square with essentially the same meaning that we use when we describe the mixing of cream as a cup of coffee is stirred. More formally, an area preserving map $\mathbf{M}$ of a compact region S is mixing on S if given any two subsets σ and σ' of S where σ and σ' have positive Lebesgue measure ($\mu_L(\sigma) > 0, \mu_L(\sigma') > 0$), then

$$\frac{\mu_L(\sigma)}{\mu_L(S)} = \lim_{m\to\infty} \frac{\mu_L[\sigma' \cap \mathbf{M}^m(\sigma)]}{\mu_L(\sigma')}. \tag{7.54}$$

As another example of an area preserving mixing two-dimensional hyperbolic map, we mention the generalized baker's map (Figure 3.4) in the area preserving case, $\lambda_a = \alpha$, $\lambda_b = \beta$. (It is easy to check that the Jacobian determinant, Eq. (4.28), is one in this case.)

Another group of strongly chaotic systems can be constructed from certain classes of 'billiard' problems. A billiard is a two-dimensional planar domain in which a point particle moves with constant velocity along straight line orbits between specular bounces ((angle of incidence) = (angle of reflection)) from the boundary of the domain, Figure 7.24(a). Figures 7.24(b)–(g) show several shapes of billiards. The circle (b) and the rectangle (c) are completely integrable. The two constants of the motion for the circle are the particle energy and the angular momentum about the center of the circle. The two angular frequencies associated

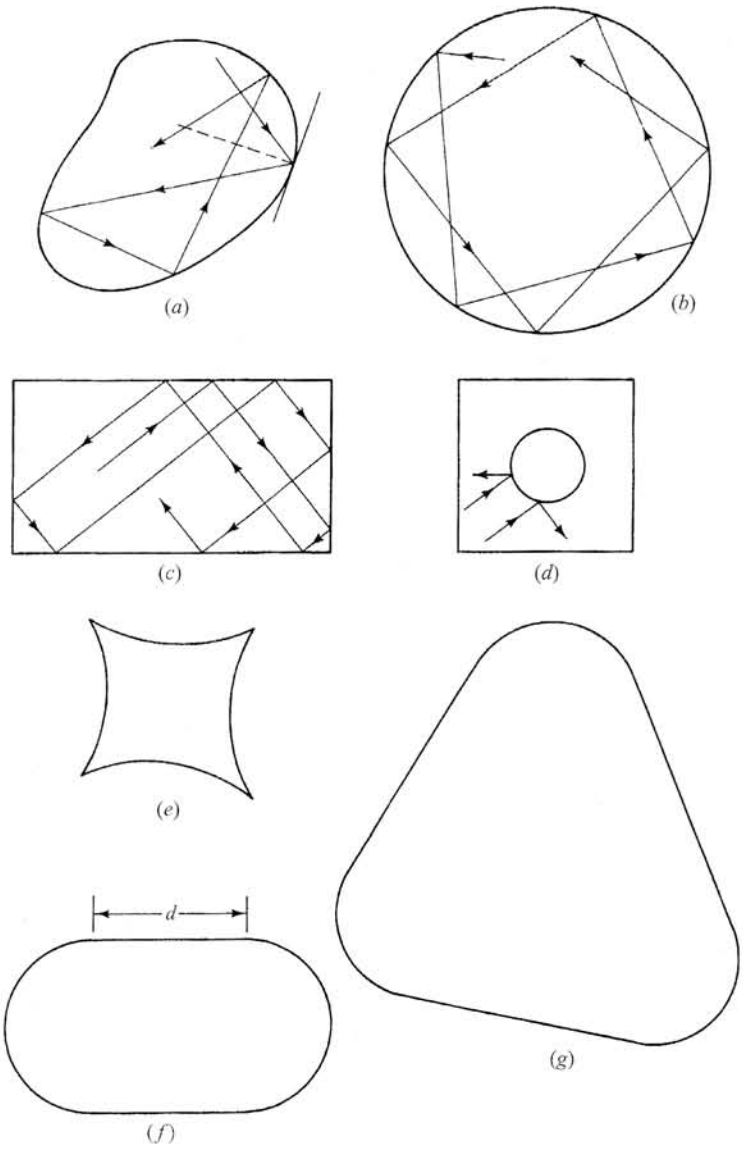


Figure 7.24 Billiard systems.

