4.14 The baker's transformation is two-dimensional and therefore has two Lyapunov exponents, λ_x and λ_y. Determine these exponents by considering the stretching process in the x direction and the contraction in the y direction. (Answer: $\lambda_x = \log_e 2$, $\lambda_y = \log_e a$)

4.15 Another map that can produce horseshoes is the Hénon map:

$$x_{n+1} = 1 - a x_n^2 + y_n,$$

$$y_{n+1} = b x_n.$$

If $b = 1$ then the map preserves areas, and if $|b| < 1$ then the map is dissipative. Write a program to transform an initial block of (x,y) coordinates according to the Hénon map.

4.16 For $g > 1.4$, the pendulum exhibits period doubling. Use the BIFURCATION program through a range of g that covers the doubling region and check the ratio:

$$\frac{g_{n+1} - g_n}{g_{n+2} - g_{n+1}},$$

where g_n is the value of g at the nth bifurcation. Does this ratio appear to approach the Feigenbaum ratio?

---

5   The characterization of chaotic attractors

Many of the geometric structures generated by chaotic maps or differential dynamical systems are extremely complex. For example, the chaotic attractor of the pendulum (Figure 3.3(d)) in its three-dimensional phase space typically consists of an infinite number of infinitely thin layers. Its Poincaré section (Figure 3.5) reveals this structure clearly. These sets are called *fractals*. In this chapter we discuss fractals, their dimensions, and the relation between the fractal character of chaotic attractors and the underlying dynamics.

An elementary example of a fractal is the *Cantor set*. This is a prototype of complex geometric structures in much the same way that the logistic map is a prototype for chaotic dynamical systems. The Cantor set is generated by iteration of a single operation on a line of unit length, as shown in Figure 5.1. The operation consists of removing the
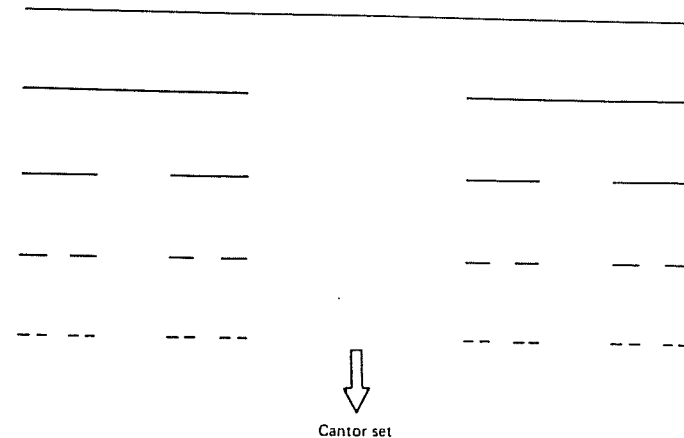

Fig. 5.1 The Cantor set, produced by the iterated process of removing the middle third from previous segments. The Cantor set has zero length and noninteger dimension.

Cantor set

middle third from each line segment of the previous set. As the number of
iterations increases, the number of separate pieces tends to infinity, but
the length of each one approaches zero. Furthermore, if the set is
examined under high magnification, its structure is essentially indistin-
guishable from the unmagnified version. This property of invariance
under a change of scale is called *self-similarity* and is common to many,
although not all, fractals. (Within the resolution of the numerical
simulation, the pendulum's Poincaré section of Figure 3.6 appears to
exhibit this property.)

In contrast to a line with its infinite number of points and finite length,
the Cantor set has an infinite number of points but zero length.
Therefore it seems plausible that the dimension of the Cantor set should
be less than 1 but greater than zero, the dimension of a finite set of points.
The possibility of noninteger dimension requires a more sophisticated
concept of dimension than that associated with lines, surfaces, and solids.

5.1 Dimension

There are many ways to define the dimension, $d(A)$, of a set A. (The
particular set will often be omitted in the notation.) One approach is the
capacity dimension, d_C.

Consider a one-dimensional figure such as a straight line or curve of
length L, as shown in Figure 5.2(a). This line can be 'covered' by $N(\varepsilon)$
one-dimensional boxes of size ε on a side. If L is the length of the line then

$$N(\varepsilon) = L(1/\varepsilon). \tag{5.1}$$

Similarly, a two-dimensional square of side L can be covered by
$N(\varepsilon) = L^2(1/\varepsilon)^2$ boxes as shown in Figure 5.2(b). For a three-dimensional
cube the exponents would be 3, and so on for higher dimensions. In
general,

$$N(\varepsilon) = L^d(1/\varepsilon)^d. \tag{5.2}$$

Taking logarithms one obtains

$$d = \frac{\log N(\varepsilon)}{\log L + \log(1/\varepsilon)}, \tag{5.3}$$

and in the limit of small ε, the term involving L becomes negligible. The
capacity dimension is defined as

(a)

(b)
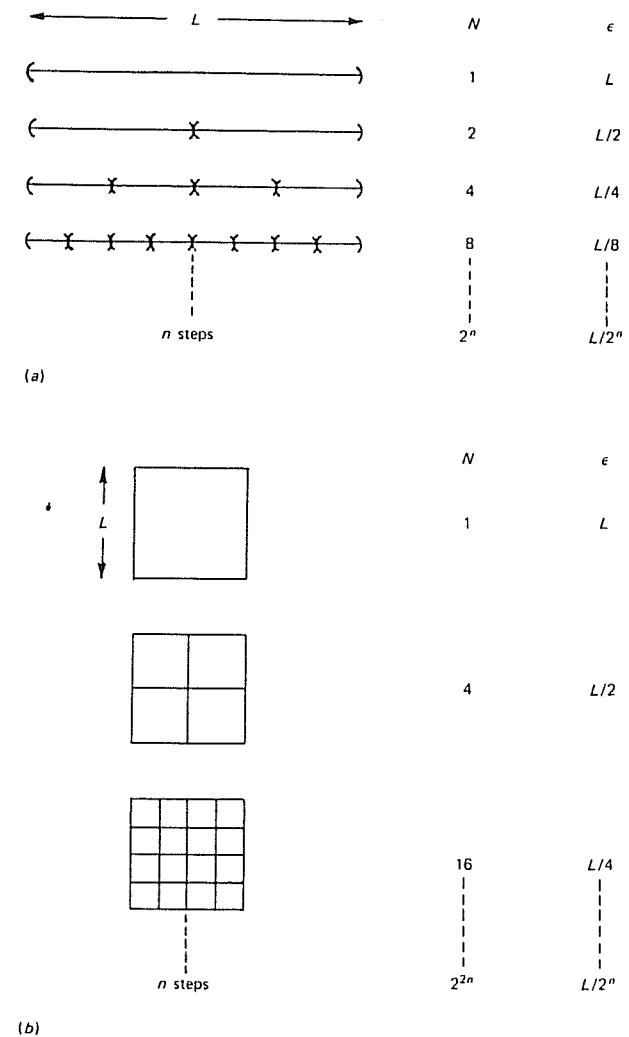
Fig. 5.2 Application of the box covering method to calculate capacity
dimension. Boxes of decreasing size ε are used, leading to increasing numbers
of boxes. The scaling exponent gives the dimension.

$$d_C = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}. \tag{5.4}$$

An equivalent approach is to regard d_C as the slope of the $\log N$ versus $\log(1/\varepsilon)$ curve as $\varepsilon \to 0$. As shown in Figure 5.3 the Cantor set has capacity dimension $d_C = \log 2/\log 3$.

Example 5.1

Let us calculate the dimension of the *Koch* curve. As with the Cantor set of Figure 5.3 we develop a general expression for N and ε

	N	ε
	1	1
	4	1/3
	16	1/9
	⋮	⋮
	4^n	$(1/3)^n$

which leads to $d_C = \dfrac{\log 4}{\log 3} = 2\,\dfrac{\log 2}{\log 3}$. (Note that this result is twice the dimension of the Cantor set.)

For a dissipative dynamical system such as the pendulum, the attractor resides in an n-dimensional phase space, but its dimension is less than n. For example, the periodic limit cycle of a lightly driven pendulum in (θ,ω,ϕ) space is an elliptical spiral with $d = 1$, as seen in Figure 3.3(a). The chaotic attractor of Figure 3.3(d), consisting of an infinite number of closely spaced sheets, has zero volume and d_C between 2 and 3. Chaotic attractors generally have noninteger dimension, and are called *strange attractors*.

Dimension calculations for dynamical systems with periodic forcing may utilize the Poincaré section rather than the full attractor (Moon and Li, 1985). The time or ϕ direction contributes 1 to the full attractor

	N	ε
	1	1
	2	$\frac{1}{3}$
	4	$\frac{1}{9}$
	8	$\frac{1}{27}$
n steps	2^n	$\frac{1}{3^n}$

$$d = \lim_{\varepsilon \to 0} \left(\frac{\log N}{\log(1/\varepsilon)} \right)$$
$$= \lim_{n \to \infty} \left(\frac{\log 2^n}{\log 3^n} \right)$$
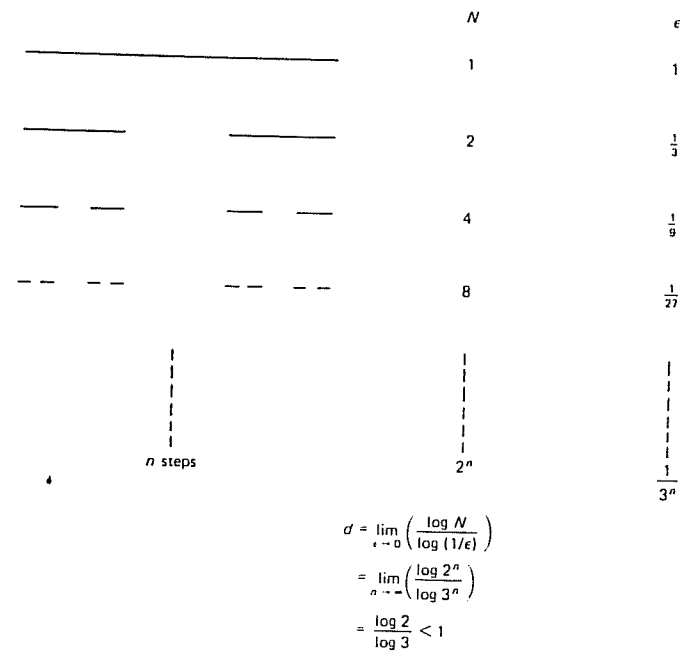$$= \frac{\log 2}{\log 3} < 1$$

Fig. 5.3 Calculation of the capacity dimension for the Cantor set.

dimension, and therefore the dimension of the Poincaré section is $D_C = d_C - 1$. Hence, the Poincaré section of the lightly driven, periodic pendulum has $D_C = 0$ (it's just a point), whereas the dimension of the Poincaré section for chaotic states is between 1 and 2. Furthermore, a transverse cut across the closely spaced lines of the Poincaré section reveals a Cantor-like structure. One can think of this transverse direction as contributing the fractional part to the overall fractional dimension of the Poincaré section.

The capacity dimension is only one of several types of dimension (Farmer, Ott, and Yorke, 1983). However, any reasonable definition is expected to have the following properties:

(a) If $A \subset B$ then $d(A) \leq d(B)$. In the case of the pendulum, the Poincaré section is a subset of the full attractor, and, in agreement with this property, the dimension of the section is less (by 1) than that of the attractor.

(b) The null set has zero dimension.

(c) $d(A \times B) = d(A) + d(B)$. (The symbol × indicates the Cartesian prod-

uct. If $a = \{x_i | i = 1, \cdots, n\}$ and $B = \{y_j | j = 1, \cdots, m\}$ then $A \times B = \{(x_i, y_j) | i = 1, \cdots, n; j = 1, \cdots, m\}$.) For the Cantor set in two-space, the dimension is just the sum of the dimensions of two Cantor sets in one-space:

$$\frac{\log 2}{\log 3} + \frac{\log 2}{\log 3} = \frac{\log 4}{\log 3} \qquad \text{(Problem 5.3)}$$

(*d*) If *f* is a differentiable and invertible function on the set *A*, then $d(f(A)) = d(A)$. For the pendulum the Poincaré section represents the state at a specific phase in the drive cycle. At a slightly later phase, the section is modified according to some unknown function, *f*, acting upon the previous state. Numerical simulation shows that the dimension remains constant, in conformity with this property.

The capacity dimension may be estimated from experimental or numerically generated data. (Implicit in such estimates is the assumption that the finite set of data points is only a sampling of an infinite set. If the data were actually limited to a finite set, then its dimension would be zero.) The technique of *box counting*, previously applied to the Cantor set, can be used for experimental or numerical data. In Figure 5.4, a portion of the pendulum's Poincaré section is shown covered with square boxes of length ε. The required number of boxes of size ε is $N(\varepsilon)$, and ε is varied to obtain $N(\varepsilon)$. From the definition given above, d_C is the slope of the graph of $\log[N(\varepsilon)]$ versus $\log[1/\varepsilon]$ as $\varepsilon \to 0$. For this choice of parameters of the pendulum $d_C = 1.3 \pm 0.1$.

For experimental data or higher-dimensional dynamical systems another type of dimension is more efficient to compute than the capacity dimension. This is the *correlation dimension*, d_G (Grassberger and Procaccia, 1983a). Suppose that many points are scattered over a set. The typical number of *neighbors* of a given point will vary more rapidly with distance from that point if the set has high dimension than otherwise. The correlation dimension may be computed from the correlation integral $C(R)$ defined by

$$C(R) = \lim_{N \to \infty} \left[\frac{1}{N^2} \sum_{i,j=1}^{N} H(R - |x_i - x_j|) \right], \qquad (5.5)$$

where x_i and x_j are points on the attractor, $H(y)$ is the Heaviside function (1 if $y \geq 0$ and 0 if $y < 0$), and N is the number of points randomly chosen from the entire data set. The Heaviside function simply counts the number of points within a radius R of the point denoted by x_i, and $C(R)$ gives the *average* fraction of points within R. Figure 5.5 illustrates the
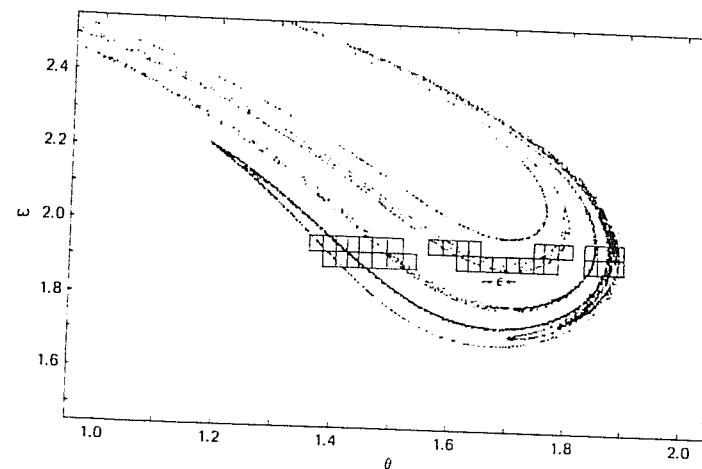

Fig. 5.4 A portion of the pendulum's Poincaré section with some representative boxes covering the points of the section. $g = 1.4954$, $q = 4$.
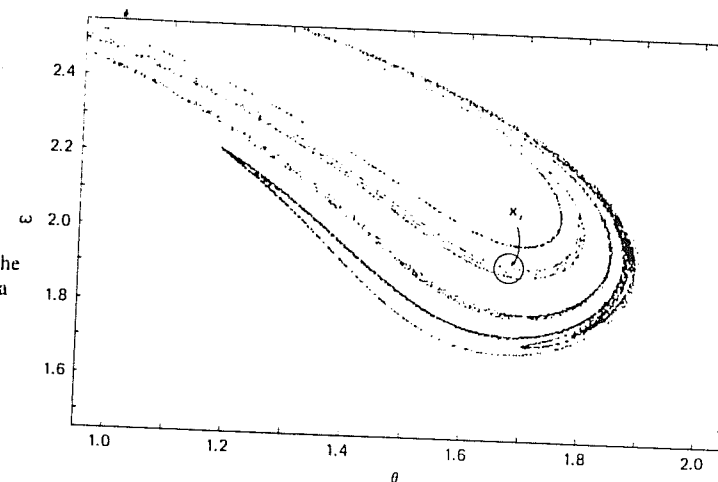

Fig. 5.5 A portion of the Poincaré section with a representative circle centered on the *i*th point. All the other points within the circle are counted by the Heaviside function. $g = 1.4954$, $q = 4$.

method; each point has a circle of radius R drawn around it, and then all points within all circles of size R contribute to $C(R)$. This method of estimating the dimension of a set has the advantage of using less computer memory and computational time. The correlation dimension is defined by the variation of $C(R)$ with R:

$$C(R) \sim R^{d_G} \text{ as } R \to 0. \qquad (5.6)$$

Therefore the correlation dimension is the slope of a graph of $\log C(R)$ versus $\log R$. Figure 5.6 shows such a graph for the driven pendulum, and
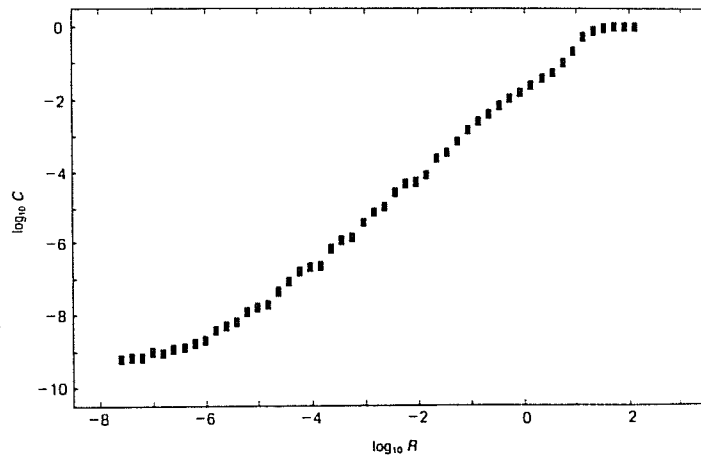
Fig. 5.6 Graph of $\log_{10}C$ versus $\log_{10}R$ for a Poincaré section of the driven pendulum. The correlation dimension is the slope of the straight line portion.

its straight line portion has a slope $d_G = 1.3 \pm 0.1$. The line differs from a straight line for high and low values of R, for the following reasons. When R approaches the size of the phase space, $C(R)$ saturates at unity since all points are then included in the circle. On the other hand, when R is smaller than the spacing between the data points, only one point lies in each circle, and $C(R)$ levels off at $1/N^2$.

