Chapter 1

Fundamentals

This chapter will describe the fundamental concepts of continuous-time chaotic systems so as to make the book self-contained and readable by someone unfamiliar with the subject using the damped, forced pendulum as the main example. It will also define the sense in which a chaotic system is deemed to be elegant since that term is rather subjective with no universal definition. Most of the material in this chapter is covered in more detail in Sprott (2003).

1.1 Dynamical Systems

A *dynamical system* is one whose state changes in time. If the changes are determined by specific rules, rather than being random, we say that the system is *deterministic*; otherwise it is *stochastic*. The changes can occur at discrete time steps or continuously. This book will be concerned with continuous-time, deterministic, dynamical systems since they arguably best approximate the real world. This view represents the prejudice of most physical scientists, but it is also the case that chaos is relatively too easy to achieve in discrete-time systems, and hence it is less of a challenge to find elegant examples of chaos in such systems, and those that are found have less apparent relevance to the natural world. Also, discrete-time systems have already been extensively explored, in part because they are more computationally tractable.

Stochastic systems mimic many of the features of chaos, but they are not chaotic because chaos is a property of deterministic systems. Furthermore, introducing randomness into a dynamical model is a way of admitting ignorance of the underlying process and obtaining plausible behavior without a deep understanding of its cause. For these reasons, stochastic systems will not be discussed in this book.

The most obvious examples of dynamical systems are those that involve something moving through space, like a planet orbiting the Sun, a pendulum swinging back and forth, or an animal exploring its habitat. But dynamical systems can be more abstract, such as money flowing through the economy, information propagating across the Internet, or disease spreading through the population. In each of these examples, we characterize the state of the system at each instant by a set of values of the time-changing *variables* that collectively define not only the current conditions but that uniquely determine what will happen in the future.

For a planet orbiting the Sun, six variables are required, three components to describe its position relative to the Sun at each instant, and three to describe its velocity in the three-dimensional space in which it moves. Newton's second law ($F = ma$) coupled with the universal law of gravitation ($F = GmM/R^2$) provides the deterministic rule whereby its future state is completely and uniquely determined.

Calculation of celestial motion is one of earliest problems solved by scientists, and the calculations are some the most precise in all of science, allowing, for example, the prediction of Solar eclipses many years in advance both in time and space. Not only can one predict when such an eclipse will occur, but where on the Earth to stand to get the best view of it. Contrast this strong predictability with the difficulty of predicting even a few days in advance whether the sky will be clear enough to observe that eclipse using extremely detailed weather models and vast computational resources. Therein lies the difference between regular and chaotic dynamics.

For the simpler example of a pendulum swinging in a plane (Baker and Gollub, 1996; Baker and Blackburn, 2005), only two variables are required, one to specify its angle with respect to the vertical, and a second to specify its velocity. Such a system has one spatial degree of freedom (the arc of a circle), in contrast to the planet orbiting the Sun, which has three degrees of freedom, and this one degree of freedom is more than sufficient to produce chaos under some conditions. Hence it is a simpler and more elegant example with which to begin.

1.2 State Space

Consider a 1-kilogram mass on the end of a 1-meter-long, massless, rigid rod (so as to allow oscillations of arbitrarily large amplitude, including ones where the pendulum goes 'over the top') as shown in Fig. 1.1, and located on

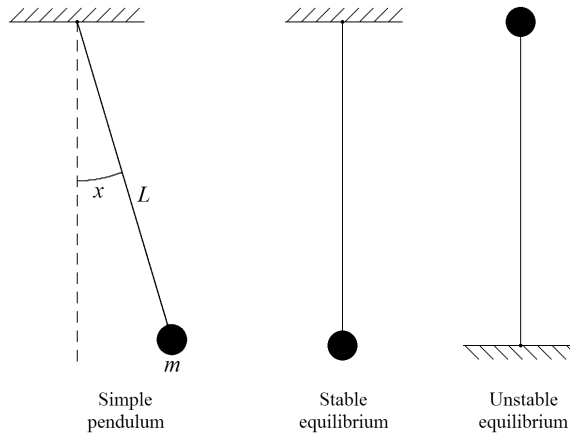

Fig. 1.1 A pendulum consisting of a mass m on the end of massless, rigid rod of length L at an angle (in radians) of x. The quantities m and L are fixed parameters of the system, and x is a time-dependent variable with $x = 0$ corresponding to a stable equilibrium and $x = \pi$ corresponding to an unstable equilibrium.

a distant planet where the acceleration due to gravity is 1 meter/second^2 (about one tenth of Earth's gravity). These particular values make the resulting equations more elegant in a sense to be described later and adequately serve to illustrate the ideas, which are very general. Of course such an idealized pendulum could never be built, not only because massless rods are unavailable, but because any rod with a mass much less than 1 kilogram would not likely have sufficient strength to withstand the buckling forces when the pendulum is inverted, but these practical limitations need not constrain our imagination or deter the discussion.

The mass, length, and acceleration due to gravity are considered *parameters* of the system that can be adjusted but that do not change during the time the motion is being examined. A major theme of this book is to find values of the parameters that make the equations simple but that allow the system to behave chaotically. In the equations that model the system, the parameters will usually be denoted by letters near the beginning of the alphabet $(a, A, b, \ldots)$ or sometimes by Greek letters $(\alpha, \beta, \gamma, \Omega, \sigma, \ldots)$, whereas the variables will be denoted by letters near the end of the alphabet (t, u, v, w, x, y, z). For a dynamical system, the time (here always denoted by t) is the *independent variable* upon which the other *dependent variables* depend.

The distinction between parameters and variables is usually clear in a mathematical model, but is much less so in a real dynamical system. In a radio receiver, the parameters might be the position of the frequency and volume knobs, whereas the variables might be the position and velocity of the cone of the loudspeaker as it vibrates in response to the audio signal that is received for particular settings of the knobs. In a climate model, a parameter might be the concentration of carbon dioxide in the atmosphere, or perhaps its rate of change, whereas a variable might be the globally averaged temperature. However, in a more complicated climate model, the carbon dioxide might itself be a variable determined by the model.

Think of the parameters as time-independent inputs to the model that are specified by factors not included in the model and the variables as time-dependent outputs that are determined by the equations that specify the model. In the real world, almost everything is a variable expect perhaps a few fundamental constants such as the speed of light and Planck's constant, but the systems in this book are gross simplifications of nature in which only a few quantities are allowed to vary while others are held rigidly constant. The elegance comes from showing that interesting and realistic behavior can result from such extremely simple models.

In the pendulum swinging in a plane, one time-dependent variable is the angle x (in radians) that the pendulum makes from the vertical. At each instant, there is a force equal to $-\sin x$ pulling the mass back to its stable equilibrium position at $x = 0$. The virtue of using the parameters $m = L = g = 1$ is that the coefficient multiplying $\sin x$, which would normally be g/L, is eliminated. Newton's second law then leads to a system of equations for the motion given by

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= -\sin x, \end{aligned} \tag{1.1}$$

where v is the angular velocity and the overdot denotes a time derivative ($\dot{x} = dx/dt$). (Isaac Newton used this notation when he invented calculus, whereas his rival, Gottfried Leibniz, used the more familiar but more clumsy notation.) This is the sense in which a mechanical system with one degree of freedom has a *state space* of two dimensions (x and v). A state space such as this, in which one variable is proportional to the time derivative of another ($v = dx/dt$), is usually called *phase space*, but we will keep the more general term throughout this book since the state space variables are not always related in this way.