with the action-angle variables are the inverses of the time between bounces and the time to make a complete rotation around the center of the circle. For the rectangle the two constants are the vertical and horizontal kinetic energies, and the two frequencies are the inverses of twice the time between successive bounces off the vertical walls and twice the time between successive bounces off the horizontal walls. The billiard shapes shown in Figures 7.24(d)–(g) are strongly chaotic (Bunimovich, 1979;

Sinai, 1970) in the same sense as the cat map and the area preserving generalized baker's map: almost every initial condition yields a chaotic orbit which eventually comes arbitrarily close to every point in the phase space. (In Figures 7.24(d)–(g) the curved line segments are arcs of circles.) In particular, for these chaotic billiard problems the orbit generates a density which is uniform in the angle of the velocity vector and is also uniform in the accessible area of the billiard. Note that, unlike the cat map and the area preserving generalized baker's map, the billiard systems shown in Figures 7.24(d), (f) and (g) are not hyperbolic. This is because, in violation of the definition of hyperbolicity in Section 4.3, they possess periodic orbits with zero Lyapunov exponent. (For (d) and (f) there are simple zero exponent periodic orbits bouncing vertically between the straight parallel walls (and, for (d), not intersecting the interior circle). For (g) there is a zero exponent periodic orbit tracing out a triangle as it bounces between points on the straight line component segments of the boundary.)

For the cases (d) and (e) in Figure 7.24 the origin of the chaos is intuitively fairly clear. The curved boundary segments have curvature vectors pointing out of the billiard area. As shown in Figure 7.24(d) this has a 'dispersion' effect on two parallel orbits, and one expects this to lead to rapid separation of nearby orbits after a few bounces. Indeed, Sinai (1970) showed that such systems were chaotic. Further work by Bunimovich (1979) showed that a certain class of nondispersing billiards, of which (f) and (g) of Figure 7.24 are members, are also chaotic. The so-called stadium billiard (f) is of particular interest in that it has been the subject of a number of numerical studies, particularly in the context of quantum chaos (Chapter 11). Note that the stadium is chaotic for any value of $d > 0$, where d is the length of the two parallel line segments joining the two semicircular segments at the left and right end of the stadium. Thus, we have the somewhat surprising result that the stadium is chaotic for any $d > 0$, but, as soon as we make $d = 0$, we have the circular billiard, Figure 7.24(b), which is integrable.

The author is unaware of any example of a smooth two-dimensional potential $V(x, y)$ in Cartesian spatial variables $\mathbf{q} = (x, y)$ such that the Hamiltonian $H = (p_x^2/2m) + (p_y^2/2m) + V(x, y)$ yields chaotic orbits that are ergodic on the energy surface. For a time it was thought that the potential $V(x, y) = x^2y^2$ would yield chaotic orbits for almost all initial conditions, but very careful numerical analysis (Dahlqvist and Russberg, 1990) has found KAM surfaces surrounding a small region about an elliptic periodic orbit of high period (period 11 in the surface of section).

An example of a continuous time two-dimensional dynamical system which is completely chaotic, in the sense we have been using, is the case of a point mass moving along geodesics on a closed two-dimensional

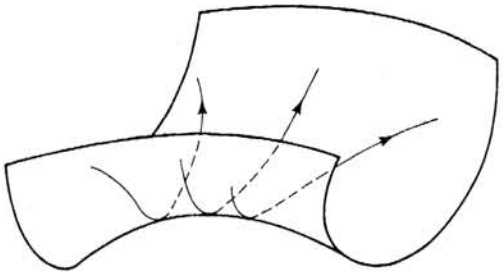


Figure 7.25 Geodesics on a surface of negative curvature.

surface of negative Gaussian curvature (the two principal curvature vectors at each point on the surface point to opposite sides of the surface, Figure 7.25) (Hadamard, 1898). Two geodesics on such a surface that are initially close and parallel separate exponentially as they are followed forward in time. Note, however, that a closed surface of negative curvature cannot be embedded in a three-dimensional Cartesian space (four dimensions are required), so this example is somewhat nonphysical, although it has proven to be very fruitful for mathematical study.