The quantities d_C and d_G are not equivalent. The capacity dimension depends only on whether small elements of phase space contain *any* points and does not take into account the differing numbers of points in the various elements. That is, small scale variations of the *density* of points are ignored. On the other hand, the correlation dimension does include this effect. Because of these differences, d_C is called a *metric* dimension and d_G is called a *frequency dimension*. (See, for example, Farmer *et al.* (1983).)

A third quantity, the *information dimension* d_I, is related to the entropy defined in Chapter 4. Like d_G, the information dimension depends on the distribution of points on the attractor. Suppose the attractor is covered by a set of n boxes of size ε, and let the probability that a point is in the ith box be p_i. Then the metric entropy or missing information (see Chapter 4) is

$$I(\varepsilon) = -\sum_{i=1}^{n} p_i \log p_i. \tag{5.7}$$

If the points are equally distributed on the fractal, $I(\varepsilon)$ takes its

maximum value of $I_0 = \log n$. The information dimension d_I is defined by the equation:

$$d_I = \lim_{\varepsilon \to 0}\left[-\frac{I(\varepsilon)}{\log \varepsilon}\right] = \lim_{\varepsilon \to 0}\left(\frac{\sum_{i=1}^{n} p_i \log p_i}{\log \varepsilon}\right). \tag{5.8}$$

The information dimension is the 'scaling exponent' in the variation of the entropy, or missing information, with ε. The remarkable feature of these different definitions of dimension is that they have been shown to be related:

$$d_C \leq d_I \leq d_G. \tag{5.9}$$

If the points are uniformly distributed on the fractal, the equality holds. In practice the numerical values are usually close together (Grassberger and Procaccia, 1983a).

The above definitions of dimension have been consolidated into a general definition of dimensions of *order q*:

$$d^{(q)} = \frac{1}{q-1} \lim_{\varepsilon \to 0}\left(\frac{\log \sum_{i=1}^{n} p_i^{q}}{\log \varepsilon}\right), \tag{5.10}$$

where n is the number of phase space elements and p_i is the probability that an attractor point falls in the ith element. (A concise review is given in Atmanspacher, Scheingraber, and Voges 1988). It can be shown that d_C, d_I, and d_G are equivalent to $d^{(0)}$, $d^{(1)}$, and $d^{(2)}$, respectively (see Problem 5.9 and below). The higher order dimensions are sensitive to nonuniformities in the distributions of points on the attractor.

However, for a uniform distribution of points on the attractor it can be shown that all orders of generalized dimension are the same; namely, $d^{(q)} = d^{(0)}$. The derivation of this result follows directly from the fact that $p_i = 1/n$ for a uniform distribution over all n elements of the phase space for which p_i is nonzero. Then

$$d^{(q)} = \frac{1}{q-1} \lim_{\varepsilon \to 0} \frac{\log \sum_{i=1}^{n} p_i^{q}}{\log \varepsilon}$$

$$= \frac{1}{q-1} \lim_{\varepsilon \to 0} \frac{\log[n(1/n)^q]}{\log \varepsilon}$$

$$= \frac{1}{q-1} \lim_{\varepsilon \to 0} \frac{\log[(1/n)^{q-1}]}{\log \varepsilon}$$

$$= \frac{1}{q-1} \lim_{\varepsilon \to 0} \frac{(q-1) \log(1/n)}{\log \varepsilon}$$

$$= \lim_{\varepsilon \to 0} \frac{\log(n)}{\log(1/\varepsilon)} = d^{(0)}. \qquad (5.11)$$

But in general, $d^{(q)} \geq d^{(q')}$ where $q \leq q'$.

Let us now show the equivalence of the correlation dimension and the generalized dimension of order $q = 2$. That is, we wish to show equality of

$$d^{(2)} = \frac{1}{2-1} \lim_{\varepsilon \to 0} \frac{\log \sum_{i=1}^{N} p_i^2}{\log \varepsilon} \quad \text{and} \quad d_G = \lim_{R \to 0} \frac{\log[C(R)]}{\log R}.$$

The box size ε and the radius R have roughly equivalent roles, and therefore we take $R = \varepsilon$. With this assumption the proof of equivalence amounts to showing that

$$C(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^{N} H(\varepsilon - |\mathbf{x}_i - \mathbf{x}_j|) \text{ is equal to } \sum_{i=1}^{n} p_i^2. \qquad (5.12)$$

Let us start with the correlation integral $C(\varepsilon)$. In the summations for $C(\varepsilon)$, N is both the number of circles (labeled i) and the number of points (labeled j) tested for each circle. Furthermore, the total number of circles, N, should be large enough to cover the entire attractor. In essence the circles form a covering set for the attractor. In the other summation, p_i is nonzero only for those elements of phase space which contain attractor points and therefore the upper limit of the right hand summation is also N. For these conditions

$$\sum_{i=1}^{n} p_i^2 = \sum_{i=1}^{N} p_i^2 \text{ for large enough } N. \text{ Furthermore,}$$

$$\sum_{i=1}^{N} p_i^2 = \sum_{i=1}^{N} p_i p_i = \langle p_i \rangle, \text{ the average value of } p_i, \text{ and}$$

$$\langle p_i \rangle = \frac{1}{N} \sum_{i=1}^{N} p_i \text{ also.} \qquad (5.13)$$

We now connect this result to the circles of the correlation integral.

$$\frac{1}{N} \sum_{i=1}^{N} p_i = \frac{1}{N} \sum_{i=1}^{n} \text{ (probability of points in } i\text{th circle)}$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} \frac{\text{(average number of points in any circle)}}{N}$$

$$= \frac{1}{N^2} \sum_{i>j}^{N} H(\varepsilon - |\mathbf{x}_i - \mathbf{x}_j|), \qquad (5.14)$$

which is the desired result. (We emphasize that this result is rigorously true only as both N and $1/\varepsilon$ approach infinity. See Ding *et al.* (1993b) for a fuller discussion of this derivation.)

Are so many dimensions really needed to describe an attractor? The answer is yes; strange attractors, like many other fractals, contain more geometrical information than can be described by a single number, in much the same way that the mass distribution of an extended object requires an infinite number of parameters (the 'moments' of a mass distribution) for a complete characterization.

In recent studies of strange attractors, an alternative to the use of these generalized dimensions has emerged. It is based on the idea that complex sets may often be described as 'multifractals', that is, sets consisting of many interwoven fractals, each with a different fractal dimension. The multifractal concept, which is now well substantiated experimentally, leads to characterization of the set by a function called the multifractal spectrum. This function contains the same information as the generalized dimensions $d^{(q)}$, so we will not discuss it in this book. However, an accessible account is provided in Glazier and Libchaber (1988). (See also Halsey *et al.* (1986).)

While the calculation of dimension is a common feature of the research literature on strange attractors, misleading results are often obtained, as shown by Arneodo, Grasseau, and Kostelich (1987), and microcomputers are not well suited to these calculations. Therefore we approach the dimension calculation through a remarkable connection between the dimension of the attractor and the time variation of the motion, as characterized by the Lyapunov exponents. In the next two sections we develop this aspect of the subject.

5.2   Lyapunov exponents

In Chapter 4 we introduced the idea of the Lyapunov exponent and gave an example for the one-dimensional map. Constraints on the exponents were also given for chaotic, dissipative systems in higher-dimensional spaces. For these systems, of which the pendulum is an example, the conditions are that the sum of all the exponents must be negative, but at least one must be positive to allow sensitive dependence on initial conditions.

For higher-dimensional systems, the calculation of Lyapunov exponents is more challenging than in the one-dimensional case. However, the idea is the same: measurement of the average rate of divergence of neighboring trajectories on the attractor. (See Wolf (1986) for a review article which gives an overview of the subject.)

The direction of maximum divergence or convergence is a changing local property on the attractor. The motion must be monitored at each point along the trajectory. Therefore, a small sphere is defined whose center is a given point on the attractor and whose surface consists of phase points from nearby trajectories. As the center of the sphere and its surface points evolve in time, the sphere becomes an ellipsoid, with principal axes in the directions of contraction and expansion. The evolution of the ellipsoid is illustrated in Figure 5.7(a). The average rates of expansion or contraction along the principal axes are the Lyapunov exponents. For the ith principal axis, the corresponding exponent is defined as

Fig. 5.7 A two-dimensional example of the calculation of Lyapunov exponents. (a) The evolution of a sphere of initial point to an ellipsoid. (b) Readjustment of the size of vectors along the principal axes. Beginning at t_1, an orthonormal set of vectors from the center of the sphere evolves by stretching and contracting along the axes of the developing ellipsoid. At t_2, a new set of orthonormal vectors is generated such that one of the new vectors is parallel to the previous stretching direction. The process is repeated over many drive cycles to determine average rates of divergence and convergence of nearby trajectories.

(a)

(b)
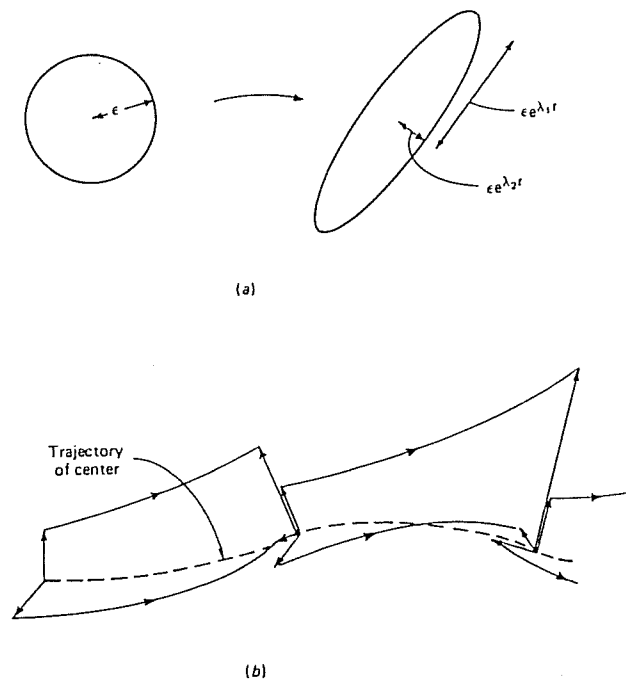
$$\lambda_i = \lim_{t \to \infty} \{ (1/t) \log_e [L_i(t)/L_i(0)] \}, \tag{5.15}$$

where $L_i(t)$ is the radius of the ellipsoid along the ith principal axis at time t (Wolf *et al.*, 1985). In this expression, the growth rate is always measured along the ith principal axis, but the absolute orientation in phase space of that axis is not fixed. It is impractical to perform the actual computation in the way suggested by the definition, because the initially close phase points would soon diverge from each other by distances approaching the size of the attractor, and the computation would then fail to capture the local rates of divergence and contraction. Therefore, vectors connecting the surface of the ellipsoid to the center must be shrunk periodically or *renormalized* to ensure that the size of the ellipsoid remains small and that its surface points correspond to trajectories near that of the center point. The renormalization is illustrated in Figure 5.7(b). The Lyapunov exponents are taken to be the *averages* of those obtained over many segments of the central trajectory.

The algorithm for the Lyapunov exponent calculation in more than one dimension is somewhat beyond the scope of this text, although a computer program that calculates the exponents for the pendulum is included in Appendix B. The method is described in some detail in Wolf *et al.* (1985), and FORTRAN code given in that article is the basis for the program listing.

Nevertheless, it is useful to discuss some of the results for the pendulum. Figure 5.8 shows a typical graph of the three computed Lyapunov exponents for the pendulum, plotted as a function of the number of drive cycles utilized for the computation. The initial transients gradually decay as the calculation is extended over many cycles, so that the structure of the attractor is fully explored.

There are three Lyapunov exponents because the pendulum equations have three variables, and their sum should be negative since the system is dissipative. One exponent corresponds to the direction parallel to the trajectory. It contributes nothing to the expansion or contraction of phase volumes, and therefore the corresponding Lyapunov exponent (λ_2) is zero.

The remaining exponents are negative or zero in the periodic states, whereas in the chaotic state one exponent is positive, indicating divergence of trajectories. For the parameters used to generate Figure 5.8 the system is chaotic, and (from the figure) the exponents are estimated to be: $\lambda_1 = 0.16$, $\lambda_2 = 0$, and $\lambda_3 = -0.42$, with an uncertainty of about 0.02.
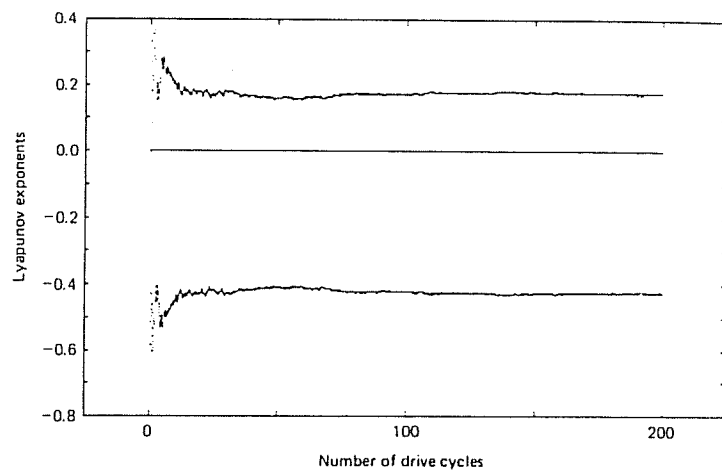
Fig. 5.8 Graph of the three Lyapunov exponents for the chaotic pendulum versus the number of orbits (or drive cycles) used in the computation. Here $g = 1.4954$ and $q = 4$. After a few cycles, the steady state values are reached.