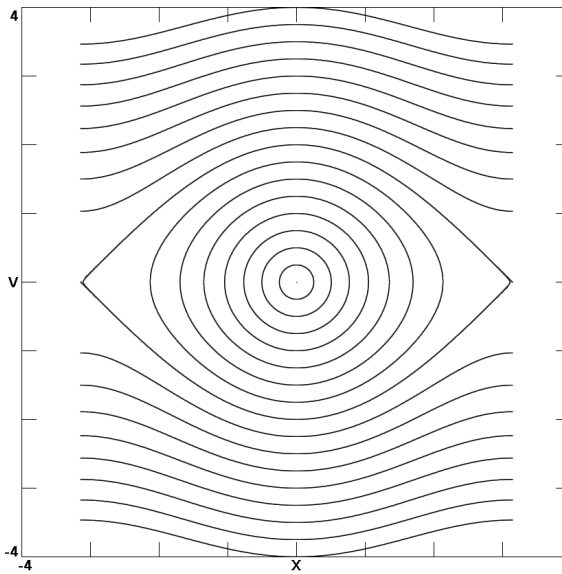

Fig. 1.2 State space plot for the frictionless pendulum from Eq. (1.1) with many different initial conditions showing representative streamlines of the flow. The closed curves near the origin represent librations, and the open lines at large values of $|v|$ represent rotations.

Equations (1.1) imply that at every point in this abstract two-dimensional state space of x and v, there is a unique direction and amplitude of the motion given by the vector whose components are $\dot{x}$ and $\dot{v}$. The simultaneous motion of all the points in this space resembles a flowing fluid, and hence systems such as Eq. (1.1) are usually called *flows*. Figure 1.2 shows some of the streamlines for this particular flow. The streamlines are simply the clockwise trajectories followed by small particles moving with the flow for various initial conditions. In this case, each streamline is a curve of constant total energy (kinetic + potential) given by $E = v^2/2 + 1 - \cos x$, with the different curves corresponding to different values of E. Of course, there is a streamline through every point in the state space, only a representative sample of which is shown in the figure.

There are several things to note about Fig. 1.2. There are two kinds of streamlines — those that encircle the origin at $x = v = 0$, and those that bump along at large positive or negative velocities. The former represent back and forth oscillations of the pendulum (so-called *librations*), and the

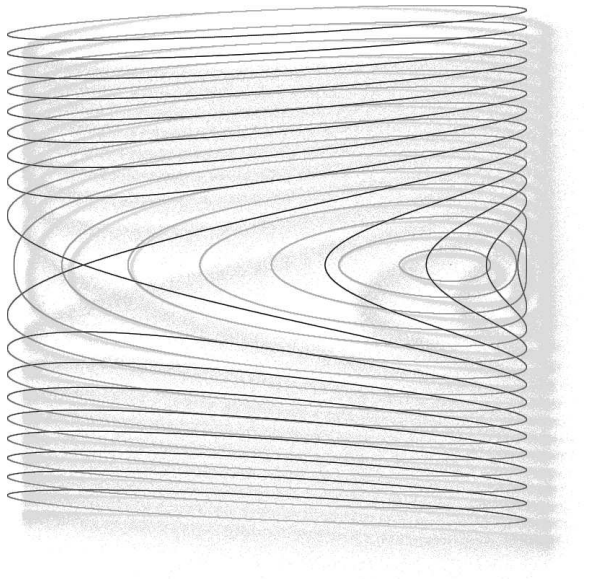

Fig. 1.3 State space plot for the frictionless pendulum from Eq. (1.1) with many different initial conditions showing representative streamlines of the flow as in Fig. 1.2 but wrapped onto a cylinder and then projected back onto the plane of the page.

latter represent *rotations* (sometimes called 'hindered rotations' since gravity affects the rotational velocity) in which the pendulum goes 'over the top.' For rotational motion, the angle continually increases or decreases, and the streamlines are plotted mod 2π so that when they reach $+\pi$ they wrap around and continue at $-\pi$ but going in the same direction, and similarly when they reach $-\pi$. Curiously, the cases are counterintuitive since the streamlines that are closed loops correspond to back-and-forth oscillations, while those that appear to end at $+\pi$ and $-\pi$ are rotations.

If the streamlines were drawn on the surface of a cylinder of unit radius and circumference 2π, they would automatically wrap around in the correct way as Fig. 1.3 shows. Of course, the cylinder is projected back onto the surface of the page, and so a bit of imagination is required to see how the trajectories wrap around it. The visualization is aided by making the lines darkest where they are closest to the viewer and by showing a subtle shadow as if the cylinder were illuminated by a light coming from the top left of the figure.

The two types of streamlines are separated by a line called the *separatrix* that appears to have a sharp bend in Fig. 1.2. That bend is called an

X-point since it appears to cross itself when drawn on the surface of a cylinder as in Fig. 1.3, and it corresponds to the pendulum swinging with just enough energy ($E = 2$) to reach the inverted position with zero velocity. In such a case, it would stall and remain there forever. The X-point and the *O-point* at the origin (also called a *center*) are equilibrium points for the pendulum, with the former being unstable since a slight perturbation will grow, while the latter is stable since a small perturbation will simply cause a small rotation around the center. These two equilibrium conditions are shown in Fig. 1.1.

The X-point is also called a *saddle point* since, like the saddle on a horse, there are two directions (the *stable manifold*) from which a ball rolling downhill will approach the equilibrium and two other directions (the *unstable manifold*) toward which a ball released from near the equilibrium point will roll downhill. Potato chips also sometimes have this shape. The analogy is not perfect, however, since the velocity of the flow goes to zero at the equilibrium point, whereas a ball rolling along the stable manifold toward the equilibrium point of a real saddle will shoot right past it without slowing down. Saddle points also exist in higher-dimensional state spaces where the manifolds have a correspondingly higher dimension and are thus harder to visualize.

1.3 Dissipation

In the previous example, there was no friction, and hence the mechanical energy is conserved and the pendulum swings or rotates forever. The streamlines on a cylinder are closed curves. Such systems are said to be *conservative* since they conserve energy. According to Liouville's theorem (Lichtenberg and Lieberman, 1992), they also have the property that the area occupied by a cluster of initial conditions remains constant in time, implying that the flow is incompressible, much like a swirling liquid. The flow is also time-reversible since changing the sign of t corresponds to the transformation $v \to -v$, which flips the figure vertically and preserves the shape of the streamlines. Strictly conservative systems are rare in nature, with the standard examples being the motion of astronomical bodies in the near vacuum of outer space and the nonrelativistic motion of individual charged particles in magnetic fields where electromagnetic radiation can be neglected.

For the pendulum, it is more realistic to include a friction term in the

equations of motion, due primarily to air resistance (or whatever gas exists in the atmosphere of the distant planet). This friction term is typically assumed to be proportional to the velocity, which is reasonable approximation if the velocity is not too large, and directed opposite to it, so that the equations of motion become

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= -bv - \sin x, \end{aligned} \tag{1.2}$$

where b is a measure of the friction. The quantity b is a parameter that is assumed to have a constant value as the motion continues, but that can be changed to produce different types of motion. More generally, such a term is called *damping*, and it can be a nonlinear function of v and can also depend on x.

Systems with damping are said to be *dissipative*, in contrast to the conservative system previously discussed. Dissipative systems do not conserve mechanical energy and are not time-reversible. Their streamlines are not contours of constant energy. They are described by a *compressible* flow as shown in Fig. 1.4 for a case with a small damping of $b = 0.05$ and an initial condition of $(x_0, v_0) = (0, 4)$. In this case, the pendulum rotates eight times and then swings back and forth with a decreasing amplitude, approaching ever closer to the stable equilibrium point at $(x, v) = (0, 0)$. This point acts as an *attractor* for all initial conditions since they are drawn to it in the limit of $t \to \infty$, and it is called a *sink* for obvious reasons. Imagine the fluid swirling as it goes down the drain. Hence instead of just saying 'the pendulum slows down,' we can now say, in the pompous language of dynamical systems, that 'all initial conditions in state space are attracted to the stable equilibrium at the origin.'