7.6 The succession of increasingly random systems

In the previous section we have discussed 'strongly chaotic' systems by which we meant chaotic systems that were mixing throughout the phase space. One often encounters in the literature various terms used to describe the degree of randomness of a Hamiltonian system. In particular, one can make the following list in order of 'increasing randomness:'

ergodic systems,
mixing systems,
K-systems,
C-systems,
Bernoulli systems.

We now discuss and contrast these terms, giving some examples of each.

Ergodicity is defined in Section 2.3.3 for maps. For an ergodic invariant measure of a dynamical system, phase space averages are the same as time averages. That is, for the case of a continuous time system, ergodicity implies

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(\bar{\mathbf{x}}(t))\mathrm{d}t = \langle f(\bar{\mathbf{x}})\rangle, \tag{7.55}$$

where $f(\bar{\mathbf{x}})$ is any smooth function of the phase space variable $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}(t)$

represents a trajectory in phase space, $\langle f(\tilde{\mathbf{x}}) \rangle$ represents the average of $f(\tilde{\mathbf{x}})$ over the phase space weighted by the invariant measure under consideration, and (7.55) holds for almost every initial condition with respect to the invariant measure. As an example, consider the standard map, Eqs. (7.15). In the case $K = 0$, we have (Eq. (7.40)) $\theta_{n+1} = (\theta_n + p_n)$ modulo 2π, $p_{n+1} = p_n$. Thus the lines $p = \text{const.}$ are invariant. Any region $p_a > p > p_b$ is also invariant. Orbits are not ergodic in $p_a > p > p_b$. Orbits are, however, ergodic on the lines $p = \text{const.}$, provided that $p/2\pi$ is irrational. Ergodicity is the weakest form of randomness and does not necessarily imply chaos. This is clear since the example, $\theta_{n+1} = (\theta_n + p)$ modulo 2π, $p/2\pi$ irrational, is ergodic on the line $p = \text{const.}$ but is nonchaotic (its Lyapunov exponent is zero). In the case of $K > 0$ there are connected regions of positive Lebesgue measure in the (θ, p)-space over which orbits wander chaotically (see Figure 7.13). In this case, for each such region, ergodicity applies (where the relevant measure of a set A is just the fraction of the area (Lebesgue measure) of the ergodic chaotic region in A).

Mixing is defined by (7.54). An example of a nonmixing system is the map, $\theta_{n+1} = (\theta_n + p)$ modulo 2π, $p/2\pi = \text{irrational}$, which is just a rigid rotation of the circle by the angle increment p. An example, which is chaotic but not mixing, occurs when we have a $\tilde{p}/\tilde{q}$ island chain ($\tilde{q} > 1$) for the standard map. In this case, there are typically chaotic sets consisting of component areas (within the $\tilde{p}/\tilde{q}$ island chain), each of which map successively one to another, returning to themselves after $\tilde{q}$ iterates. To see that these chaotic sets are not mixing according to (7.54), let σ' be the area of one of the $\tilde{q}$ components, and σ be the area of another. Then, as m (time) increases, the quantity $[\mu_L(\sigma' \cap \mathbf{M}^m(\sigma)]$ $/\mu_L(\sigma')$ is equal to one once every $\tilde{q}$ iterates and is equal to zero for the other iterates. Hence, the limit in (7.54) does not exist.[8]

A system is said to be a K-system if every partition (see Section 4.5) has positive metric entropy. Basically, in terms of our past terminology, this is the same as saying that the system is chaotic (possesses a positive Lyapunov exponent for typical initial conditions).

A C-system is one which is chaotic and is hyperbolic at every point in the phase space (not just on the invariant set). Examples of C-systems are the cat map, a compact surface of negative geodesic curvature, and the billiard of Figure 7.24(e). An example, which is a K-system, but not a C-system, is the stadium billiard, Figure 7.24(f).

A Bernoulli system is a system which can be represented as a symbolic dynamics consisting of a full shift on a finite number of symbols (see Section 4.1). An example, of such a system, for the case of an area preserving map, is the generalized baker's map with $\lambda_a = \alpha$ and $\lambda_b = \beta$.