An interesting relationship can be developed between the Lyapunov exponents and the autonomous set of differential equations describing the pendulum. This relation depends on the fact that both the Lyapunov exponents and the differential equations contain information on the change of volumes in phase space, and the relation provides a check on the numerical values developed from the Lyapunov algorithm. First we note that, according to the definition of Lyapunov exponents, a small volume in phase space will change in time as

$$V(t) = V_0 e^{(\lambda_1 + \lambda_2 + \lambda_3)t},\tag{5.16}$$

and therefore the rate of change of $V(t)$ is simply

$$dV/dt = \sum_{i=1}^{c} \lambda_i V(t).\tag{5.17}$$

Furthermore, the discussion of Chapter 2 concerning the change of phase volume in a time δt showed that the time rate of change of the phase volume can also be written as

$$dV/dt = \int \nabla \cdot F \, dV,\tag{5.18}$$

where F is the vector function forming the right side of the autonomous set of first order differential equations for the dynamical system. In the case of the pendulum, $\nabla \cdot F$ equals $-1/q$, independently of position and time. The two expressions for rate of change of volume are combined and lead to the simple result:
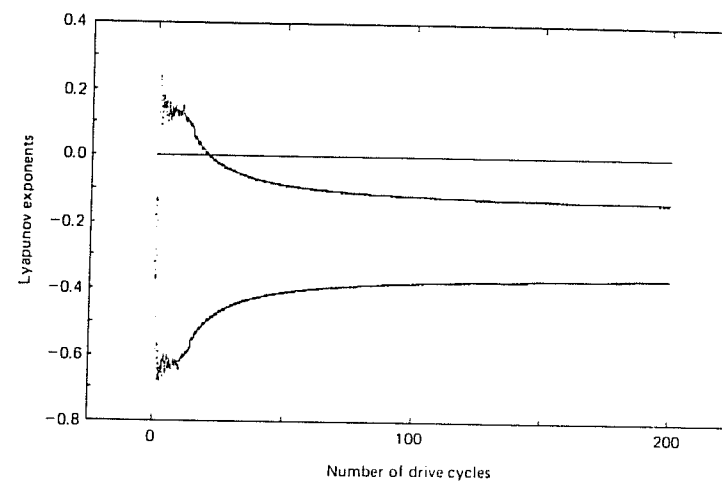
Fig. 5.9 Graphs of Lyapunov exponents computed for a periodic pendulum, with $g = 1.125$ and $q = 2$. The steady state values can be used to check the relation,

$$1/q = -\sum_{i=1}^{3} \lambda_i.$$

$$1/q = -\sum_{i=1}^{3} \lambda_i.\tag{5.19}$$

The exponents obtained from Figure 5.8 with $q = 4$ satisfy this relation. The Lyapunov exponents for a periodic state (Figure 5.9) also satisfy the relation to within numerical precision.

5.3 Lyapunov exponents and dimension

An interesting and important relationship between Lyapunov exponents and dimension was proposed by Kaplan and Yorke (1979). This conjecture can be developed heuristically by reference to Figure 5.10 where a region in (θ, ω) phase space evolves in time under the influence of a dissipative, chaotic system. In this simple example, one direction stretches by a factor of $e^{\lambda_1 t}$ with $\lambda_1 > 0$, and the other direction shrinks according to the factor $e^{\lambda_2 t}$ with $\lambda_2 < 0$. Therefore the area evolves as $A(t) = A_0 e^{(\lambda_1 + \lambda_2)t}$. Now we define a Lyapunov dimension d_L by analogy with the capacity dimension as

$$d_L = \lim_{\varepsilon \to 0} \left[\frac{d(\log(N(\varepsilon)))}{d(\log(1/\varepsilon))} \right],\tag{5.20}$$

where $N(\varepsilon)$ is the number of squares with sides of length ε required to cover $A(t)$. Both $N(\varepsilon)$ and the length ε depend on time as follows:

Fig. 5.10 Schematic diagram of the role of Lyapunov exponents in the stretching and contraction of an area in phase space.
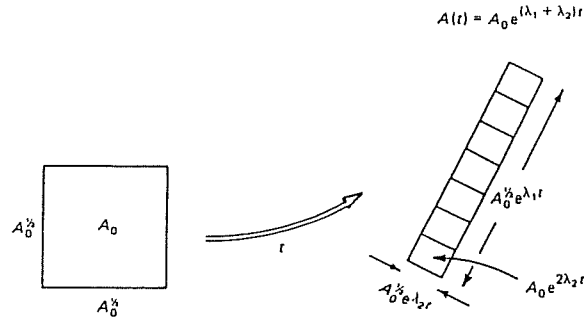
$$N(t) = \frac{A(t)}{\text{square area } (t)} = \frac{A_0 e^{(\lambda_1 + \lambda_2)t}}{A_0 e^{2\lambda_2 t}} = e^{(\lambda_1 - \lambda_2)t} \tag{5.21}$$

and

$$\varepsilon(t) = A_0^{\frac{1}{2}} e^{\lambda_2 t}.$$

Combining these expressions with the definition of d_L and using the chain rule, it follows that

$$d_L = 1 - \frac{\lambda_1}{\lambda_2}. \tag{5.22}$$

This is the Kaplan–Yorke relation. Kaplan and Yorke proposed that $d_L \geq d_G$ and that if the points on the fractal are approximately uniformly distributed, then the equality should hold. (The above derivation depended on a simple example of contraction and expansion in fixed, mutually perpendicular directions. A discussion of the conditions under which the conjecture should be true has been given by Grassberger and Procaccia (1983a).)

The Kaplan–Yorke relation may be generalized to higher-dimensional spaces. The formula becomes

$$d_L = j + \frac{\lambda_1 + \lambda_2 + \lambda_3 + \cdots + \lambda_j}{|\lambda_{j+1}|}, \tag{5.23}$$

where the λ_i are ordered (λ_1 being the largest) and j is the index of the smallest nonnegative Lyapunov exponent. For example, in the phase space (θ, ω, ϕ) the Lyapunov exponents for the pendulum were found to

Table 5.1. *Dimensions of the Poincaré section for various damping factors. All states shown are chaotic. Decreasing the damping allows the Poincaré section to spread and the dimension to increase.* ($g = 1.5$)

q	λ_1	λ_2	$\sum_i \lambda_i$	$\frac{1}{q}$	$d = 1 + \frac{\lambda_1}{|\lambda_2|}$
4.0	0.16	−0.42	−0.26	0.25	1.38
3.7	0.16	−0.43	−0.27	0.27	1.37
3.0	0.11	−0.44	−0.33	0.33	1.25
2.8	0.09	−0.45	−0.36	0.36	1.2
2.0	0.12	−0.58	−0.46	0.5	1.2

be $\lambda_1 = 0.16$, $\lambda_2 = 0$, and $\lambda_3 = -0.42$ (when $g = 1.4954$ and $q = 4$). Therefore $j = 2$ and the predicted dimension is

$$d_L = 2 + \frac{0.16 + 0}{0.42} \approx 2.4.$$

In the two-dimensional subspace (θ, ω) the exponents are $\lambda_1 = 0.16$ and $\lambda_2 = -0.42$ and the calculated dimension is $d_L = 1 + \frac{0.16}{0.42} \approx 1.4$, with an uncertainty of 0.05. (Remember that the ϕ dimension corresponds to time and does not contribute to the expansion or contraction of phase volume.) The result $d_L = 1.4$ correlates well with values of both the capacity and correlation dimension calculated directly for this particular state of the pendulum.

The Lyapunov exponents, in conjunction with the Kaplan–Yorke relation, can be used to study chaotic pendulum states with different amounts of damping. For example, when $q = 2$ the pendulum is highly damped, points in the Poincaré section are tightly packed, and the dimension is small. Table 5.1 shows Lyapunov exponents for chaotic states with various damping factors. The corresponding calculated dimensions are also given. While the changes indicated are not large, there is a general increase in dimension with decreasing damping. However, the pendulum is not chaotic for every value of q in the range shown.

From this discussion it is evident that Lyapunov exponents provide an important link between the fractal geometry of the attractor and the property of sensitive dependence on initial conditions.

5.4   Information change and Lyapunov exponents

The condition of sensitivity to initial conditions that is characteristic of chaotic systems implies divergence of initially adjacent dynamical states. While an initial state of the system may be known with a high but finite degree of precision, the ability to predict later states diminishes because of trajectory divergence. Information is lost, or conversely, more information is required to specify the system with the original precision; the entropy has increased.

For many systems the information function has a simple linear time dependence (Atmanspacher and Scheingraber, 1987)

$$I(t) = I_0 + Kt. \tag{5.24}$$

In this expression, $I(t)$ depends on the probabilities p_i, which change with time as the system evolves. Note that I also depends on the length, ε, of the elemental unit that is used to partition the space. The rate of information change or *Kolmogorov entropy* K is an average rate taken over long times T, although in a numerical calculation T would have to be less than the time taken to cover the entire attractor. Subject to these conditions, K is defined as

$$K = \lim_{\varepsilon \to 0} \lim_{T \to \infty} [I(\varepsilon, T)/T]. \tag{5.25}$$

(See, for example, Grassberger and Procaccia (1984).)

As an illustration of Kolmogorov entropy, consider an imaginary dynamical model which is capable of (a) deterministic, nonchaotic motion, (b) deterministic, chaotic motion, and (c) random motion, for which future states are completely unknown. The value of K will depend on the type of motion. We assume that the phase space of the system may be divided into N small regions, each with probability p_i. Suppose that the system is known to be in a particular initial phase region, and that all initial probabilities are zero except the corresponding probability for that region, which is one. Therefore the initial information (entropy) is zero. The three kinds of motion will cause different types of evolution and therefore different changes in the information function.

For nonchaotic time evolution the system's phase trajectories remain close together. After a time T, nearby phase points are closely grouped in some other small region of phase space and there is no change in information. Therefore the Kolmogorov entropy is zero. For chaotic

evolution phase trajectories diverge, and the number of phase space regions available to the system after a time T is

$$e^{\lambda_+ T}$$

where λ_+ is a positive Lyapunov exponent. Assuming that all of these regions are equally likely, the information function now becomes

$$I(T) = -\sum_{i=1}^{N} p_i(T) \log_e p_i(T) = \log_e e^{\lambda_+ T} = \lambda_+ T. \tag{5.26}$$

Therefore the Kolmogorov entropy in the chaotic case is λ_+. Finally, for random evolution all phase space regions become possible after a very short time, and assuming equal probabilities, the information entropy is

$$I = \log N \tag{5.27}$$

where N is the number of phase space cells. Although N is finite, the fact that each cell has a finite probability of being instantly accessible implies an infinite rate of information expansion. Therefore, in the case of random evolution the Kolmogorov entropy is infinite.

From this discussion it is evident that the Kolmogorov entropy for chaotic systems depends on the positive Lyapunov exponent. In systems with several positive exponents,

$$K \le \sum_{i=1}^{j} \lambda_i \tag{5.28}$$

where j is the index of the smallest positive λ_i. Grassberger and Procaccia (1983a) suggest that equality usually holds. Hence knowledge of the Lyapunov exponents provides a good estimate of Kolmogorov entropy, and positive but finite K implies chaotic behavior. Furthermore, the Kolmogorov entropy provides a means of categorizing the motion of dynamical systems (Atmanspacher and Scheingraber, 1987). If $K = 0$, then the motion is regular (periodic, quasiperiodic, or stationary). If $K > 0$, then the motion is chaotic, and if $K = \infty$, then the motion is random.

Another feature of the Kolmogorov entropy is that it may be used to estimate the time for which future predictions of the state of a chaotic system are valid. After this prediction time, the system's uncertainty has diverged to the size of the phase space (at least in some direction), and the folding action has to occur. Then the system can be described only by a probability density that depends on location in phase space. For purposes of prediction the system becomes probabilistic.

Suppose the system is characterized by a positive Lyapunov exponent, λ_+, and its initial state is defined to within a size ε. Then, in a time, T, the uncertainty in the coordinates will have expanded to the size L of the attractor:

$$L \sim \varepsilon e^{\lambda_+ T} \text{ or } L \sim \varepsilon e^{KT}. \tag{5.29}$$

Either of these relations may be solved for the prediction time:

$$T \sim (1/\lambda_+)\log_e(L/\varepsilon) \text{ or } T \sim (1/K)\log_e(L/\varepsilon). \tag{5.30}$$

Therefore, the prediction time increases only logarithmically with the precision of the initial measurement. For this reason, chaotic states allow only short-term prediction.

In chapter 3 we raised a question about the efficacy of very long computer simulations of chaotic systems when such simulations introduced errors that would apparently be rapidly multiplied in chaotic systems. Our derivation of a formula for prediction time allows us to calculate the effect. For example, if the phase space coordinates are in error by only 1 part in 10^8 of the attractor size and $\lambda_1 = 0.16$ then the prediction time is about 115 time units. A computer simulation that introduces errors of this magnitude will then be accurate for only this relatively short prediction time. We can illustrate the effect by providing an artificial 'error' in the form of a small change of initial conditions. In Figure 5.11 we show time series of two pendula with slightly different initial conditions. The starting values of the angular velocity differ by
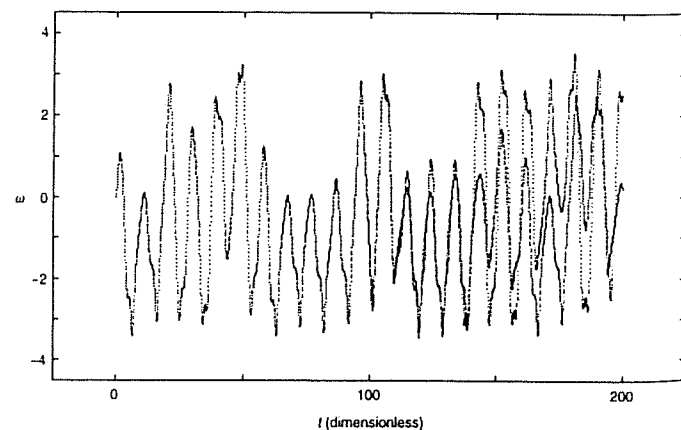

Fig. 5.11 Time series for two pendula whose initial angular velocities differ by $\Delta\omega/\omega = 10^{-8}$. The time for divergence of these systems is estimated to be 115 time units.

4×10^{-8} or one part in 10^8 relative to the attractor size. As expected, the two time series begin to diverge very close to the prediction time.

Does the divergence inherent in chaotic systems mean that strange attractors produced by long simulations are meaningless? The answer appears to be no (Grebogi *et al.*, 1990, Fryska and Zohdy, 1992). For many chaotic systems the 'true' orbits (error-free) are very dense on the attractor in the phase space. Therefore a 'false' trajectory – one from a simulation with inherent error – is often very close to a true orbit that will *shadow* the false orbit. Occasionally false and true orbits will break apart and then another nearby true orbit will shadow the false orbit. The result is that, in many circumstances, an attractor developed from a computer simulation will have the same structure as an attractor for true orbits. While an individual time series from a computer simulation may not match a true orbit, the resultant attractor still has the appropriate geometry as well as the correct dimension, Lyapunov exponents, and other more sophisticated invariant measures.

Problems

5.1 The length of the Cantor set may be determined by subtracting the length of each segment taken out of the set during each step in its formation. For example, a length $\frac{1}{3}$ is taken out in the first step, $2(\frac{1}{9})$ in the second step, $4(\frac{1}{27})$ in the third step, and so on. Form the infinite geometric series which this process describes and show that the sum approaches 1. Therefore the Cantor set has no length.

5.2 Construct a fractal that is similar to the Cantor set, but instead remove the middle $\frac{1}{2}$ from each previous section. Show that its dimension is $\frac{1}{2}$.

5.3 Construct a 'two-dimensional' Cantor set in the following way. Draw two Cantor sets at right angles to each other so they just touch at one corner. Then fill in those two-dimensional regions where each 'one-dimensional' set intersects the other. In other words form the set which is the Cartesian product of the two original sets, as in Figure (*a*) below. Show that the capacity dimension is 2(log2/log3). (This set resembles the invariant set of the horseshoe transformation; that is, the points which remain from the original set after many iterations.)

5.4 Following a procedure analogous to that used in Problem 5.1, calculate the area of the two-dimensional Cantor set defined in
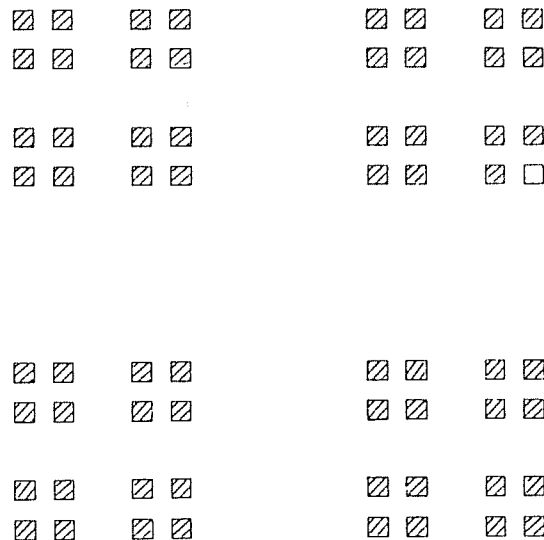
Fig. (*a*)

Problem 5.3.

Fig. (*b*)

5.5 Construct a Cantor-like set by taking squares of relative area $\frac{1}{9}$ out of the center of larger squares. The process is illustrated in Figure (*b*). Show that the dimension of the structure is 3(log2/log3).