1.4 Limit Cycles

The equilibrium at the origin in the previous example is a point attractor for $b > 0$, but other types of attractors also exist. Consider a situation in which the damping is negative ($b < 0$) near the origin but then becomes positive when the trajectory gets too far from the origin. In such a case, the trajectory is drawn to the region where the two effects offset. A system with this property is

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= (1 - x^2)v - \sin x, \end{aligned} \tag{1.3}$$

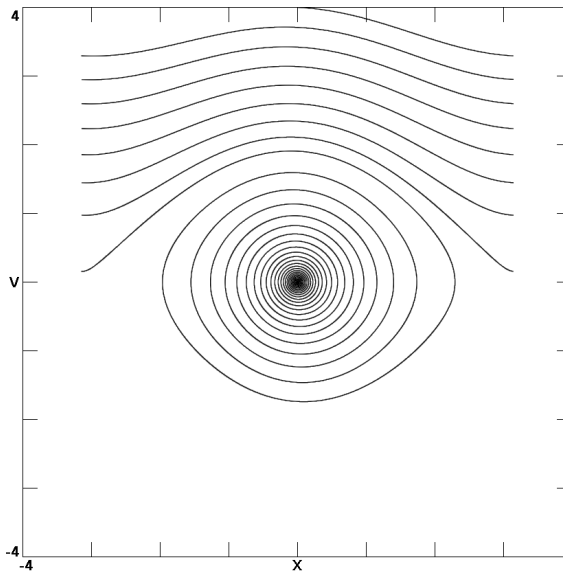

Fig. 1.4 State space plot showing a point attractor for the damped pendulum in Eq. (1.2) with $b = 0.05$ for an initial condition of $(x_0, v_0) = (0, 4)$.

where a coefficient of unity has been chosen for the damping, which causes rapid convergence to a kind of attractor called a *limit cycle*. Figure 1.5 shows the trajectories for eight different initial conditions, all of which approach the attractor, as do almost all initial conditions in the state space. The exceptions are the O-point and the X-point, along with its stable manifold. The O-point is still an equilibrium, but it is now unstable and is called a *repellor* rather than an attractor. It is sometimes called a *source* for obvious reasons. Similar behavior also occurs in a simpler system called the *van der Pol oscillator* (van der Pol, 1920, 1926) where the $\sin x$ in Eq. (1.3) is replaced by x, which eliminates the X-point.

Note that attractors in flows can be zero-dimensional (an equilibrium point) or one-dimensional (a limit cycle). Note also that the curves in Fig. 1.2 are not limit cycles, but rather are *invariant circles* since they do not attract nearby initial conditions. They are circles only in the topological sense of being closed loops and can be greatly distorted in shape. Attractors are a feature of dissipative systems and do not occur in conservative systems.

The antidamping that occurs near the origin implies an external source of energy, not explicit in the equations, and corresponds to positive feed-

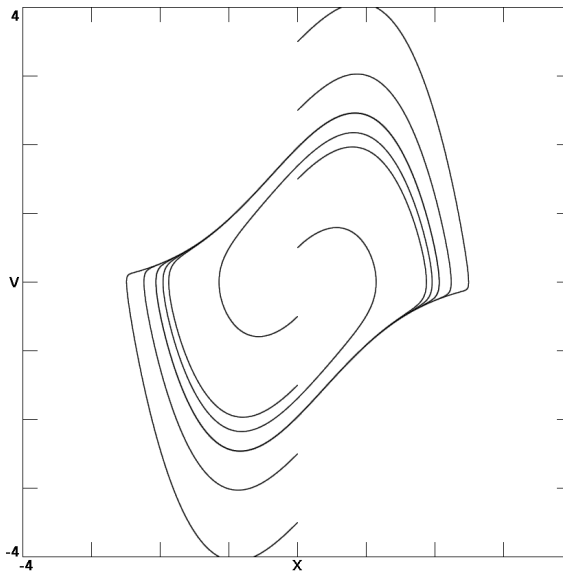

Fig. 1.5 State space plot showing a limit cycle for the damped pendulum in Eq. (1.3) for various initial conditions.

back. In fact, a public address system where the microphone gets too close to the speaker is an excellent example of a limit cycle, producing an unpleasant periodic audio oscillation of high amplitude (known as the *Larsen effect*). In such a case, the electrical energy comes from the wall socket. If the amplifier is unplugged, there is no antidamping, and the oscillations will cease. Such hidden energy sources will characterize nearly all of the examples in this book.

1.5 Chaos and Strange Attractors

An important feature of flows in two-dimensional state space is that the streamlines cannot cross except at an X-point, which is always an unstable equilibrium with a vanishing flow. The implication of this fact is that the preceding figures represent all the types of bounded flows that can occur in two-dimensional systems of ordinary differential equations. In particular, chaos cannot occur (However, see Chapter 5 for some exceptions). Chaos requires at least three dimensions so that the streamlines can cross by passing behind one another. This notion has been formalized in the *Poincaré–Bendixson theorem* (Hirsch *et al.*, 2004).

As an example of such a three-dimensional flow, consider the simple damped pendulum but with a sinusoidal forcing given by

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= -bv - \sin x + \sin \Omega t, \end{aligned} \tag{1.4}$$

where Ω is the frequency of the forcing, which is another parameter of the system. Such a system in which t appears explicitly on the right-hand side of the equations is said to be *nonautonomous*, but the time dependence can always be removed by adding another equation as in

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= -bv - \sin x + \sin z \\ \dot{z} &= \Omega, \end{aligned} \tag{1.5}$$

where z is the phase of the forcing function, which can be restricted to the range $0 \leq z < 2\pi$ (or $-\pi \leq z < \pi$) as was the case with x. The state space is now three-dimensional, and the governing equations are *autonomous* since t does not appear explicitly. Autonomous systems are important and desirable because there is a unique direction of the flow at almost every point in their state space independent of when the point is visited.

Note that the $\sin \Omega t$ in Eq. (1.4) could equally well have been replaced by $\cos \Omega t$ since the difference just amounts to starting the clock at a different time. Indeed, an arbitrary phase angle ϕ could have been added to give $\sin(\Omega t + \phi)$ without changing the dynamics. In terms of Eq. (1.5), the phase can be accounted for by changing the initial condition z_0. Furthermore, Ω can be restricted to positive values since the sine function has the same shape whether its argument is positive or negative.

Figure 1.6 shows the streamlines for such a flow with $b = 0.05$ and $\Omega = 0.8$ for an initial condition of $(x_0, v_0, z_0) = (0, 1, 0)$ projected onto the xv-plane with the z variable shown in shades of gray and with a subtle shadow to give the illusion of depth. The complicated and tangled behavior is a signature of chaos, and the trajectory winds around forever, never repeating, on an object called a *strange attractor*. Such an attractor has a noninteger dimension greater than 2.0 but less than 3.0. Only a portion of the state space trajectory is shown since it would otherwise densely cover a region of the plane and the individual streamlines would not be visible. As with the limit cycle, almost all initial conditions are drawn to the strange attractor. Dissipative chaotic flows generally have such strange attractors. Said differently, the appearance of a strange attractor in the state space dynamics is a signature of chaos, and they have been called the 'fingerprints of chaos' (Richards, 1999).

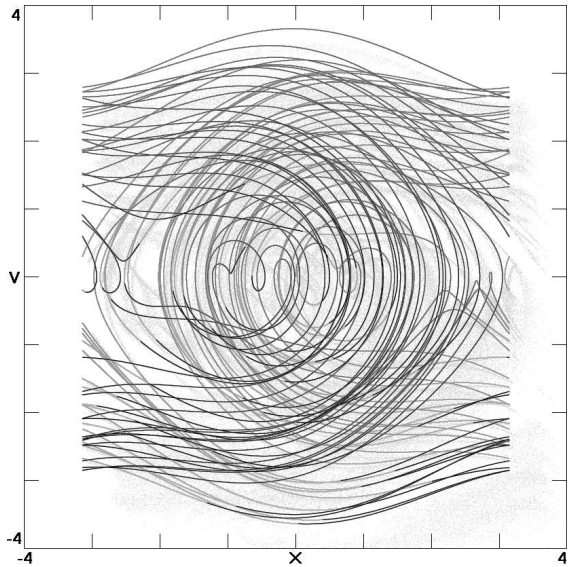

Fig. 1.6 State space plot showing a strange attractor for the damped forced pendulum in Eq. (1.5) with $b = 0.05$ and $\Omega = 0.8$ for an initial condition of $(x_0, v_0, z_0) = (0, 1, 0)$, $\lambda = (0.1400, 0, -0.1900)$.