5.6 The Cantor set may be used to study some properties of information dimension. Recall that $I = -\Sigma_{i=1}^{N} p_i \log p_i$. Calculate I for each state in the development of the set (as shown in Figure 5.1), assuming that each line segment is equally probable at each iteration. Then use the defining expression for information dimension to show that $d_1 = d_C = \log2/\log3$, in this case.

5.7 Repeat the calculation of Problem 5.6 but do not assume equal probabilities. At every iteration of the set let the right segment have twice the probability of the left segment. For example, when there are four segments the probabilities are, from left to right, $\frac{1}{9}, \frac{2}{9}, \frac{2}{9}$, and $\frac{4}{9}$. Develop the expression for entropy when there are 2^n segments. Show that, in the limit as n tends to infinity, $d_1 = 1 - (\frac{2}{3})(\log2/\log3)$. (Hint: Be careful with the factorial terms.)

5.8 By finding the maximum of the function $I = -\Sigma_{i=1}^{N} p_i \log p_i$, show that d_1 is maximized when $d_1 = d_C$ (See Problem 4.5.)

5.9 Show that the generalized dimensions are equivalent to d_C and d_1 for $q = 0$ and $q = 1$, respectively. (Hint: For $q = 1$, use L'Hôpital's rule.)
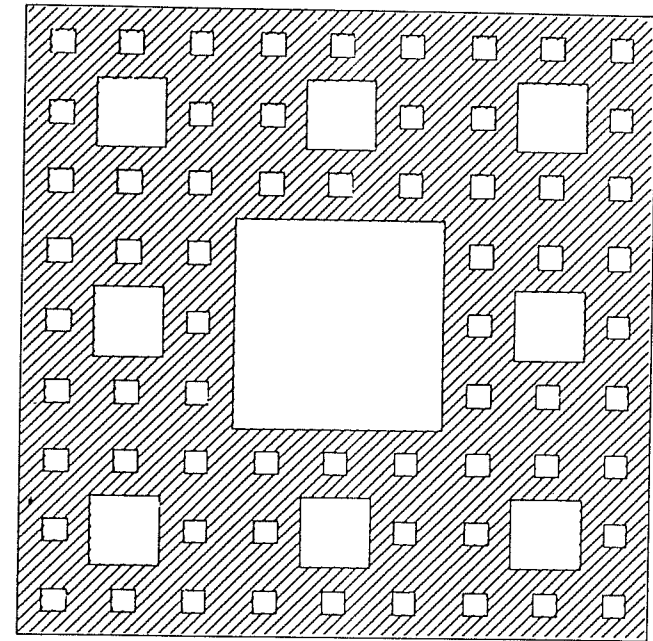
5.10 For a pendulum with $q = 5$ and $g = 1.5$ the Lyapunov exponents are $\lambda_1 = 0.06, \lambda_2 = 0, \lambda_3 = -0.26$. Verify the relation between Lyapunov exponents and the damping factor. Calculate the Lyapunov dimension for the attractor in (θ, ω, ϕ) space.

5.11 Estimate the time for predictability of the pendulum of Problem 5.10, assuming its state is initially known to within 1% of the range of the phase variables. How would you expect this time to change if the damping were increased?

5.12 Moon and Li (1985) suggested a way to estimate Lyapunov exponents. The method is based on the fact that at a saddle point the stable and unstable trajectories meet and define directions of convergence and divergence of trajectories, respectively. The rates of these behaviors approximate the respective Lyapunov exponents. To study the behavior of the trajectories in the saddle point region the equation of motion is linearized at the saddle point and then solved. Furthermore it is only necessary to solve the homogeneous part of the differential equation because the saddle point is a fixed point in any Poincaré section taken at frequency ω_D. An examin-

ation of Poincaré sections taken at various phases shows that $\theta = \pm \pi$ appears to be the θ saddle point coordinate and therefore the required homogeneous, linearized equation is

$$d^2\theta/dt^2 + (d\theta/dt)/q - \theta = 0.$$

Substitute the solution $\theta = e^{mt}$ into the differential equation to obtain the relation

$$m = -\frac{1}{2q} \pm \left(\frac{1}{4q^2} + 1\right)^{\frac{1}{2}},$$

which has both positive and negative values. These values are the estimates of λ_+ and λ_-. Check that they sum to $\nabla \cdot F$. Substitute these values into the Kaplan–Yorke relation for d_L, thereby obtaining a relation for d_L in terms of q. Find d_L in the limiting cases where $q = 0$ (infinite damping) and $q = \infty$ (no damping). Calculate d_L in the cases where $q = 2$ and $q = 5$. Do your results match those given in the text and developed in a previous problem? If not suggest a reason. (Moon and Li change the value of the damping parameter to an *effective* value giving numerical results that better match the dimension. If you were to follow a similar procedure, what value of q would you choose for essential agreement with the numerically derived d_L?)

5.13 Use the program PENDLYAP in Appendix B to calculate Lyapunov exponents in various cases. Examine both chaotic and nonchaotic states. Use the bifurcation diagrams to choose the appropriate values of g.

6

Experimental characterization, prediction, and modification of chaotic states

Chaotic states occur widely in natural phenomena, but closed form mathematical models are rarely available. This situation leads to a number of related problems that bear on the use of experimental data. First, how is it possible to tell whether a set of apparently noisy data in fact arises from chaotic dynamics? Second, how can chaotic data be used to make short-term predictions or forecasts? Finally, how can experimental data be used to influence and control nonlinear systems? We address these questions in the present chapter.

6.1 Characterization of chaotic states

In this section, we consider the use of experimental data to *test* for the existence of chaos, to *reconstruct* the chaotic attractor if it exists, and to *characterize* its structure quantitatively.

One rarely has complete information about all of the degrees of freedom in a complex dynamical system. For example, in a chaotic fluid system, this information would include the velocity of the fluid at many different positions as a function of time. Even for the pendulum, a complete specification would seem to require measurement of three distinct time-dependent quantities (the angle, the angular velocity, and the phase of the forcing function).

Although this information can easily be obtained for a pendulum, it is more informative to use the pendulum to learn how to handle situations where a fuller description is unavailable. It is frequently possible to learn a considerable amount from a single *time series*, a list of successive values of one dynamical quantity. This is the case even if the measured variable is not one that appears in the dynamical equations describing the system. Consequently, we focus attention on methods for analyzing time series.

There are quite a few conventional approaches to analyzing time series. One approach, developed for linear systems, is based on spectral analysis, and has been discussed in Chapters 1 and 3. The power spectra of oscillatory *linear* systems can often be interpreted directly in terms of normal modes of oscillation. Furthermore, some of the measurement noise typically occurs outside the frequency range of interest; if so, it can be removed by filtering.

On the other hand, the power spectra of nonlinear systems are more problematic to interpret. It is sometimes difficult to distinguish conclusively between (*a*) a deterministic periodic oscillation that appears noisy due to a *fluctuating parameter*, and (*b*) a chaotic oscillation. Filtering generally does not help; it may destroy information about the dynamics. (Kostelich and Schreiber, 1993, Abarbanel *et al.*, 1993)

If a dynamical model is available, one can imagine simply computing the variable that is measured and comparing the computed and measured time series. But how should such a comparison be carried out? They cannot be expected to agree, both because of sensitive dependence on initial conditions, and also because the parameters of the model may not be known sufficiently well. One could compare the statistical properties (such as moments and probability distributions) of computed and measured time series, but even this approach is difficult to carry out successfully.

Consequently, it is useful to consider methods of analyzing experimental data that do not rely on a particular model, but which can be used to detect chaotic dynamics and characterize it quantitatively. Methods for dynamical analysis of experimental time series are still developing, but a common method is a two-step process: (1) reconstruction of the strange attractor of the unknown dynamical system from the time series, and (2) determination of certain invariant quantities of the system from the reconstructed attractor. These invariants might include one or more Lyapunov exponents and the dimension of the attractor. In principle, one can then compare these invariants to their corresponding values as derived from some model system. A close match would lead to

further insight into the dynamics of the experimental system.[†] However, the success of this approach has generally been limited.

In the next few sections of this chapter we describe some of these techniques of dynamical data analysis using a time series for $\omega(t)$, the angular velocity of the driven pendulum. A typical example is shown in Figure 6.1 where the angular velocity values are taken at intervals of 7 ms. We use experimental data provided by James A. Blackburn, coinventor (with H.J.T. Smith) of the mechanical pendulum whose motion is simulated by equations that are essentially the same as those used in this book (Blackburn *et al.*, 1989).[‡] Throughout this chapter reference to experimental data implies use of this data set. Reference to simulated data, both periodic and chaotic, implies use of data obtained from numerical solution of the differential equation for the pendulum.

The experimental time series becomes the basis for reconstruction of the attractor and calculation of the dimension and positive Lyapunov exponent. The reconstructed attractor and the invariant quantities are then compared to those same quantities computed directly from the pendulum equations using the methods of Chapter 5.

6.1.1 Experiment and simulation

The Blackburn pendulum is constructed so as to be modeled by the equation

$$I\frac{d^2\theta}{dt^2} + b\frac{d\theta}{dt} + \omega_0{}^2 I\sin\theta = T\sin\omega_f t, \tag{6.1}$$

where I is the moment of inertia of the physical pendulum, b is a damping factor, ω_0 is the natural frequency, T is the amplitude of the forcing torque, and ω_f is the forcing frequency.

The main parts of the pendulum are shown in Figure 6.2 (Blackburn *et al.*, 1989.) A small pendulum bob is attached to a shaft with damping and forcing mechanisms. Damping occurs through the induction of eddy currents in a copper plate whose position may be adjusted relative to a magnet attached to the moving shaft. The forcing frequency and amplitude are adjusted electronically. Optical encoding of the pendulum's motion allows data to be transmitted to a computer for storage on

---

[†] For example, Ciliberto and Gollub (1985) compared a model of chaotic surface waves with experimental results.

[‡] This pendulum is produced commercially by the Daedelon Corporation and the data for this chapter were gathered from a production model.
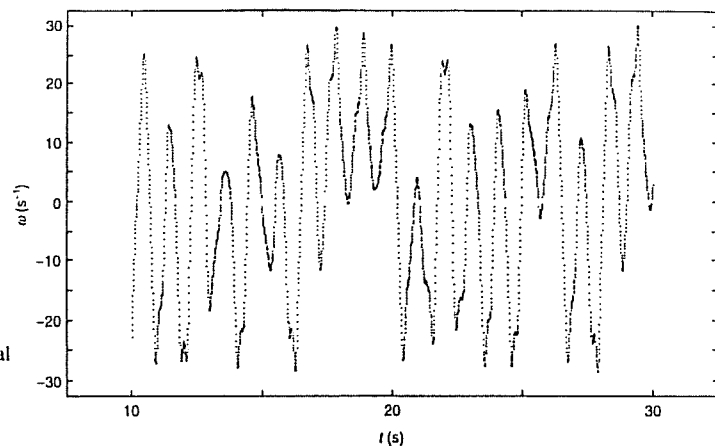
Fig. 6.1 An experimental time series of the angular velocity of the driven pendulum. The sampling interval is 7 ms.

Fig. 6.2 Exploded view of the experimental pendulum. Approximate dimensions are 5 cm × 9 cm × 9 cm. Labeled components are: M, pendulum mass; A, code wheel; B, copper damping plate; C, ring magnet; D, encoder module; E, motor coils; and F, micrometer. (From Blackburn et al., 1989, reprinted with permission.)

diskette and eventual analysis. Procedures described in the manual that accompanies the commercial version are used to calibrate the pendulum and thereby to obtain conditions that correspond to the chosen parameter values.

With the exception of a phase difference of $\pi/2$ in the forcing term, equation (6.1) may be transformed to our previous dimensionless form,

$$\frac{d^2\theta}{dt'^2} + \frac{1}{q}\frac{d\theta}{dt'} + \sin\theta = g\cos\omega_D t', \qquad (6.2)$$

by a time transformation $t' = \omega_0 t$. (See Problem 6.1.) As a result, the dimensionless parameters q, ω_D, and g are expressed in terms of the physical parameters b, I, T, ω_0, and ω_f. (See Problem 6.2.) All simulations and experimental data from the pendulum were obtained so that the parameters of equation (6.1) corresponded to the values $g = 1.5$, $\omega_D = 2/3$, and $q = 4$, in equation (6.2). In this way experimental data may be directly compared with some of the figures presented in Chapters 3 and 5. The corresponding parameters for the physical pendulum are $b/I = 2.24 \pm 0.1$ s^{-1}, $T/I = 121 \pm 6$ s^{-2}, $\omega_0 = 8.98 \pm 0.05$ s^{-1}, and $\omega_f = 5.98 \pm 0.02$ s^{-1}.

An initial comparison of the experimental pendulum and its simulation using equation (6.1) may be made by comparing their Poincaré sections shown in Figure 6.3. While there are some similarities, the experimental Poincaré section is somewhat noisy and contains features that are not in the simulated section. Despite these differences, we will see that the
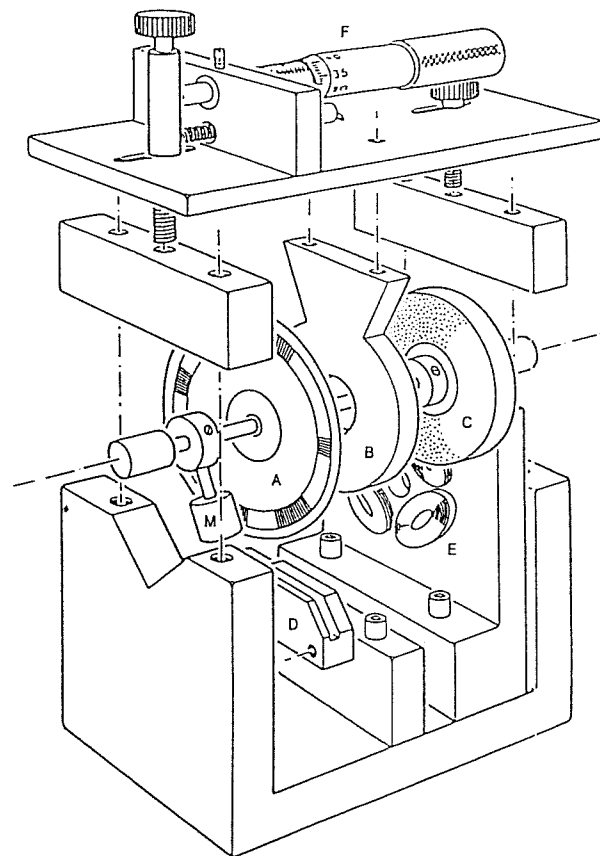
discrepancies between the corresponding invariant quantities are remarkably small.

6.1.2 Reconstruction of the attractor

In this experiment we *could* have access to all of the variables. Nevertheless we do not utilize this complete set of information in order to illustrate situations where only a single time series is obtainable. A single experimental time series is affected by all of the relevant dynamical variables, and therefore it contains a relatively complete historical record of the dynamics. It is possible to glean the dynamics from a single

Fig. 6.3 (a)
Experimentral Poincaré
section and (b)
simulation using
equation (6.1). The
parameters for the
simulation are
$h/I = 2.245$ s^{-1},
$T/I = 120.691$ s^{-2},
$\omega_0 = 8.98$ s^{-1}, and
$\omega_f = 5.987$ s^{-1}.

(a)

(b)
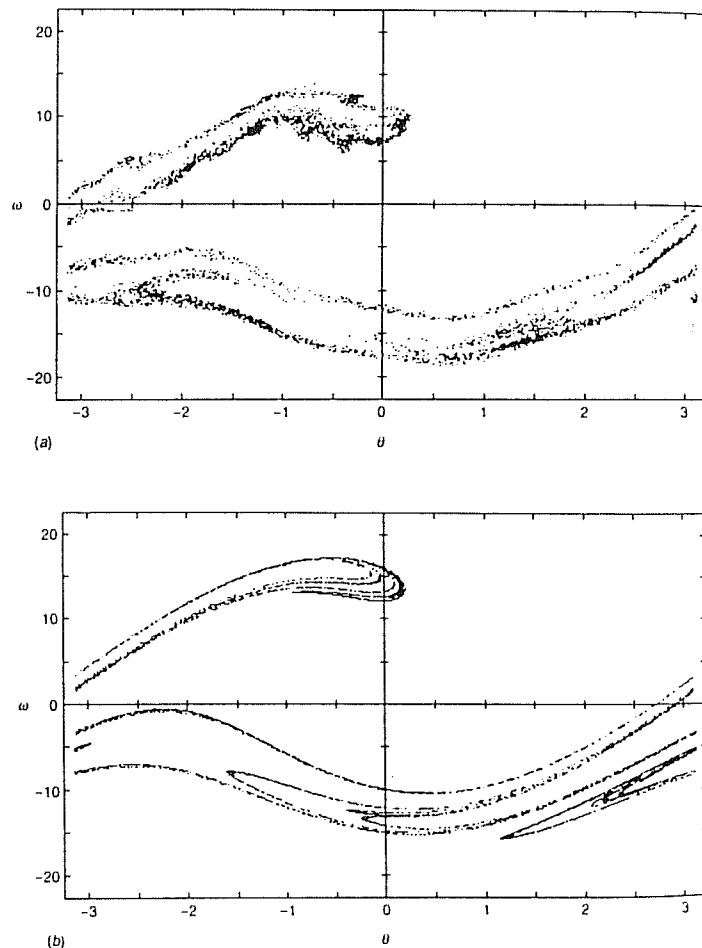
time series without reference to other physical variables. This concept was first illustrated by Packard *et al.* (1980) and given a rigorous mathematical basis by Takens (1981) and Mané (1981).