Strange attractors are also called *chaotic attractors*, although the two terms are not precisely the same. There are examples of strange attractors that are not chaotic (Grebogi *et al.*, 1984; Ditto *et al.*, 1990; Feudel *et al.*, 2006) and chaotic attractors that are not strange (Anishchenko and Strelkova, 1997), but they are rare. The terminology depends on whether the geometric (strange) or dynamic (chaotic) property of the attractor is of primary interest. All the attractors in the remainder of this book are both strange and chaotic.

1.6 Poincaré Sections and Fractals

Given that the strange attractor for a three-dimensional chaotic flow must have a dimension greater than 2.0, it is useful to devise a method for displaying it in a two-dimensional plot. The *Poincaré section* provides such a method. Instead of projecting the entire state space trajectory onto the xv-plane as was done in Fig. 1.6, one can sample it at a particular value of the third variable (z in this case). Imagine that the three-dimensional trajectory in state space is illuminated briefly and periodically

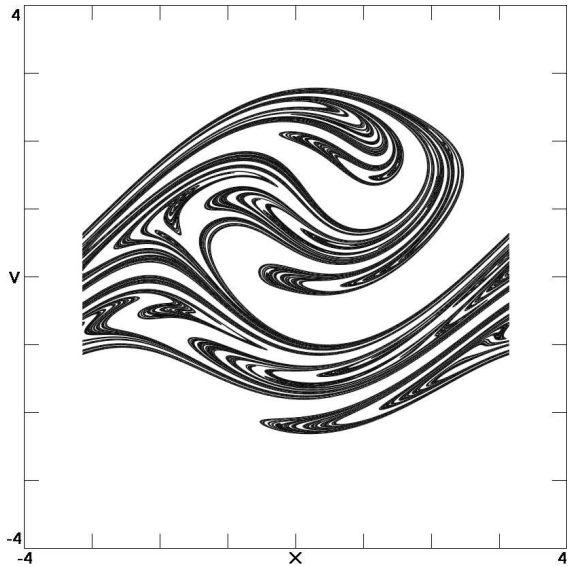

Fig. 1.7 Poincaré section for the damped forced pendulum in Eq. (1.5) with $b = 0.05$ and $\Omega = 0.8$ at $z \pmod{2\pi} = \pi/2$.

with a strobe light synchronized with the forcing function. This method amounts to taking a plane slice through the strange attractor at a fixed value of the phase z, which reduces the dimension of the attractor by 1.0. Since the attractor for a dissipative three-dimensional chaotic flow must lie between 2 and 3, its Poincaré section will have a dimension between 1 and 2. Figure 1.7 shows such a Poincaré section for Eq. (1.5) with $b = 0.05$ and $\Omega = 0.8$ at the times when $z \pmod{2\pi} = \pi/2$. Other values of z give similar results. In fact, a succession of Poincaré sections at successively increasing values of z can be used as frames of an endlessly repeating animation, some examples of which can be found at `http://sprott.physics.wisc.edu/fractals/animated/`.

A practical issue when calculating a Poincaré section is that the chosen time does not usually occur exactly at one of the finite iteration steps, but rather between two of them. In such a case, it is generally necessary and usually suffices to interpolate linearly between the two adjacent time steps. Failing to do so will smear out the fine-scale detail and will usually result in ugly plots. Of course it also helps to use relatively small time steps. Other more accurate methods are also available such as one suggested by Hénon (1982).

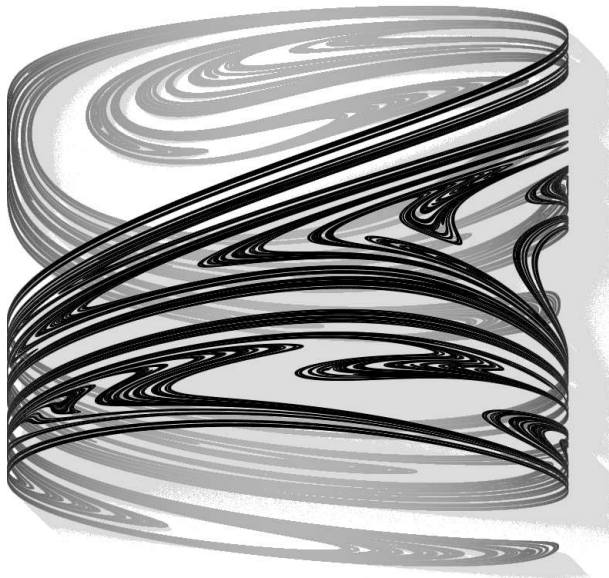

Fig. 1.8 Poincaré section for the damped forced pendulum in Eq. (1.5) with $b = 0.05$ and $\Omega = 0.8$ at $z \pmod{2\pi} = \pi/2$ as in Fig. 1.7 but wrapped onto a cylinder.

A Poincaré section can also be wrapped onto a cylinder as was done with the streamlines in Fig. 1.3. Figure 1.8 shows the result for the Poincaré section in Fig. 1.7. Note that there are many other sections that could have been taken, for example by fixing the angle of the pendulum (x) and plotting the velocity of the pendulum and phase of the forcing function on the surface of the cylinder, with a similar result.

Figures 1.7 and 1.8 illustrate why the attractor is *strange* and what it means to have a noninteger dimension. This object is an example of a *fractal*, about which books have been written, for example Mandelbrot (1983). This particular fractal has a dimension of about 1.7, which implies that the strange attractor for which it is a section has a dimension of about 2.7. The Poincaré section is somehow more than a long curvy line, but less than a surface with holes in it. Since its dimension is less than 2.0, it has no area. If one were to throw infinitely sharp darts at it, they would all miss. By the same token, it could not be seen except for the fact that the infinite collection of points that make it up are drawn with a nonzero size due to the limitation of computer graphics in the same way that a true mathematical line with zero width would not be visible if drawn on

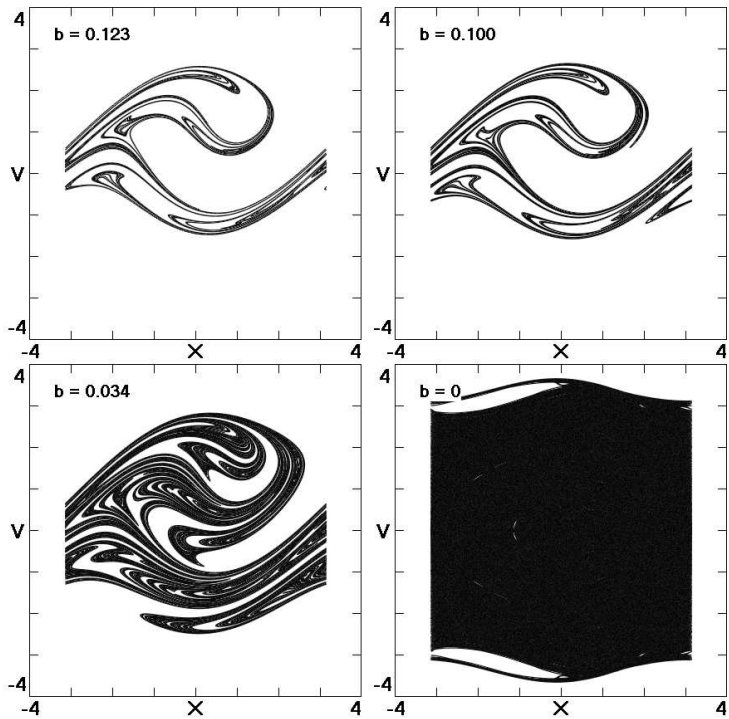

Fig. 1.9 Poincaré sections for the damped forced pendulum in Eq. (1.5) with $\Omega = 0.8$ for various values of damping (b) at $z \pmod{2\pi} = \pi/2$ showing how the dimension increases from 'line-like' to 'surface-like' with decreased damping.

a surface even though it contains infinitely many points. The dimension of the Poincaré section is the same no matter what section is taken with rare exceptions, and wrapping the object onto a cylinder as in Fig. 1.8 also preserves the dimension.