Reconstruction of the attractor from a single time series requires the generation of additional variables. For some experimental systems the effective number of degrees of freedom is relatively small. Consequently, it is possible to define a low-dimensional 'phase' space that captures the dynamics in a geometric structure embedded in that space. The embedded geometric set is called the *reconstructed attractor* and it is

usually topologically equivalent to that attractor which would be produced by numerical solution of the dynamical system equations if they were known. In particular, the dimension and Lyapunov exponents are usually approximately the same for both the original and reconstructed attractors.

The dimensionality of the required phase space is frequently not known in advance; here it is 3. It must be relatively small for reconstruction to be practical. In choosing the coordinate axes of this space we look for relatively independent pieces of information. For example, we might choose the time series variable $\omega(t)$ and its $(n-1)$ derivatives to span an n-dimensional space. If data points are separated by an interval Δt then the derivatives may be approximated by finite differences:

$$\left.\begin{aligned}\frac{\mathrm{d}\omega}{\mathrm{d}t} &\approx \frac{\omega(t+\Delta t)-\omega(t)}{\Delta t}, \\ \frac{\mathrm{d}^2\omega}{\mathrm{d}t^2} &\approx \frac{\omega(t+2\Delta t)-2\omega(t+\Delta t)-\omega(t)}{\Delta t^2},\end{aligned}\right\} \tag{6.3}$$

and so on. In practice, high order derivatives are quite noisy. It is simpler and better to use the set of coordinates introduced in Chapter 4; namely, *time delay coordinates*, where each point in the space is an ordered n-tuplet of consecutive values of the time series: $(\omega(t), \omega(t+\tau), \omega(t+2\tau), \cdots, \omega(t+[n-1]\tau))$. The time τ is typically some multiple of the spacing Δt between the time series points.

In the next sections we discuss the mechanics of attractor reconstruction in the phase space of time-delay coordinates, and calculation of the attractor dimension and the largest Lyapunov exponent.

6.1.3 Time-delay coordinates

Consider the time series $\{\omega(t_1), \omega(t_2), \cdots\}$. Successive points in the phase space formed from time-delay coordinates can be written as vectors, $\mathbf{y}_j$;

$$\left.\begin{aligned}\mathbf{y}_1 &= (\omega(t_1), \omega(t_1+\tau), \omega(t_1+2\tau), \cdots, \omega(t_1+(n-1)\tau)), \\ \mathbf{y}_2 &= (\omega(t_2), \omega(t_2+\tau), \omega(t_2+2\tau), \cdots, \omega(t_2+(n-1)\tau)), \\ &\vdots \\ \mathbf{y}_j &= (\omega(t_j), \omega(t_j+\tau), \omega(t_j+2\tau), \cdots, \omega(t_j+(n-1)\tau)), \\ &\vdots\end{aligned}\right\} \tag{6.4}$$

(Note that $(t_{j+1}-t_j)=\Delta t$ whereas $\tau=$ integer$\cdot\Delta t$.) Because the time series is presumed (by hypothesis) to be the result of a deterministic process, each y_{j+1} is the result of a mapping, $\mathscr{M}$ – generally unknown – from the vector, y_j. That is,

$$y_{j+1}=\mathscr{M}(y_j). \tag{6.5}$$

Note the similarity of this relation to that defining a Poincaré section and to the other mappings we have discussed.

Often the appearance of reconstructed and numerically simulated attractors are similar. Yet sometimes the attractors may, for various reasons, look very different. We use the Lorenz model as an example of the first case and the pendulum as an example of the second case.

The Lorenz model

Let us consider the dynamical system introduced by Edward Lorenz (1963) as a simple model of convection. These equations were the first to illustrate chaotic behavior in a model inspired by fluid dynamics. They have the form

$$\left.\begin{array}{l} dx/dt=-\sigma(x+y), \\ dy/dt=-xz+rx-y, \\ dz/dt=xy-bz, \end{array}\right\} \tag{6.6}$$

where σ, r, and b are parameters that characterize the properties of the fluid and of the thermal and geometric configuration of the experiment. The variable x is related to the fluid's *streamfunction* (a function which characterizes the fluid flow), y is proportional to the temperature difference between the upward and downward moving parts of a convection roll, and z describes the nonlinearity in temperature difference along the roll. Numerical solution of these equations with parameter values $\sigma=10$, $r=28$, and $b=8/3$ leads to an attractor embedded in a three-dimensional space with coordinates (x,y,z), as shown in Figure 6.4(a). The trajectory rotates about one of two unstable fixed points (where the derivatives are zero) and eventually escapes to orbit the other fixed point. This process is repeated indefinitely.

An attractor of similar appearance can be reconstructed from the time series of $x(t)$ alone. Figure 6.4(b) shows the reconstructed Lorenz attractor in a space with coordinates, $x(t)$, $x(t+\tau)$, and $x(t+2\tau)$. (The choice of delay τ is a nontrivial one and some guidance is provided in the next section.) Note that both figures have the same 'butterfly' appearance.
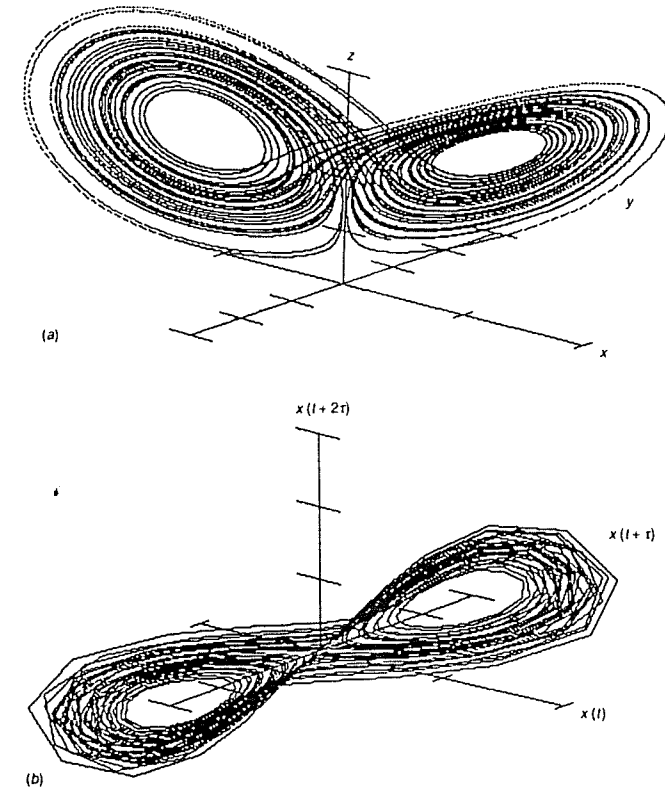

Fig. 6.4 Chaotic attractors for the Lorenz system. (a) Attractor generated from numerical solution of the differential equations. (b) Attractor reconstructed with time-delay coordinates from the time series for $x(t)$.

This qualitative geometrical similarity is also found for many other dynamical systems.

The pendulum

We recall that the pendulum attractor is portrayed in the three-dimensional space with coordinates (ω,θ,ϕ) and that both θ and ϕ have periodic boundary conditions. These periodic boundary conditions cause the original and reconstructed attractors to differ qualitatively even for the nonchaotic pendulum as shown in Figure 6.5. Figure 6.5(a) shows the numerically simulated spiral attractor for a periodic state of the pendulum. If the periodic motion becomes more complex through period doubling, additional open-ended spirals appear along with the original one, as shown in Figure 3.3.
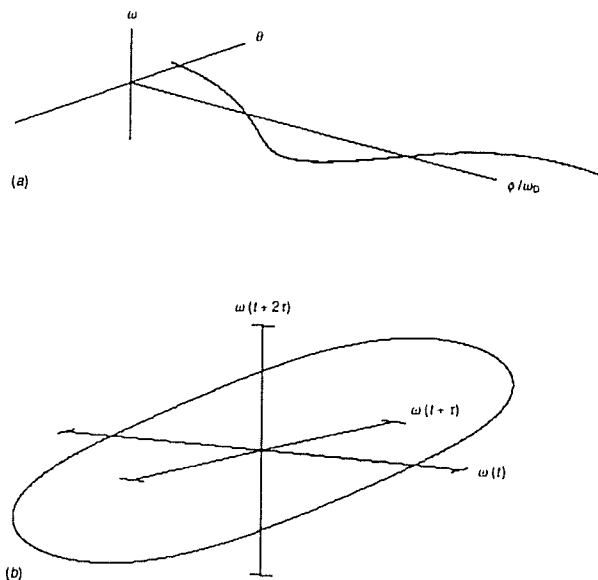
Fig. 6.5 Periodic attractors for the driven pendulum. (*a*) The attractor is generated from numerical solution of the differential equations. Because of the periodic boundary conditions on the time variable $\phi = \omega_D t$, the orbit is actually closed. (*b*) Attractor generated by the periodic time series $\omega(t)$.

A single time series of the angular velocity from the numerical simulation of Figure 6.5(*a*) can be used to reconstruct the attractor. (An experimentally determined time series of the angular velocity for a periodic pendulum would give the same reconstructed attractor.) Unlike θ and ϕ, this coordinate does not have periodic boundary conditions, and therefore has no discontinuities in its time series. In the regime of small forcing amplitude, $\omega(t)$ is periodic and its reconstructed attractor, shown in Figure 6.5(*b*), is a *closed* loop. Period doubling is easily recognized in the reconstructed attractor as a closed loop of two turns (one twist), whereas in (θ,ω,ϕ) space the attractor consists of two spirals. Therefore, even in the periodic regime, the original attractor and the reconstructed attractor have a different appearance.

In the chaotic regime the contrast between the numerically simulated attractor with coordinate axes (θ,ω,ϕ) and the attractor reconstructed from an experimental time series $\omega(t)$ is quite marked. The attractor generated from a numerical simulation of equation (6.1) is shown in Figure 6.6(*a*), and it is very similar to the attractor of Figure 3.3(*d*) which is a simulation of equation (6.2). On the other hand, the attractor reconstructed from experimental data, shown in Figure 6.6.(*b*), is quite different in topology from either of the previous figures. As in the periodic case the differences in appearance are primarily due to the
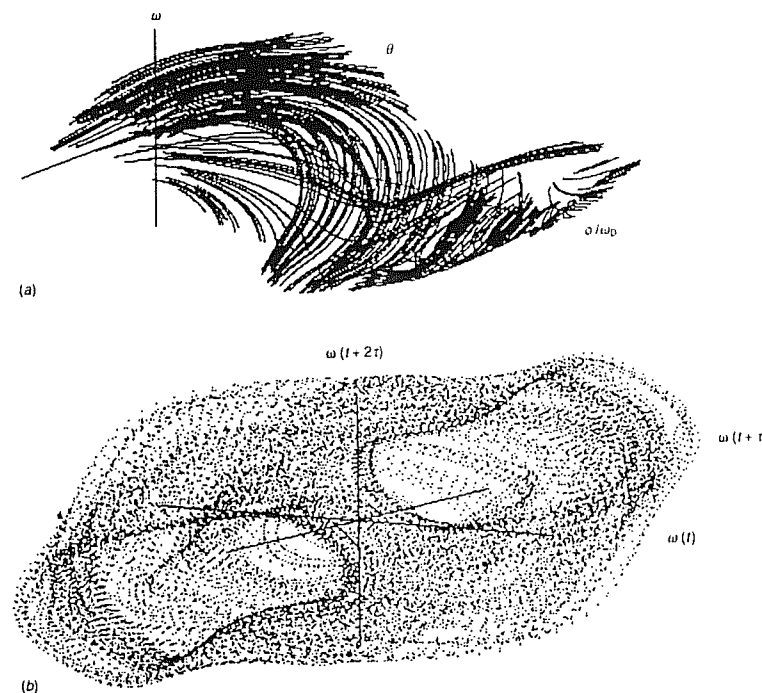
Fig. 6.6 Chaotic attractors for the driven pendulum. (*a*) Attractor generated from numerical solution of the differential equations with periodic boundary conditions on *t* and the angle *θ*. (*b*) Attractor reconstructed from the time series $\omega(t)$. It consists of an infinite number of orbits that never completely close.

existence of periodic boundary conditions in the original attractor, but not in the reconstructed attractor.

6.1.4   Choosing the time delay

The reconstruction of the attractor with time-delay coordinates is not an automatic process. The choice of an appropriate delay τ is important to the success of the reconstruction. If τ is too short then the coordinates $\omega(t_j)$, $\omega(t_j+\tau)$, $\omega(t_j+2\tau)$, ... of a given vector y_j are almost equal to each other. The tip of each vector in the space of delay coordinates will lie near the diagonal, and the reconstruction will be useless. In essence, the sampled points from the time series for $\omega(t)$ are too closely spaced to provide information about the dynamics of the system. (The use of very closely spaced points is somewhat equivalent to the use of derivatives as phase coordinates – a process described earlier. Coordinates of derivatives tend to produce high levels of noise (Eckmann and Ruelle, 1985).)

On the other hand, we know from Chapter 5 that correlations between dynamical states in chaotic systems last for a relatively short time that depends on the positive Lyapunov exponents. If τ is too large then the coordinates are so far apart as to be uncorrelated. Significant folding and stretching will have occurred and there is no causal relation between the data used to form the coordinates of a point in phase space.

Several methods for choosing τ have been suggested (Fraser and Swinney, 1986, Fraser, 1989, Albano *et al.*, 1988). If the system has some rough periodicity, as occurs with the forced motion of the pendulum, then a value comparable to but somewhat less than that period is typically chosen. When there is no dominant period more sophisticated methods are used. One straightforward approach is to examine the correlation between pairs of data points as a function of their time separation. Define a correlation function

$$f(\mathscr{T}) = \frac{\langle \omega(t)\cdot\omega(t+\mathscr{T})\rangle_t}{\langle \omega(t)^2\rangle_t}, \tag{6.7}$$

where $\langle\ \rangle_t$ denotes an average over all data points in a time series. Then determine the time $\mathscr{T}_0$ of the first zero crossing of $f(\mathscr{T})$ as a measure of the correlation time. Since we seek a value of τ that yields high correlation and yet also allows some time development of the system, a modest fraction of $\mathscr{T}_0$ would be a reasonable choice. Figure 6.7 shows the correlation function for the angular velocity from the experimental chaotic pendulum. The first zero crossing occurs at about $\mathscr{T}_0 = \frac{1}{3}$ s which, since the data sampling time is 7 ms, corresponds to a delay of 47 points.
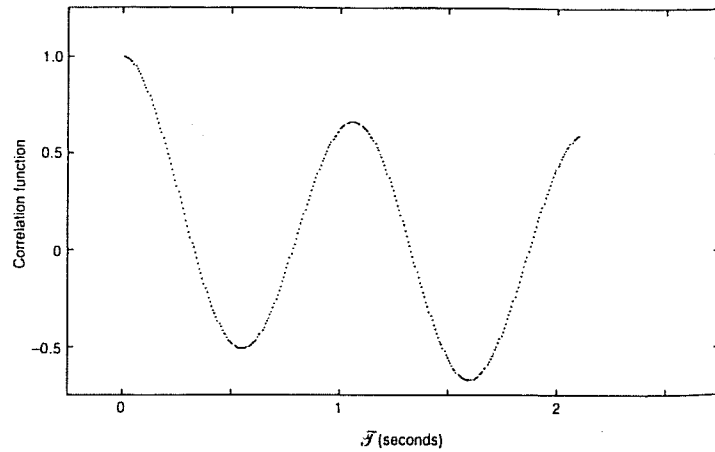
While a typical correlation function might be expected to decay monotonically, the correlation function in Figure 6.7 has a slowly decaying oscillation at the forcing period of 0.95 s. It eventually decays to zero for sufficiently large $\mathscr{T}$. In general, the correlation function and the Fourier spectrum provide the same information (see, for example, Press *et al.* (1986)). This oscillation of the correlation function is related to the strong spectral peak at the forcing frequency, shown in Figure 3.9.