It is often the case that the dimension of a strange attractor increases as the damping is reduced, approaching an integer value in the limit of zero damping, and such is the case for this example as Fig. 1.9 shows. The four Poincaré sections have approximate dimensions of 1.4, 1.6, 1.8, and 2.0, respectively, and represent a transition from an object that is 'line-like' (dimension 1.0) to 'surface-like' (dimension 2.0). However, it is rare for the dimension of an attractor to increase continuously and monotonically as the damping is reduced while remaining chaotic all the way to the limit of zero damping.

1.7 Conservative Chaos

Systems without damping can also exhibit chaos as the case with $b = 0$ in Fig. 1.9 shows, but they do not have attractors. Instead, some initial conditions lie within the *chaotic sea* and eventually come arbitrarily close to every point in the sea. The chaotic sea is a region with integer dimension, although it can have infinitely many holes (or islands) of various sizes. If the number of islands of each size scales as a power of their size, like craters on the Moon, the chaotic sea is an example of a *fat fractal* (Farmer, 1985; Grebogi *et al.*, 1985), also called 'dusts with positive volume' by Mandelbrot (1983). Initial conditions within these islands or otherwise outside the chaotic sea typically have periodic orbits as indicated by closed loops (called *drift rings*) in the Poincaré section. In the full three-dimensional state space, these drift rings correspond to quasiperiodic orbits (two incommensurate frequencies) that lie on the surface of a 2-torus (a doughnut), on the axis of which is a simple periodic orbit as will be described in the next section.

The simplest example of the conservative forced pendulum is Eq. (1.4) with $b = 0$ and $\Omega = 1$, which can be written compactly as

$$\ddot{x} + \sin x = \sin t, \tag{1.6}$$

where $\ddot{x} = d^2x/dt^2 = d\dot{x}/dt = dv/dt$. Its Poincaré section at $z \pmod{2\pi} = \pi/2$ is shown in Fig. 1.10 for various initial conditions. Remember that a chaotic sea whose Poincaré section is two-dimensional is actually a three-dimensional region of chaos sampled in cross section. That a system as simple and elegant as Eq. (1.6) can have a chaotic solution as complicated as Fig. 1.10 is the theme of this book.

Note that when a system is said to be *conservative*, it does not necessarily mean that the mechanical energy is conserved, but only that there is no damping (friction). In fact the mechanical energy of the forced pendulum described by Eq. (1.6) varies drastically in time as Fig. 1.11 shows. When averaged over a sufficiently long time, the energy is conserved, but it can vary from zero to a rather large value as the forcing function supplies and removes energy from the system. The forcing function does work $W = x \sin t$ on the system for this case, and that work can be positive or negative, but it averages to zero and causes the energy to fluctuate. If it were an electrical circuit instead of a pendulum, we would say that the circuit is purely *reactive*, storing energy temporarily but not dissipating it through electrical resistance. Furthermore, in a dissipative system the

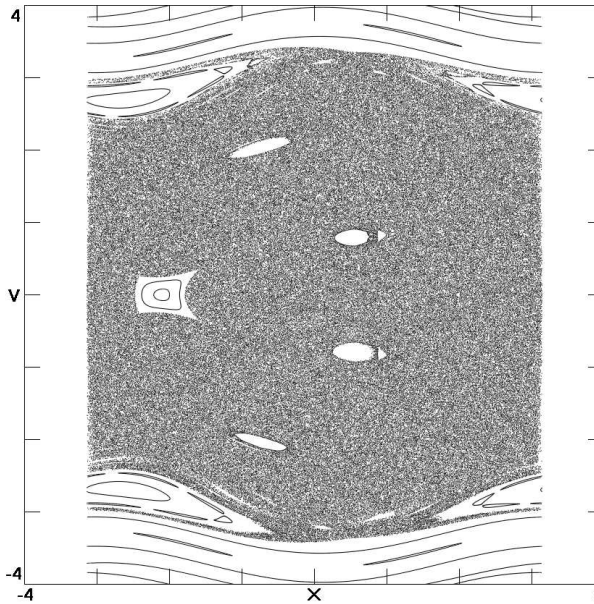

Fig. 1.10 Poincaré section at $z \pmod{2\pi} = \pi/2$ for the conservative forced pendulum in Eq. (1.6) for various initial conditions showing the chaotic sea with islands of periodicity.

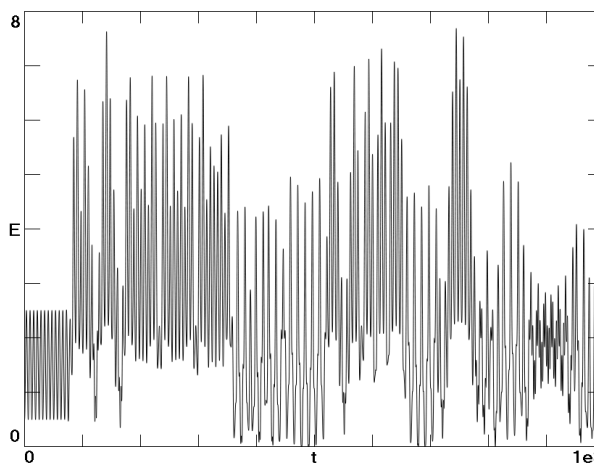

Fig. 1.11 Mechanical energy $E = v^2/2 + 1 - \cos x$ versus time for the conservative forced pendulum in Eq. (1.6) with initial conditions $(x_0, v_0, z_0) = (0, 1, 0)$ in the chaotic sea.

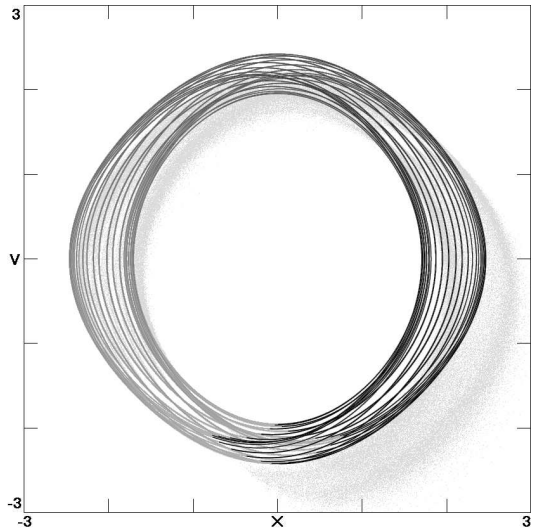

Fig. 1.12 State space plot showing an invariant 2-torus for the conservative forced pendulum in Eq. (1.6) with initial conditions $(x_0, v_0, z_0) = (-1.8, 0, \pi/2)$, $\lambda = (0, 0, 0)$.

energy decays to zero only when there is a point attractor. Otherwise, the forcing function or antidamping term supplies energy to replace that lost through dissipation, and the energy fluctuates about some average value as the trajectory moves around on the attractor.

1.8 Two-toruses and Quasiperiodicity

The holes in the Poincaré section in Fig. 1.10 represent regions in which the trajectories lie on the surface of a 2-torus, like the surface of a doughnut or inner tube, although the shape is only topologically equivalent to a doughnut and may be greatly distorted. Figure 1.12 shows the trajectory in the small hole in the vicinity of $(x, v) = (-2, 0)$ of Fig. 1.10. This trajectory is for an initial condition of $(x_0, v_0, z_0) = (-1.8, 0, \pi/2)$. Other initial conditions would lie on 2-toruses nested with the one shown. These nested toruses are successively smaller, ultimately forming an invariant circle at the center of the island. Each trajectory eventually comes arbitrarily close to every point on the surface of its torus, mapping out a two-dimensional surface in the three-dimensional state space. The surface has a cut where $z \pmod{2\pi} = 0$, but this cut could be removed by plotting the torus in cylindrical coordinates as was done in Fig. 1.3.

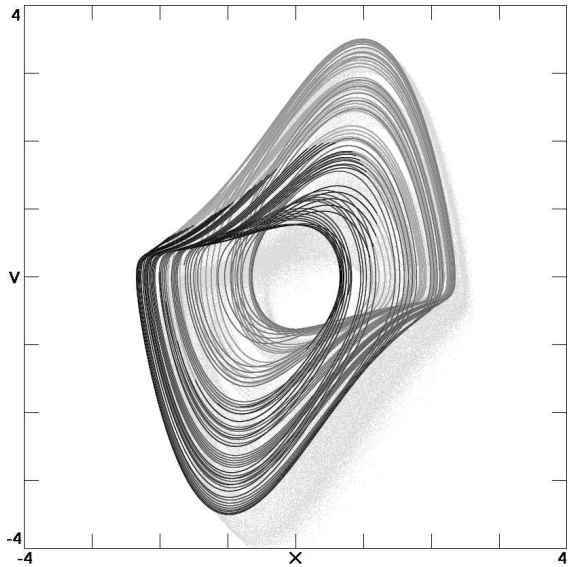

Fig. 1.13 State space plot showing an attracting 2-torus for the forced van der Pol oscillator in Eq. (1.7) with $A = 0.9$, $\Omega = 0.5$ and initial conditions $(x_0, v_0, z_0) = (1, 0.1, 0)$, $\lambda = (0, 0, -0.8264)$.

A 2-torus can also be attracting, just as a limit cycle, which can be considered as a 1-torus, is an attracting analog of the invariant circle. A simple example of an attracting 2-torus is the forced van der Pol oscillator

$$\ddot{x} + (x^2 - 1)\dot{x} + x = A \sin \Omega t \tag{1.7}$$

with $A = 0.9$ and $\Omega = 0.5$. Its state space attractor as shown in Fig. 1.13 resembles a strange attractor, but it is not. Rather, the trajectory will eventually fill the entire two-dimensional surface of a torus, and hence it is not a fractal. For other values of the parameters, the trajectory closes on itself after a finite number of transits around the torus, and thus fails to fill the torus, producing a limit cycle. For yet other values of the parameters, the attracting torus develops wrinkles that eventually give way to a strange attractor.

For both the invariant torus and the attracting torus, the trajectory on the torus consists of two independent periodic motions, one the short way around the torus, and the other the long way. For the trajectory to fill the torus, the periods of these motions must be *incommensurate*,

which means that their ratio is an irrational number (not a ratio of two integers). One would think that such a case would be overwhelmingly dominant, but for a nonlinear system like the forced van der Pol oscillator with two characteristic periods, the motion in xv-space tends to lock to a multiple of the period of the forcing function ($2\pi/\Omega$), in which case there is only a single period producing a limit cycle (or an invariant circle in the case of a conservative system). When the periods are incommensurate, the motion is said to be *quasiperiodic*, and in the simplest case, it consists of a superposition of two sine waves of incommensurate frequencies. In higher dimensions, one can have 3-toruses, 4-toruses, and so forth, but they are less common than chaos.

As Figs. 1.12 and 1.13 suggest, it is often difficult to distinguish quasiperiodicity from chaos. In this case, taking a Poincaré section makes it easy, but when there are more than two frequencies or when the Poincaré section itself is not a simple closed loop, it can be difficult to distinguish between the two. Better methods are available as we now describe.

1.9 Largest Lyapunov Exponent

Chaos can often be identified with some confidence by observing the strange attractor or chaotic sea in a state space plot or Poincaré section. However, it is useful to have a more objective and quantitative measure of chaos, and for that purpose the *Lyapunov exponent* is ideal. The main defining feature of chaos is the sensitive dependence on initial conditions. Two nearby initial conditions on the attractor or in the chaotic sea separate by a distance that grows exponentially in time when averaged along the trajectory, leading to long-term unpredictability. The Lyapunov exponent is the average rate of growth of this distance, with a positive value signifying sensitive dependence (chaos), a zero value signifying periodicity (or quasiperiodicity), and a negative value signifying a stable equilibrium. The Lyapunov exponent is named in honor of the Russian mathematician Aleksandr Lyapunov who was one of the first to explore dynamic stability over a hundred years ago. His name is sometimes spelled as 'Liapunov.'

Calculation of the Lyapunov exponent is conceptually simple since one only needs to follow two initially nearby trajectories and fit the logarithm (base-e throughout this book) of their separation to a linear function of time. The slope of the fit is the Lyapunov exponent. However, there are a number of practical difficulties.

First, one has to follow the trajectory for a sufficiently long time to be sure it is on the attractor before beginning to calculate the exponent. There is no way to know for sure how long this will take, and there are situations, thankfully somewhat rare, where the time required is infinite. For a conservative system, it is not necessary to wait since there is no attractor, but one needs to ensure that the initial condition lies in the chaotic sea and is not on one of the equilibrium points or stable manifolds.

Second, the initial conditions chosen for the two trajectories will not in general be oriented in the direction of the most rapid expansion, and their distance may in fact initially decrease before beginning to increase. Fortunately, they tend to orient in the direction of the most rapid separation (or the least rapid contraction) rather quickly (but not always!). Therefore, it is advisable to discard some of the early points from the fit. There is no harm in letting this reorientation occur while the trajectory is approaching the attractor, thereby accomplishing two things at once.

Third is the fact that the rate of separation usually varies considerably with position on the attractor or in the chaotic sea, and thus it is necessary to average over a long time, to achieve a proper weighting of the points in each region of the attractor or sea. Since the calculated value usually converges slowly, it is useful to develop some criterion for convergence as will be described shortly.

Finally, and most seriously, is the fact that the initial conditions must have a separation large enough to be expressed accurately in whatever precision the computer is using, while not shrinking to zero or growing to a significant fraction of the size of the attractor or sea. This requirement is nearly always incompatible with the requirement of following the trajectories long enough for the calculated value to converge.

This final problem is solved by moving the trajectories back to some appropriate small separation along the direction of their separation whenever they get too far apart or too close together. If the separation is readjusted to d_0 after a time δt (which can also be the integration step size) and the separation at the end of the time step is d_1, the Lyapunov exponent is given by $\lambda = \langle \log(d_1/d_0) \rangle / \delta t$, where the angle brackets $\langle\rangle$ denote an average along the trajectory. The *ergodic hypothesis* (Ruelle, 1976) asserts that a time average along a single representative trajectory is equivalent to a spatial average over the attractor, weighted by the density of points on the attractor.

Figure 1.14 shows a graph of the Lyapunov exponent versus b for the damped, forced pendulum in Eq. (1.5) with $\Omega = 0.8$ and initial conditions

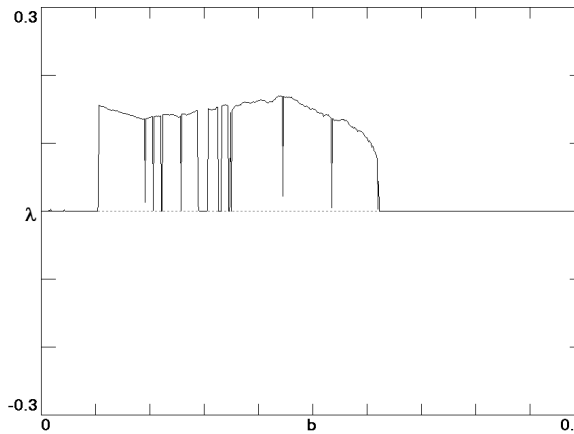

Fig. 1.14 Largest Lyapunov exponent for the damped, forced pendulum in Eq. (1.5) with $\Omega = 0.8$ and initial conditions $(x_0, v_0, z_0) = (0, 1, 0)$ as a function of the bifurcation parameter b.