Fraser and Swinney (1986) use a more sophisticated approach for selecting τ that involves the concept of mutual information. Mutual information provides a measure of the probability that neighboring data points, $\omega(t)$ and $\omega(t+\mathscr{T})$, are not statistically independent. If $P(\omega(t))$ and $P(\omega(t+\mathscr{T}))$ are separate probability distributions of the neighboring data points, and if $P(\omega(t), \omega(t+\mathscr{T}))$ is the joint probability distribution then the mutual information may be defined as

$$I(\mathscr{T}) = \sum_i P(\omega(t),\ \omega(t+\mathscr{T}))\ \log\frac{P(\omega(t)\cdot\omega(t+\mathscr{T}))}{P(\omega(t))\cdot P(\omega(t+\mathscr{T}))}. \tag{6.8}$$

If $\mathscr{T}$ is large then the joint probability is equal to the product of the separate distributions and the mutual information is zero. Therefore the delay τ is sometimes chosen to be equal to $\mathscr{T}_{min}$ where the mutual information achieves its first minimum. However, in some cases where there is no minimum (such as with the Hénon map) other criteria are used (Abarbanel *et al.*, 1993.)

Because there is no simple rule for choosing τ in all cases (Albano *et al.*, 1988, Abarbanel *et al.*, 1993), investigators sometimes adjust τ until the results seem satisfactory. This procedure could introduce bias, but the invariant quantities computed from reconstructed attractors are often not too sensitive to τ (within a reasonable range) if the number of degrees of freedom is small. For the reconstructions in this chapter a value of τ corresponding to about $\frac{2}{5}\mathscr{T}_0$ seems to work well. The experimental configuration recorded data every 7 ms and therefore the delay τ is about 140 ms, or 20 data points. Since the forcing frequency was 0.953 Hz this delay is 15% of the forcing period, $2\pi/\omega_F$.

Fig. 6.7 Correlation function for the time series $\omega(t)$ for the pendulum.

6.1.5 Embedding dimension and attractor dimension

The Lorenz attractors in Figure 6.4 are well represented in a three-dimensional space. Similarly we represent the pendulum attractor in the three-dimensional space (ω,θ,ϕ) and therefore suspect that its delay

coordinate reconstruction can also be fully represented in a three-dimensional space. However, in reconstructing an attractor from an arbitrary experimental time series of unknown dynamics, the dimensionality of the attractor is unknown. It is important that the reconstruction be *embedded* in a space of sufficiently large dimension to represent the dynamics completely.

One criterion for determining this dimension utilizes the *noncrossing* property (Chapter 2) that ensures the deterministic nature of trajectories. The map for the reconstruction in time-delay coordinates,

$$y_{j+1} = \mathcal{M} y_j, \tag{6.9}$$

must be invertible since it is the result of some underlying dynamical system. Takens (1981) suggested that it is sufficient for the embedding dimension D to be greater than the attractor dimension d according to the relation

$$D \geq 2d + 1. \tag{6.10}$$

Figure 6.8 illustrates the situation for a hypothetical one-dimensional attractor. The embeddings in one- and two-dimensional spaces lead to spurious crossings that suggest noninvertibility of the map. Furthermore, spurious crossings lead to spurious near neighbors – an effect that produces errors in calculation of the Lyapunov exponent. Only the three-dimensional phase space displays this attractor fully.

(Calculation of the correlation dimension does not require that D be large enough to satisfy equation (6.10). Ding *et al.* (1993a) have shown that, for long time series with little noise, it is often sufficient that the phase space dimension be greater than the attractor dimension. Therefore the stringent Takens condition may not be required for dimension calculations.)

Nevertheless, the attractor dimension is usually unknown for experimental data and therefore the required embedding dimension is unknown. Fortunately there is a procedure that yields both quantities simultaneously. If the attractor is embedded in spaces of increasingly higher dimension it exhibits an increasingly complex structure as it unfolds, and its measured dimension increases. This progression continues until the attractor is fully revealed, at which point the dimension saturates (in principle). When this condition is reached the embedding dimension is sufficiently high.

The process (first proposed by Grassberger and Procaccia (1983b)) is as follows. We reconstruct the attractor in a low-dimensional space and calculate its apparent dimension. The correlation dimension, described
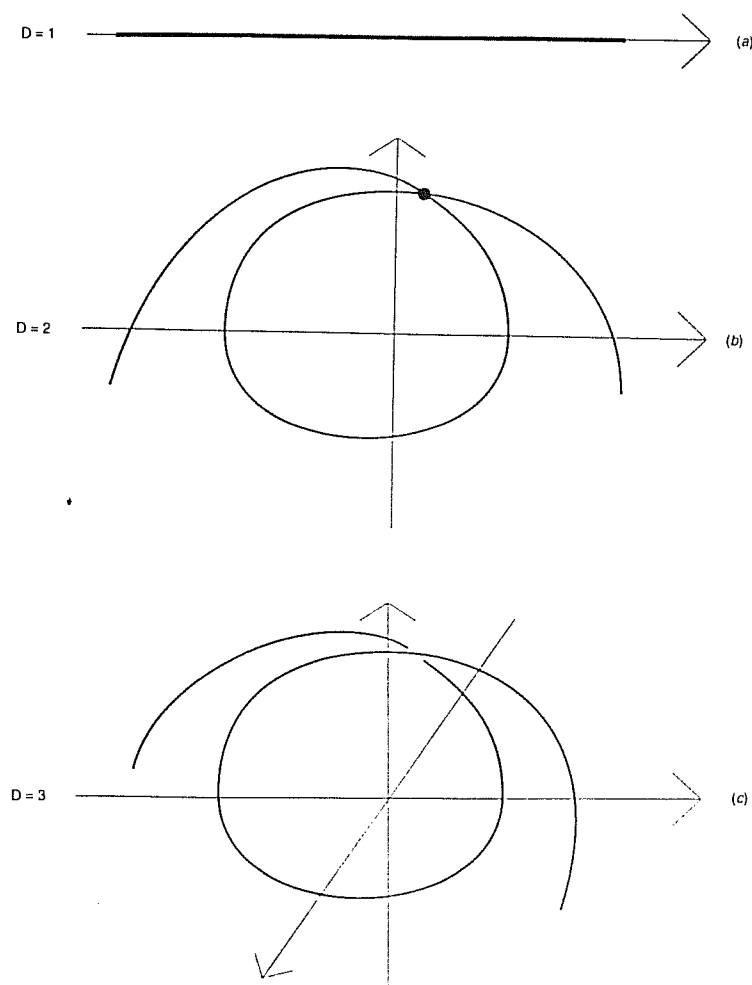

Fig. 6.8 Effect of embedding a hypothetical one-dimensional attractor in spaces of different dimensionality. In (a) and (b) spurious crossings occur, whereas the attractor is fully represented in the three-dimensional space (c).

in Chapter 5, is the typical algorithm used for reconstructed attractors. The dimension of the embedding space is now increased by 1; the attractor is then reconstructed in this new space and its dimension is recalculated. The process is continued until a limiting value of the correlation dimension d_G is reached.

Such an analysis, using data from the experimental pendulum, is shown in Figures 6.9 and 6.10. The graphs of $\log C$ versus $\log R$ appear to
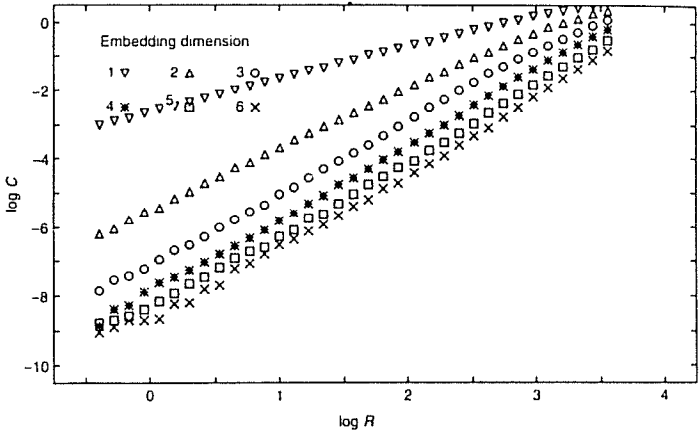
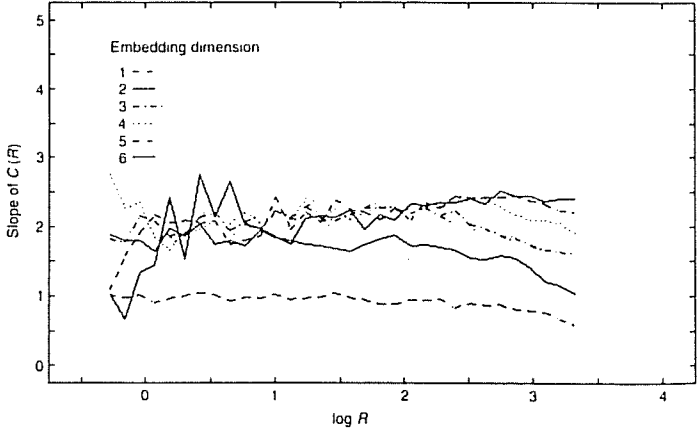Fig. 6.9 Correlation integral for embedding spaces of increasing dimension, from the experimental data.

Fig. 6.10 Slopes of correlation integral graphs for $1 \leq D \leq 6$. The middle sections of these graphs are approximately horizontal, showing that $d \approx 2.3$ over a modest range. The slope saturates when $D \geq 3$.

have linear scaling regions whose slopes increase with the embedding dimension until a limiting value is reached. However, to make a quantitative test, it is best to plot the slope, $d(\log C(R))/d(\log R)$; this quantity is shown in Figure 6.10 for embedding dimensions 1–6. The slope fluctuates wildly for small R due to the limited number of points in the time series from which $C(R)$ was computed. However, the slope does become flat and essentially independent of D for $-0.5 < \log R < 2.0$, provided D is at least 3. Over this range, the fitted slope (and its estimated statistical uncertainty) is given in Table 6.1 for each choice of

Table 6.1. *Saturation of apparent attractor dimension.*

Dimension of embedding space D	Apparent attractor dimension d
1	0.9 ± 0.1
2	1.8 ± 0.2
3	2.2 ± 0.2
4	2.3 ± 0.2
5	2.3 ± 0.2
6	2.2 ± 0.2

embedding dimension. It is apparent that the attractor dimension is approximately 2.3 under these conditions. However, choosing a different scaling range might change these values appreciably, so the quoted uncertainty, which is based on statistical fluctuations only, is probably an underestimate. Furthermore, the use of even moderately noisy experimental data can reduce the scaling range and therefore make the dimension estimate much less precise. (See Ding *et al.* (1993b).)

Nevertheless, the value obtained for the dimension of the attractor reconstructed from the experimental times series compares favorably with the value of $d = 2.35$ calculated from the pendulum simulation as given in Chapter 5 for the corresponding parameters. (Recall that the experimental pendulum configuration corresponds to a simulation configuration of $g = 1.5$, $q = 4$, and $\omega_D = \frac{2}{3}$.)

It is best not to use an embedding space of higher dimensionality than is required to produce a limiting value of d because the undesirable effects of experimental noise become more pronounced for embeddings of higher dimension. Another method that also reveals the minimum embedding dimension is the method of *false nearest neighbors*. (See Brown (1992) and Abarbanel *et al.* (1993).) This procedure is based on the fact that if the embedding space has too low a dimension D then a given vector may have a neighboring vector that is a false neighbor. That is, if these vectors were embedded in a phase space of dimension $D + 1$ they would no longer be close neighbors. Therefore one computes the change in distance between these vectors as the phase space dimensionality is increased from D to $D + 1$. If the change is greater than some appropriate constant then the neighbor is presumed false. When the dimension of the phase space is sufficiently large, the number of false neighbors drops abruptly. The dimension for which the transition occurs is the minimal embedding dimension.

Attractor reconstruction provides, in principle, a way to distinguish between stochastic noise and deterministic dynamics. Data dominated by stochastic noise will populate an embedding space of arbitrarily high dimension, so that the apparent attractor dimension will not reach a limiting value as D is increased. Of course, any experimental data contain measurement noise, which primarily affects the correlation function at small scales. Therefore, it is necessary to disregard those scales in examining $C(R)$ for scaling and in determining the attractor dimension.

6.1.6 Lyapunov exponents

The spectrum of Lyapunov exponents may also be calculated from the reconstructed attractor. Most of the methods for doing so are complex and beyond the scope of this book, especially when the deterministic data are contaminated with noise. (See, for example, Brown, Bryant, and Abarbanel (1991).) However, there is a straightforward algorithm for calculating the largest positive Lyapunov exponent. This exponent measures how quickly linear distances grow (Abarbanel *et al.*, 1993) and it places an upper limit on the prediction time for the system. Smaller positive exponents have less effect but shorten the prediction time, while negative exponents especially affect the approach of trajectories to the attractor during the initial transient stages of the motion. For the pendulum, there is only one positive Lyapunov exponent.

Wolf *et al.* (1985) give an algorithm for obtaining the largest Lyapunov exponent from a time series. (In principle, this method would also yield the smaller positive exponents, but limitations of the data size often preclude accurate calculation of these smaller exponents.) The approach is based upon following the divergence of a neighboring trajectory from a selected trajecory as shown in Figure 6.11. Over a time

Fig. 6.11 Illustration of the process by which time series data can be used to calculate the largest Lyapunov exponent.
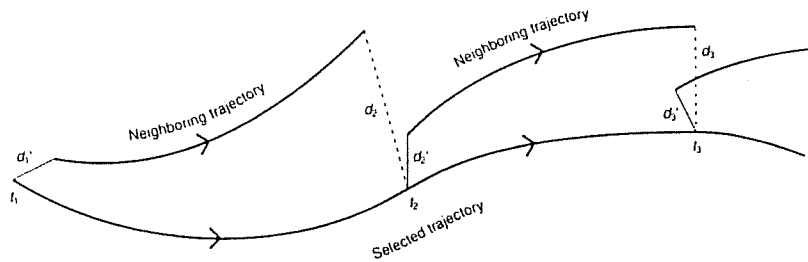
interval $t_2 - t_1$, the rate of divergence of two points that evolve from a spacing $\mathscr{L}_1'$ to a spacing $\mathscr{L}_2$ may be characterized by the quantity

$$\frac{\log_e(\mathscr{L}_2/\mathscr{L}_1')}{t_2 - t_1}. \tag{6.11}$$

Because the separation must be kept small compared with the size of the attractor, a new neighbor must be chosen periodically for subsequent estimates of the divergence rates. After n repetitions of stretching and renormalizing the spacing, the rates are weighted by the fraction of time between each renormalization and then added to yield an experimental value for the largest positive Lyapunov exponent:

$$\lambda_1 = \sum_{i=1}^{n-1}\left\{\left[\frac{(t_{i+1}-t_i)}{\sum_{i=1}^{n-1}(t_{i+1}-t_i)}\right]\left[\frac{\log_e(\mathscr{L}_{i+1}/\mathscr{L}_i')}{(t_{i+1}-t_i)}\right]\right\}. \tag{6.12}$$

Since $\sum_{i=1}^{n-1}(t_{i+1}-t_i)=t_n-t_1$ we have

$$\lambda_1 = \frac{\sum_{i=1}^{n-1}\log_e(\mathscr{L}_{i+1}/\mathscr{L}_i')}{t_n-t_1} \qquad \text{(Wolf } et\ al.,\ 1985\text{)}. \tag{6.13}$$

Figure 6.12 shows the result obtained from the experimental pendulum data for $\omega(t)$. (We used code from Wolf *et al.* (1985).) We find $\lambda_1 \approx 1.5\,\text{s}^{-1}$, but small fluctuations remain at $t = 200\,\text{s}$. In order to
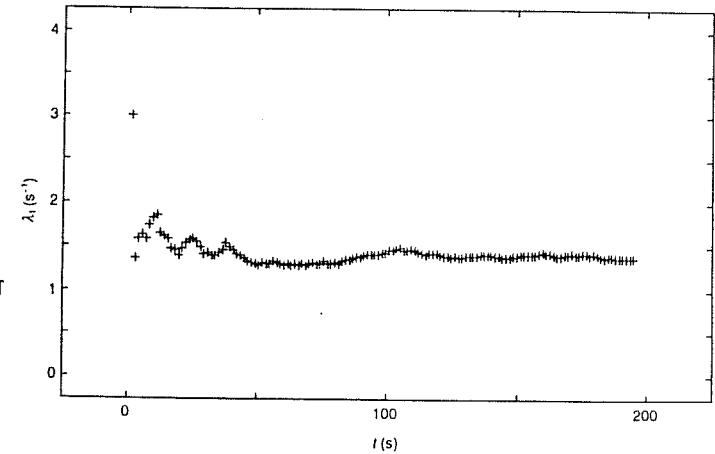
Fig. 6.12 Convergence of the largest Lyapunov exponent calculated from the experimental time series $\omega(t)$ with the algorithm of Wolf *et al.* (1985).

compare this result with that obtained numerically in Chapter 5 from equation (1.4), which is equivalent to the dimensionless model equation (6.2), we need to convert λ_1 to dimensionless form by dividing by $\omega_0 = 9\,\mathrm{s}^{-1}$:

$$\lambda_1/\omega_0 = 0.17.$$

This value is in good agreement with the result (0.16) shown in Table 5.1. Finally, using certain assumptions, it is possible to check the consistency of the measured λ_1 with the experimental value of the attractor dimension. (See Problem 6.5.)

6.1.7 Summary

In this section, we have described some methods for determining properties of chaotic attractors directly from measured time series. The values obtained for the attractor dimension and the positive Lyapunov exponent are in substantial agreement with those found from direct numerical simulation of the model equations. This agreement gives one confidence that the methods can also be applied in circumstances where suitable model equations (or appropriate parameters) are unknown, provided that the dynamics is low-dimensional.

These methods have a number of limitations. First, attractors cannot be completely described by only two or three measured invariant quantities. Therefore, measuring them does not fully characterize the dynamics or uniquely determine an appropriate model. Second, when applied to a chaos in a spatially extended system with many degrees of freedom, the embedding method can produce a severe underestimate of the dimension of the attractor (Lorenz, 1991). For example, local sampling of atmospheric data has been suggested to provide evidence for low-dimensional dynamics. However, a local variable in a strongly turbulent fluid is only slightly influenced by the dynamics even a short distance away, and hence cannot be used to reconstruct an attractor. In general, dynamical systems methods are ineffective when more than a few degrees of freedom are excited.

6.2 Prediction of chaotic states

A fundamental goal of scientific inquiry is the prediction of future states of a system. There are also many practical circumstances in which one might wish to forecast the future of a chaotic system that has been observed up to a given time. Whereas periodic systems are predictable indefinitely, the variables of a chaotic system are partially predictable only for a short time that is related to the positive Lyapunov exponents. (Stochastic systems also generally have a short correlation time during which partial prediction is possible.)

A variety of forecasting techniques have been developed for chaotic systems (Abarbanel *et al.*, 1993, Weigend and Gershenfeld, 1994). In this section we describe several methods and apply them to experimental data from the pendulum.

6.2.1 Method of analogues

Probably the simplest prediction technique for time series is the method of 'analogues' (Lorenz, 1963). The presumption is that short sequences of points in a chaotic time series are *approximately* repeated throughout the time series. In particular, the final short section of a given time series has a close relative earlier in the series. The continuation of this earlier section then provides a reasonable set of prediction points. For example, we could formulate a crude weather forecast by first looking at all previous local meteorological behavior, and then predicting tomorrow's weather as the continuation of a previous behavior that mimicked the weather of the past few days. While this method seems crude, it can (with some refinements) be used effectively if the dimension of the attractor is small and the time series contains a large number of points (Kostelich and Lathrop, 1994).

In general, prediction techniques take advantage of the geometry of the orbits on the reconstructed chaotic attractor. One approach combines the method of analogues with the attractor geometry. Recall that any trajectory on a chaotic attractor will generally exhibit near recurrences. We find the nearest earlier neighbor to the latest point on the attractor, and then use the sequence of points following this earlier neighbor as a basis for prediction. Figure 6.13 illustrates the method.

We apply this method to the pendulum using part of the experimental time series to form the attractor from which short-term predictions are made, and the remainder to test the quality of the predictions. The results are shown in Figure 6.14(a) and (b). In the first diagram prediction begins after 22 000 points in the time series and continues for another 700 points. Comparison with the experimental data shows that the forecast is remarkably good given the simplicity of the model. The
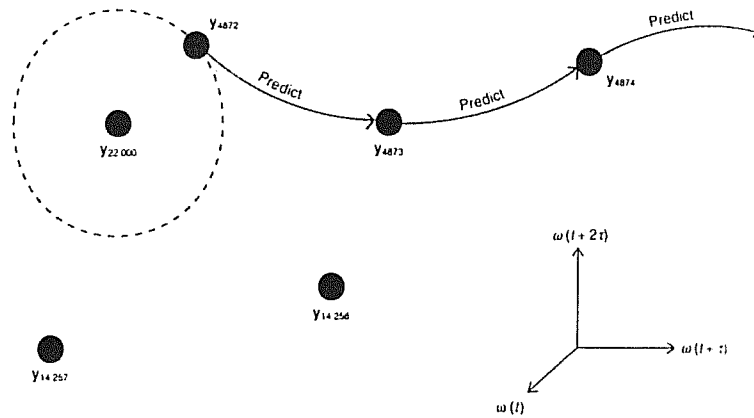
Fig. 6.13 Schematic diagram of the 'nearest neighbor' prediction algorithm.

next figure provides a quantitative measure of deviation of prediction from the experimental values. Following Farmer and Sidorowich (1987) we define an error function

$$E = \frac{\sigma_{pred}}{\sigma} \tag{6.14}$$

where $\sigma_{pred} = \langle [\omega_{pred}(t) - \omega_{obs}(t)]^2 \rangle_t^{\frac{1}{2}}$ is the cumulative root mean square deviation of predicted time series from the actual data. This error function is normalized by $\sigma = \langle [\omega(t) - \langle \omega(t) \rangle]^2 \rangle_t^{\frac{1}{2}}$, the cumulative standard deviation of the time series data. Therefore, as the prediction becomes meaningless, E approaches unity and $\log E$ approaches zero.

Figure 6.14(a) and (b) suggest that the prediction has meaning for 200 or 300 steps (several periods) in the times series. We can make a rough estimate of the maximum time over which any prediction can be expected to be valid. In Chapter 5 the approximate prediction time for a chaotic system was given as

$$T = (1/\lambda_+) \log_e(L/\varepsilon), \tag{6.15}$$

where L is the attractor size, ε is the uncertainty in initial coordinates and λ_+ is the positive Lyapunov exponent. The attractor size can be taken to be the length of a diagonal in the phase space, and the Lyapunov exponent is $1.5\,s^{-1}$. The estimation of ε is more problematic. One might interpret ε as the typical distance between a predicted point and the nearest neighbor used for the next prediction. On average this distance will roughly equal the mean separation between points on the attractor.
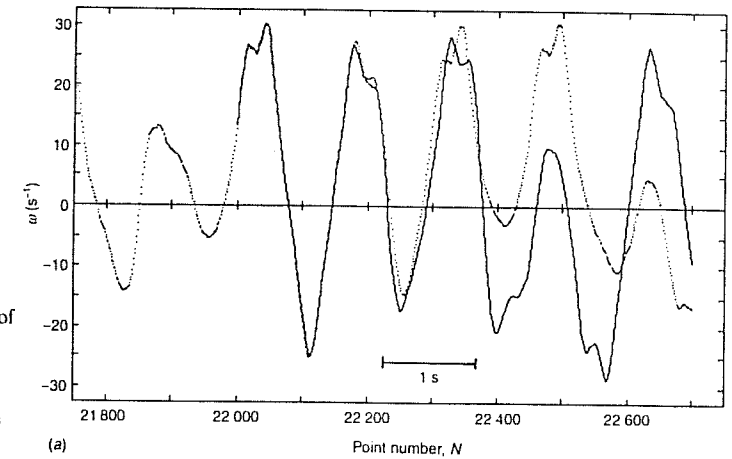

Fig. 6.14 Comparison of predicted and actual data values for the angular velocity of the experimental pendulum using the 'method of analogues' algorithm. In (a) the prediction is compared directly with the experimental points. The predicted values start at point number $N_0 = 22\,000$ and are indicated by the solid line. Experimental data are represented by the dotted line. In (b) the natural log of the error function is shown as a function of time or point number. The vertical lines show the variation due to different starting points in the region of N_0.

If N is the number of points, then for an attractor of dimension d each point occupies a 'volume' of size ε^d out of a total volume of $L^d = N\varepsilon^d$. Using these relationships we rewrite the prediction time as

$$T = (1/\lambda_+)(1/d)\log_e(N). \tag{6.16}$$

For $d = 2.3$ and $N = 22\,000$ points, the maximum prediction time is about 2.8 s, which corresponds to about 400 data points. The estimate is in reasonable agreement with the results shown in Figure 6.14.

6.2.2 Linear approximation method

A more satisfying approach to prediction is to develop an approximation for the map

$$\mathbf{y}_{j+1} = M(\mathbf{y}_j). \tag{6.17}$$

Many different mathematical forms may be used. Since the chaotic system is nonlinear one might attempt to fit the data with some combination of nonlinear functions. Undoubtedly, this set would be complex; instead we adopt a simpler approach. The behavior near each point of the attractor may be approximated by a unique *local* map. Then the evolution of points on the attractor is governed by the collection of all such local maps. Furthermore, since the points are close together the map is assumed to be a *linear* map at each point. That is,

$$\mathbf{y}_{j+1} = \mathbf{a}\ \mathbf{y}_j + \mathbf{b}, \tag{6.18}$$

where the matrix **a** and vector **b** are specific to each point. The aggregate of all these *local* linear maps forms a *global* nonlinear map (Abarbanel *et al.*, 1993). In the three-dimensional embedding space of the pendulum, the explicit form of this transformation is

$$\begin{pmatrix} \omega(t_{j+1}) \\ \omega(t_{j+1}+\tau) \\ \omega(t_{j+1}+2\tau) \end{pmatrix} = \begin{pmatrix} a_{11}\ a_{12}\ a_{13} \\ a_{21}\ a_{22}\ a_{23} \\ a_{31}\ a_{32}\ a_{33} \end{pmatrix} \begin{pmatrix} \omega(t_j) \\ \omega(t_j+\tau) \\ \omega(t_j+2\tau) \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}. \tag{6.19}$$

The components of **a** *and* **b** are determined at each point by a *least squares fit*. That is, the function

$$f(\mathbf{a},\mathbf{b}) = \sum \|\mathbf{y}_{j+1} - (\mathbf{a}\cdot\mathbf{y}_j + \mathbf{b})\| \tag{6.20}$$

is minimized, where the sum is taken over n nearest neighbors of the particular point from which the single step prediction is to be made. (See Problem 6.5.) Each $\mathbf{y}_j$ is a near neighbor of the particular phase point and the corresponding $\mathbf{y}_{j+1}$ is the successor of that same neighbor. Since a given set of **a** and **b** coefficients applies to a small region, these points must be close neighbors to the point from which the prediction is to be made. The least squares process yields three sets of four linear equations. (See Problem 6.5.) The unknowns, a_{i1}, a_{i2}, a_{i3}, and b_i are determined from equations of the form

$$\left.\begin{aligned}
a_{i1}\sum\omega_1\omega_1 + a_{i2}\sum\omega_1\omega_2 + a_{i3}\sum\omega_1\omega_3 + b_i\sum\omega_1 &= \sum\omega_i'\omega_1 \\
a_{i1}\sum\omega_2\omega_1 + a_{i2}\sum\omega_2\omega_2 + a_{i3}\sum\omega_2\omega_3 + b_i\sum\omega_2 &= \sum\omega_i'\omega_2 \\
a_{i1}\sum\omega_3\omega_1 + a_{i2}\sum\omega_3\omega_2 + a_{i3}\sum\omega_3\omega_3 + b_i\sum\omega_3 &= \sum\omega_i'\omega_3 \\
a_{i1}\sum\omega_1 \quad + a_{i2}\sum\omega_2 \quad + a_{i3}\sum\omega_3 \quad + b_i\cdot n &= \sum\omega_i',
\end{aligned}\right\} \tag{6.21}$$

where $i = \{1,2,3\}$, $\sum = \sum\limits_{k=1}^{n}$ is a sum over n nearest neighbors, $\omega_m = \omega(t_j + (m-1)\tau)$, and $\omega_m' = \omega(t_{j+1} + (m-1)\tau)$. In a refinement of this process, the terms in the sums that are associated with specific points are weighted by a Gaussian factor that depends on the distance from the point in question to a particular neighbor (Casdagli *et al.*, 1992). If that distance is d_k, the weighting factor is $e^{-d_k/\sigma}$ where σ is an adjustable cutoff length beyond which neighbors do not contribute appreciably to the summations. Gaussian weighting does improve the prediction.

Let us summarize the procedure.

(1) Start with the last known point in phase space.
(2) Search for its n nearest neighbors in phase space.
(3) Calculate the sums of coordinates of these points and their corresponding next points as described in the last paragraph.
(4) Form the linear equations (6.21).
(5) Solve the equations for the local a and b coefficients. (A Gauss–Seidel iterative method works well. See Press *et al.* (1986).)
(6) With the coefficients as components of the linear transformation, apply the linear map to the last point to obtain a predicted point.
(7) Repeat the process with the predicted point at the new starting point.

Figure 6.15 shows the resulting prediction for $\omega(t)$ with a weighting of $\sigma = 2$ and $n = 12$ neighbors.

It is interesting to compare the quality of this prediction with that given by the method of analogues. Although work on some other systems by Farmer and Sidorowich (1987) shows that analogues are less effective for short-term prediction, that approach seems surprisingly to

Fig. 6.15 Comparison of predicted and actual data values for $\omega(t)$ (experimental) using the 'least squares' algorithm. The prediction is compared directly with the experimental points. The predicted values start at $N_0 = 22\,000$ and are indicated by the solid line. The experimental data are shown by the dotted line.
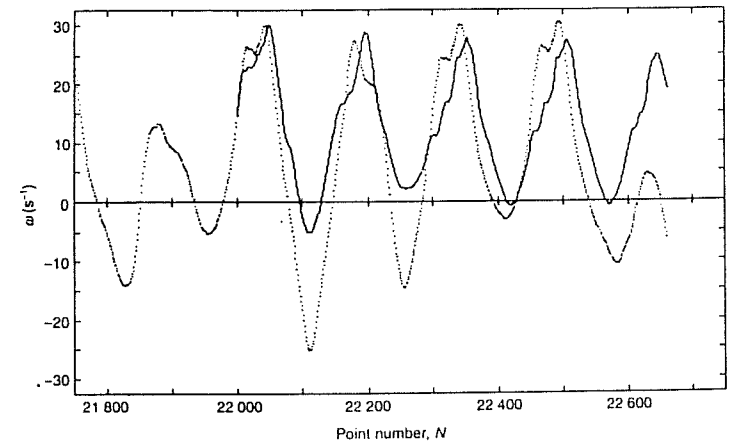
be better here. How can we understand this result? First, the result shown in Figure 6.14(a) is one of the better predictions from an ensemble of predictions done in the neighborhood of $N_0 = 22\,000$. The error bars in Figure 6.14(b) indicate that the quality of these predictions depends heavily on where the prediction begins. Second, the method of analogues naturally incorporates the nonlinearity of the system, whereas the least squares method approximates the local nonlinear dynamics by a linear map. If the actual dynamics contains strong second derivative terms in the region of prediction, then the linear mapping may be less accurate. Finally, averaging caused by use of the large numbers of neighbors required by the least squares method may result in some further insensitivity to complex dynamics.

On the other hand, the least squares method provides a basis for more sophisticated approaches. For example, least squares models can incorporate nonlinear terms in the map, an extension that is consistent with the spirit of nonlinear dynamics. Furthermore, since least squares methods are optimization techniques, one can add constraints involving invariant quantities. These quantities may include the Lyapunov exponents and the spatial density of points on the attractor. That is, the time series predictions can be constrained to be consistent with Lyapunov exponents and densities that were calculated from the experimental data (Abarbanel, Brown, and Kadtke, 1990). Of course, increasing sophistication brings increasing complexity and greatly increased computation time. Nevertheless, even the crude predictions shown here are an example of what can be achieved for short-term forecasting.