$(x_0, v_0, z_0) = (0, 1, 0)$. It is common to see considerable structure in such plots since the Lyapunov exponent is usually not a smooth function of the parameters, with the discontinuities representing *bifurcations*, about which books have been written such as Wiggens (1988) and Kuznetsov (1995). However, it is easy to identify regions of positive and zero Lyapunov exponents and to confirm that the fractal Poincaré sections in Fig. 1.9 have positive Lyapunov exponents as expected for chaos, whereas the quasiperiodic trajectories such as the one in Fig. 1.12 have zero Lyapunov exponents (to within the precision of the calculation). Note the ubiquity of periodic windows in the chaotic region as is typical for low-dimensional chaotic systems. In fact, it is usually the case that there are infinitely many such windows of ever smaller widths, but fortunately there are also infinitely many values of the bifurcation parameter (b in this case) that lie outside these windows.

Graphs such as Fig. 1.14 are useful for studying the route to chaos and for identifying parameters that give chaos, but they are most useful for systems with a single adjustable parameter. When there are two or more parameters such as b and Ω in Eq. (1.5), it is more convenient and instructive to plot the regions in which the Lyapunov exponent is positive, zero, and negative in the plane of two of the parameters. Figure 1.15 shows such a plot for the damped, forced pendulum in Eq. (1.5), with black indicating chaos (C) and light gray indicating periodicity (P).

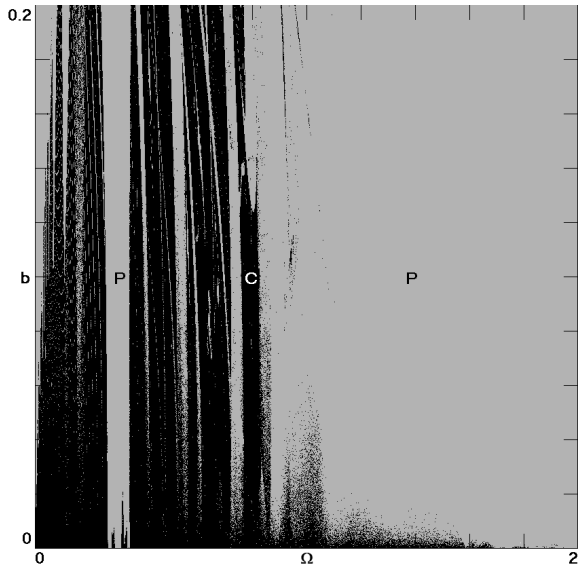

Fig. 1.15 Regions of different dynamics for the damped, forced pendulum in Eq. (1.5) with initial conditions $(x_0, v_0, z_0) = (0, 1, 0)$ as a function of the bifurcation parameters b and Ω. The chaotic regions (C) are shown in black, and the periodic regions (P) are shown in light gray.

A practical computational difficulty is distinguishing an exponent of zero from one that is small but nonzero. For the purpose of this book, values of the Lyapunov exponent for which $|\lambda| < 0.001$ will be taken as zero provided the estimated uncertainty in λ is less than $0.001 - |\lambda|$. Conversely, the Lyapunov exponent is taken as nonzero if $|\lambda| > 0.001$ provided the estimated uncertainty in λ is less than $|\lambda| - 0.001$. The sign of λ is assumed to have been determined correctly whenever one or the other of these conditions is satisfied, but there are always cases very near the bifurcation boundaries where the dynamics cannot be reliably classified.

What remains is to describe a method for estimating the best value of λ and its uncertainty $\delta\lambda$. For that purpose, a running average of $\log(d_1/d_0)$ is calculated at each time step (typically $\delta t = 0.1$) and the most recent m such values are stored in an array (typically $m = 1 \times 10^4$) as shown in Fig. 1.16. The array is divided into three equal groups ($m/3$ points each), and the average value of λ is calculated for the early (λ_e), mid (λ_m), and late (λ_l) third, as shown by the small circles in Fig. 1.16. The Lyapunov exponent is assumed to converge as $a + b/(t + C)$, and a curve of

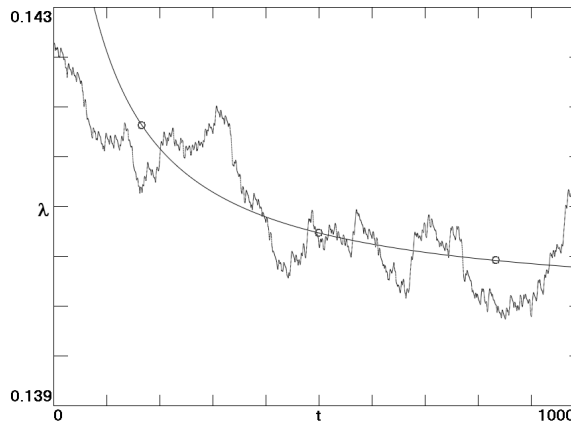

Fig. 1.16 Running average of the Lyapunov exponent over 1×10^4 time steps for the damped, forced pendulum in Eq. (1.5) with $b = 0.05$ and $\Omega = 0.8$ and initial conditions $(x_0, v_0, z_0) = (0, 1, 0)$, along with a curve fitted to the three average values shown with small circles.

that form is fitted to the data using least squares as shown by the smooth curve in Fig. 1.16, so that the extrapolated value for $t \to \infty$ is given by $\lambda = (2\lambda_e\lambda_l - \lambda_e\lambda_m - \lambda_m\lambda_l)/(\lambda_e - 2\lambda_m + \lambda_l)$ provided $|\lambda_l - \lambda_m| < |\lambda_m - \lambda_e|$. Otherwise, the Lyapunov exponent is not converging a and is estimated to be $\lambda = (\lambda_e + \lambda_m + \lambda_l)/3$. The uncertainty is assumed to be given by the range of values (largest minus smallest) stored in the array including the extrapolated value. This method has been tested and found to give reasonable results in cases where the exact Lyapunov exponent is known by other means, although the convergence is slower as the dimension of the attractor increases because it takes longer to sample the attractor or chaotic sea.

When λ is graphed versus a parameter as in Fig. 1.14, each data point is derived from a calculation that proceeds until $\delta\lambda$ is typically 10% the width of the line to ensure that the observed variations are real and not numerical. When values are quoted for the Lyapunov exponent, the uncertainty is typically less than 1×10^{-4} so that all four of the digits typically quoted are significant.

1.10 Lyapunov Exponent Spectrum

The previous discussion implied that there is a single Lyapunov exponent for a given flow. In fact, there are as many Lyapunov exponents as there

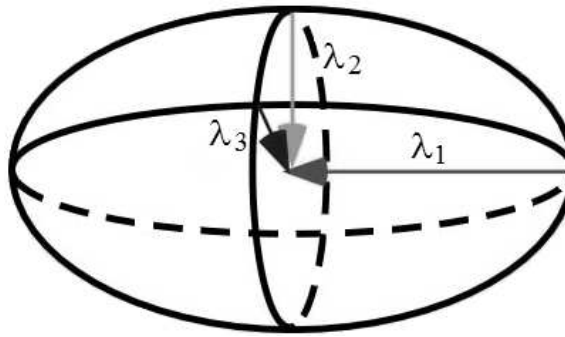

Fig. 1.17 A sphere of initial conditions distorts into an ellipsoid as time progresses, with each axis of the ellipsoid expanding at a rate given by one of the Lyapunov exponents (or contracting if that exponent is negative).

are state space variables, and what was calculated is only the largest (or least negative) of them. Fortunately, this is the only one that is required to identify chaos, since if it is positive, the system exhibits sensitive dependence on initial conditions independent of the values of the others, and if it is zero or negative, none of the others can be positive either.