Beyond these refinements there exists a variety of other forecasting methods. For example, efforts have been made to train neural networks – nonlinear devices that simulate neural input/output mechanisms – to predict chaotic states (Lapedes and Farber, 1987). In an attempt to bring together and evaluate these methods, Weigend and Gershenfeld (1994) proposed and organized a contest to predict chaotic states of data taken from such diverse sources as financial markets, fluctuations of a far-infrared laser, time variations of a white dwarf star, and even a Bach fugue. One conclusion that may be derived from the many responses to this contest is that superior short-term prediction is not automatic – computer algorithms must be used with great care and with attention to the nature of the data. The researcher's insight into the nature of the system remains an essential ingredient of the forecasting process. While these developments are quite exciting, accurate prediction is fundamentally limited in time by a chaotic system's sensitivity to initial data as manifested by the positive Lyapunov exponents.

6.3   Modification of chaotic states

Chaotic states are inherently unstable and unpredictable; therefore it is sometimes desirable to avoid them. Situations of this type occur, for example, in the design of elastic mechanical structures and electronic circuits. The simplest way to avoid chaos is to adjust the parameters of a system to a regime of periodic rather than chaotic oscillation. However, suitable values of the control parameters may not be known, or the power to change them sufficiently may be absent. One example of a practical attempt at control of chaotic states is the use of pacemakers to control cardiac arythmias, some of which are believed to be chaotic oscillations.

An alternative approach is to modify the dynamics in such a way that an unstable periodic orbit becomes stable, so that the motion (if it is in the basin of attraction of this periodic orbit) becomes periodic. Such stabilization is often given the apparently contradictory name 'control of chaos'. For the pendulum, the resulting periodic motion might occur at the forcing frequency or one of its subharmonics.

Are there ways to stabilize a chaotic system without drastically changing its dynamics? One simple method is based on a small periodic perturbation at the frequency of an unstable periodic orbit. (See Braiman and Goldhirsch (1991).) An example of this method, applied to the simulated chaotic pendulum, is given in Figure 6.16. Here a chaotic state is converted to a periodic oscillation by adding a secondary forcing at frequency $\omega_D/2$ to the basic forcing at ω_D. However, the required
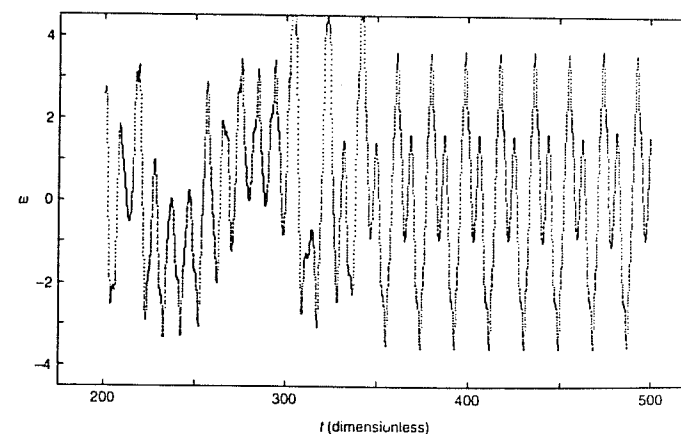

Fig. 6.16 Time series of a simulated pendulum that is perturbed by a period-2 sinusoidal forcing term. The period-2 motion becomes stable when the perturbation is about 67% of the original forcing amplitude. The control forcing was initiated at $t = 300$.

amplitude of the secondary forcing is fairly substantial; in this case the ratio of the secondary amplitude to the primary forcing is about $\frac{2}{3}$. Partial control – a reduction in the amount of chaos – may be achieved with smaller secondary forcing and may, in some contexts, be useful. For example, partial control may result in a narrowing of the Fourier spectrum of an experimental time series.

Ott, Grebogi, and Yorke (OGY 1990) proposed an elegant scheme to stabilize an unstable periodic orbit. Their idea was based on the use of feedback to induce small modifications of the control parameter. They applied the idea to simulations of the Hénon map, and stabilized both period-1 and period-2 orbits. The OGY method was subsequently demonstrated in a variety of experiments. Ditto, Rauseo, and Spano (1990) studied the buckling in a magnetic field of a thin ribbon of magnetoelastic material that is attached at its base and free to vibrate elsewhere. Since its elastic constants are affected by magnetic fields, the ribbon can be driven by application of a time-varying field. Changes in the amplitude of the field oscillation cause typical transitions from periodic to chaotic motion. A small dc magnetic field, whose size is about 5% of the amplitude of the ac field, is also present. Control of chaos – stabilization of an unstable periodic orbit – is achieved with very small changes (less than 8%) in the static magnetic field, made once during each ac cycle. This small feedback allows the chaotic motion to be stabilized indefinitely. More recently, Starrett and Tagg (1995) have stabilized an experimental parametric pendulum – a pendulum driven by periodic vertical motion of its pivot. Control is achieved through adjustment of the eddy current damping of the pendulum with proportional feedback, once each forcing cycle. The OGY method has also been modified to apply when the attractor is reconstructed from a time series (Nitsche and Dressler, 1992) and was used successfully to control a simulated Duffing oscillator. However, this last technique is beyond the scope of the present discussion.

Our primary example, the simulated pendulum, may also be controlled with the OGY method (Baker, 1995). (We summarize the main steps here and refer the reader to the previous reference for details.) Following OGY we consider a map, in this case the two-dimensional map represented by a Poincaré section of the pendulum:

$$\begin{pmatrix} \theta_{n+1} \\ \omega_{n+1} \end{pmatrix} = M(\theta,\omega) \begin{pmatrix} \theta_n \\ \omega_n \end{pmatrix},$$

(6.22)

where M is unknown and dependent on position in the phase plane.

Fixed points of this map may be found by determining the distance between consecutive iterates. If this distance is small then the given point may be a fixed point of an unstable periodic orbit. For period-2 fixed points, second iterates are compared, and so forth for higher order periodic orbits. An important property of unstable fixed points is that they are saddle points: there are both attracting and repelling directions. These directions lie along the stable and unstable manifolds, respectively. (See Chapter 4.)

Once the fixed point is chosen, M is determined near the fixed point. Although the Poincaré section is a nonlinear map, a linear approximation is found by observing the evolution of a few points in the vicinity of the fixed point with coordinates $\theta_F = 1.523$ and $\omega_F = -0.415$. (See Figure 6.17.) Vectors may be formed from pairs of these points, and the evolution of these vectors determines the elements of M. Finally, the eigenvalue problem associated with M may be solved. The eigenvectors e_s and e_u give the directions of contraction and expansion – the stable and unstable manifolds – near the fixed point. Similarly the eigenvalues L_s and L_u give the rates of contraction and expansion near the fixed point. We also determine vectors f_s and f_u that are perpendicular to e_u and e_s, respectively. All these quantities characterize the geometry near the fixed point.

Once the local geometry is determined, one of the pendulum's parameters, for example the damping constant q, is chosen as the control parameter. The effect on the fixed point coordinates of varying q is determined by simulation with q slightly displaced from its value q_0 at

Fig. 6.17 Evolution of a small patch near a fixed point of the forced pendulum. Points at the corner of a rectangle in the phase plane evolve to new locations (with corresponding symbols) after one cycle. Vectors may be formed by joining pairs of points at the vertices of the original rectangle; for example, a vector from + to O and a vector from + to ∗. (See text.) The solid square is located at the unstable fixed point.
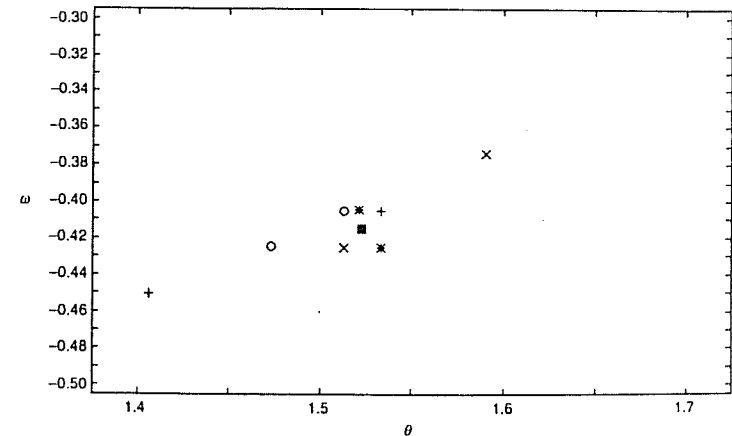
the fixed point. This variation is linearized near q_0 and described by the vector $(\partial\theta_F/\partial q, \partial\omega_F/\partial q)$.

The control algorithm uses slight changes δq in the control parameter to force the orbit toward the direction of contraction (the stable manifold) near the fixed point. The natural contraction then drives the orbit toward the fixed point. However, a small error will cause the orbit to miss the fixed point, so the adjustment process must be carried out repetitively to achieve stabilization (constant values of the phase plane coordinates). The required parameter change is given by the formula

$$\delta q = \frac{L_u}{L_u - 1} \frac{(f_{u1}, f_{u2}) \begin{pmatrix} \Delta\theta_n \\ \Delta\omega_n \end{pmatrix}}{(f_{u1}, f_{u2}) \begin{pmatrix} \partial\theta/\partial q \\ \partial\omega/\partial q \end{pmatrix}}, \tag{6.23}$$

where (f_{u1}, f_{u2}) is a vector that is perpendicular to the stable manifold and $(\Delta\theta_n, \Delta\omega_n)$ is the vector from the fixed point to the nth iterate. The change δq is reevaluated at each cycle using the current value of the latter vector.

A series of successive iterates showing control of a simulated pendulum is given in Figure 6.18. The control algorithm is turned on at $t = 1000$ and turned off at $t = 3000$. In fact, control is not immediately achieved at its activation. The control mechanism is only operable when the iterate happens to approach the fixed point. Several cycles may be required for a sufficiently close approach to occur. Moreover, the necessary closeness is determined by the maximum possible variation δq_{max} in the control

Fig. 6.18 Stabilization of a chaotic pendulum as indicated by the angular velocity at the start of each forcing cycle. The control mechanism is initiated at $t = 1000$ and terminated at $t = 3000$. During this interval the constant (cyclic) value of the angular velocity indicates the stabilization of period-1 motion. The maximum relative change in q was 4%.
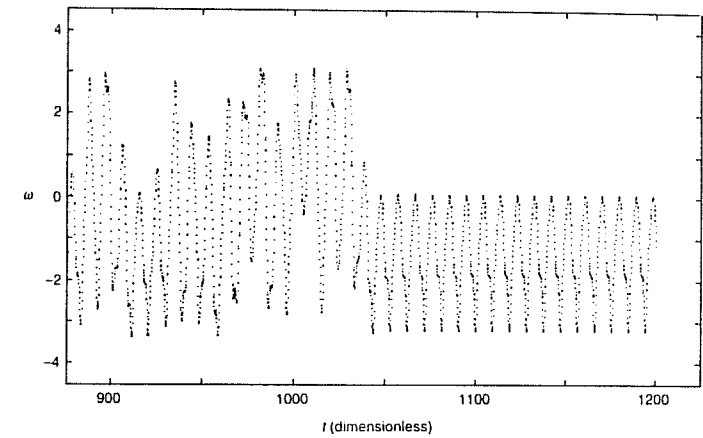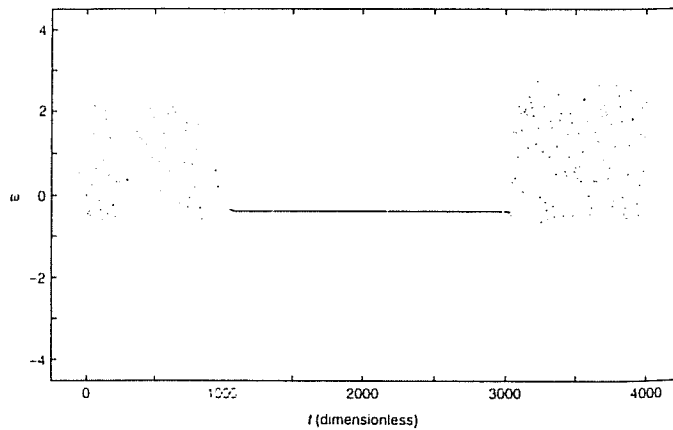

Fig. 6.19 Time series of the angular velocity during the initiation of control at $t = 1000$. Note that several cycles are required to stabilize the period-1 orbit.

parameter. Ott *et al.* (1990) have developed a relationship between δq_{max} and the average time to establish control. The delay is demonstrated in Figure 6.19, where the entire velocity time series in displayed during the initiation of control. Control lags initiation by about four forcing cycles. The techniques we have used to control the motion of the simulated pendulum are similar to those used by various investigators to control experimental systems.

We have chosen to demonstrate the basic OGY method. There are, however, other control schemes that are especially suited to nonlinear dynamics. In this connection we mention the work of Bradley (1992) on the simulated pendulum and that of Hubinger, Doerner, and Martienssen (1993) on the experimental pendulum. For an entry into the literature of this fascinating subject see, for example, Abarbanel *et al.*, 1993.

6.4 Conclusion

This chapter has introduced techniques that are commonly used for analyzing chaotic experimental data, predicting the future behavior for a limited period of time, and modifying chaotic states. These methods are often applicable even when a suitable model is unknown. In many areas of science, it has been presumed that understanding leads to prediction and control. The early development of nonlinear dynamics, on the other hand, suggested that nonlinear dynamics might be different. However,

progress in prediction and control for chaotic systems shows that such a pessimistic view is at least partially incorrect.

Problems

6.1 Use the transformation $t' = t\omega_0$ to show the equivalence of

$$I\frac{d^2\theta}{dt^2} + b\frac{d\theta}{dt} + \omega_0{}^2 I\sin\theta = T\sin\omega_f t$$

and

$$\frac{d^2\theta}{dt'^2} + \frac{1}{q}\frac{d\theta}{dt'} + \sin\theta = g\,\sin\omega_D t'.$$

(Hint: Remember to transform the derivatives.) Using intermediate results from the above exercise give values for g, q, and ω_D in terms of I, b, ω_0, T, and ω_f.

6.2 Consider a time series of experimental data $\{\omega_1, \omega_2, \cdots, \omega_n\}$, each point being separated by 7 ms. Let $\tau = 21$ ms and suppose you wish to reconstruct points in a three-dimensional phase space. Write out the coordinates of the first four points in the phase space. Using the transformation $t' = t\omega_0$ with $\omega_0 = 9\,\mathrm{s}^{-1}$, determine the value of τ' for a pendulum represented by the second, dimensionless differential equation in Problem 6.1.

6.3 Using the least squares method, derive equations (6.21) for prediction as outlined in Section 6.2.2.

6.4 For the experimental data used in this chapter the positive Lyapunov exponent is $\lambda_+ = 1.5\,\mathrm{s}^{-1}$. *Assume* that the second Lyapunov exponent is zero and calculate the third (negative) exponent from the relationship

$$\nabla \cdot F = -b/I = \sum_{i=1}^{3} \lambda_i,$$

where $b/I = 2.25\,\mathrm{s}^{-1}$. With these values for the Lyapunov exponents use the Kaplan–Yorke relation to estimate the fractal dimension of the attractor. Compare your value with that given in the text.

6.5 Derive the expression $T = (1/\lambda_+ d)\log_e(N)$ that is used as an estimate of the prediction time.

Problems 6.6, 6.7, and 6.8 require use of the data and programs found on the optional CHAOS II Diskette.

6.6 Use the program RECDIM and the data file VEL_FILE.TRU to reconstruct the attractor and calculate dimension in an embedding space of three dimensions. Use a delay (τ) of 30 data points. Comment on any differences from the results in the text, where the delay is 20 data points.

6.7 Use the program RECLYAP and the data file VEL_FILE.TRU to calculate the positive experimental Lyapunov exponent with a delay of 30 data points. Using your answer to Problem 6.6 and the Kaplan–Yorke relation, check the consistency of the Lyapunov exponent and the dimension results.

6.8 Use the program PREDICT1 with various segments of the data on the file VEL_FILE.TRU to do prediction at different epochs in the time series. Is the prediction time approximately constant for these segments. Does the number of data points 'known' prior to the beginning of prediction have an effect upon the prediction time?

6.9 Modify the program POINCARE so that you can determine the fixed points of a pendulum simulation. Use the bifurcation diagrams in the text to find a parameter set in a chaotic region and then determine some of the fixed points. Then vary one of the parameters and determine the effect on the fixed point. Next, by observing the evolution of Poincaré points near the fixed point through one forcing cycle, determine a linear map for the Poincaré section. Calculate the eigenvectors and eigenvalues of M, together with the corresponding vectors f_s and f_u. (Note that the normalization of these vectors is arbitrary.) Finally, use the program CONTROL, or a modification of it, to attempt control of your simulation.