The way to envision the Lyapunov exponents is to image a spherical ball filled with very many initial conditions in a three-dimensional state space. As the ball moves with the flow, it distorts into an ellipsoid as shown in Fig. 1.17, the longest axis of which expands at a rate given by the first Lyapunov exponent λ_1. Now consider a plane perpendicular to this expanding direction that cuts through the middle of the ellipsoid. In this plane, the points are contained within an ellipse whose major axis is expanding at a rate given by the second Lyapunov exponent λ_2 (or contracting if λ_2 is negative). The minor axis of this ellipse expands (or more likely contracts) at a rate given by the third Lyapunov exponent λ_3. From the definition, it is clear that $\lambda_1 \geq \lambda_2 \geq \lambda_3$, and this convention is universal.

As an example, Fig. 1.18 shows a succession of Poincaré sections for the conservative forced pendulum in Eq. (1.6) with a ball of radius $\sqrt{0.02}$ containing 10^6 initial conditions distributed throughout its volume. Even in a single transit ($N = 1$, corresponding to a time lapse of $\Delta t = 2\pi$), the ball has distorted into a highly elongated but very thin ellipsoid. In the direction perpendicular to the plane, there is neither expansion nor contraction since initial conditions distributed in that direction follow one another with a time lag that neither grows nor shrinks on average. Since the largest Lyapunov exponent is $\lambda_1 = 0.1491$, one would expect the ellipse

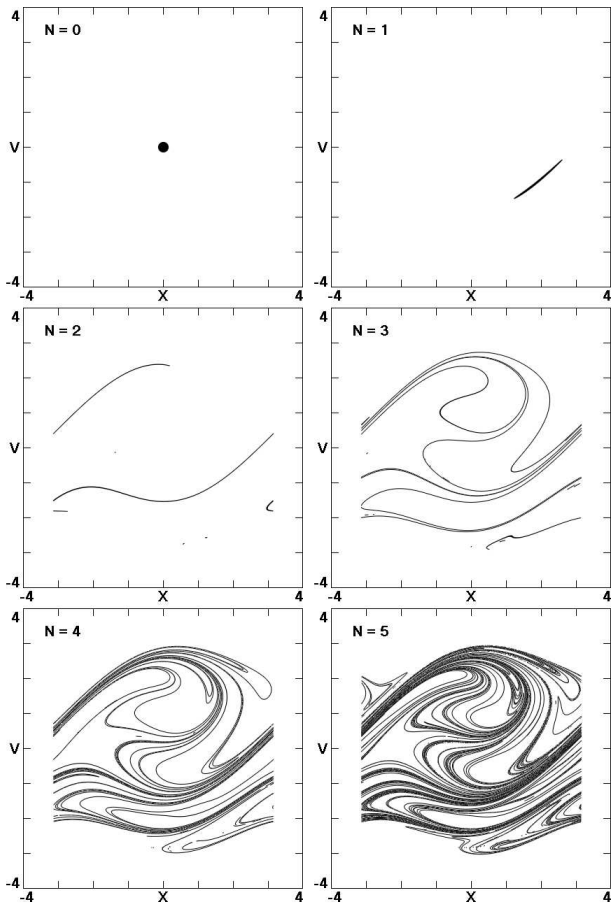

Fig. 1.18 A succession of Poincaré sections at $z \pmod{2\pi} = \pi/2$ for the conservative forced pendulum in Eq. (1.6) with a ball of initial conditions, showing the filamentation of phase space that occurs as a result of the chaos.

to elongate by a factor of $e^{0.1491\times 2\pi} \approx 2.55$ for each transit, and indeed it does on average, at least initially. However, the ellipse quickly distorts into a long, thin filament whose length grows much faster than expected because there are regions of state space near the separatrix where the local stretching is enormous. Consequently, after a very few transients, it has already begun to fill in much of the chaotic sea. The behavior is analogous to the way a drop of cream spreads throughout a cup of coffee after just a bit of stirring in a process called *filamentation*.

The idea generalizes to state spaces of any dimension, although it is hard to visualize dimensions greater than three. When we refer to *the* Lyapunov exponent λ, we will mean λ_1, although the Lyapunov exponent will usually be expressed as $\lambda = (\lambda_1, \lambda_2, \lambda_3, \ldots)$ as the occasion dictates. For example, a trajectory in the chaotic sea for Eq. (1.6) has $\lambda = (0.1491, 0, -0.1491)$. Note that the Lyapunov exponents for a conservative system are symmetric about zero because of the time-reversal invariance, and hence they necessarily sum to zero.

It is important to understand that as the ball moves with the flow, the orientation of the ellipsoid in state space is constantly changing and its axes are rarely aligned with the state space axes. Although there are as many Lyapunov exponents as there are state space variables, there is not a one-to-one correspondence between them. Furthermore, as the ellipsoid continues to stretch in one direction while contracting in another, it distorts into a long, thin filament as shown in Fig. 1.18 that eventually becomes a strange attractor as shown in cross section in Fig. 1.7 or a chaotic sea as shown in cross section in Fig. 1.10.

In a chaotic system, there must be stretching to cause the exponential separation of initial conditions but also folding to keep the trajectories from moving off to infinity. The folding requires that the equations of motion contain at least one nonlinearity, leading to the important principle that chaos is a property unique to nonlinear dynamical systems. If a system of equations has only linear terms, it cannot exhibit chaos no matter how complicated or high-dimensional it may be. However, nonlinearity does not guarantee chaos. For example, the $\sin \Omega t$ in Eq. (1.7) and elsewhere is a nonlinearity, but by itself cannot generate chaos. In fact, it can be replaced with two additional linear equations as will be shown in Chapter 6.

Note also that the volume of the ellipsoid is proportional to the product of its axes and thus increases according to $V = V_0 e^{(\lambda_1+\lambda_2+\lambda_3)t}$, so that $dV/dt = (\lambda_1 + \lambda_2 + \lambda_3)V$. The only way the volume can expand without limit is if the trajectory is unbounded and goes to infinity. The simplest example of such behavior is the one-dimensional flow given by the linear equation $\dot{x} = x$ whose solution is $x = x_0 e^t$. Two nearby initial conditions separate at the same exponential rate so that the one and only Lyapunov exponent is $\lambda = 1$. Even though the Lyapunov exponent is positive and there is sensitive dependence on initial conditions, we do not consider such cases to be chaotic. Chaotic systems must also be *recurrent*, which means that the trajectory will eventually return arbitrarily close to its starting

Table 1.1 Characteristics of the attractors for a bounded three-dimensional flow.

λ_1	λ_2	λ_3	Attractor	Dimension	Dynamic
$-$	$-$	$-$	Equilibrium point	0	Static
0	$-$	$-$	Limit cycle	1	Periodic
0	0	$-$	Attracting 2-torus	2	Quasiperiodic
0	0	0	Invariant torus	1 or 2	(Quasi)periodic
$+$	0	$-$	Strange	2 to 3	Chaotic

point (but not exactly since then it would then be periodic) and will do so repeatedly (Smale, 1967).

In practice, unbounded solutions in a model are detected and discarded whenever the absolute value of any of the state space variables exceeds some large value such as 1000. We are interested only in dissipative systems whose volume contracts onto an attractor or in conservative systems whose volume remains constant although perhaps stretching and thinning to such a degree that every point in the ball of initial conditions eventually comes arbitrarily close to any point in the chaotic sea.

Another property of bounded systems is that, unless the trajectory attracts to an equilibrium point where it stalls and remains forever, the points must continue moving forever with the flow. However, if we consider two initial conditions separated by a small distance along the direction of the flow, they will maintain their average separation forever since they are subject to the exact same flow but only delayed slightly in time. This fact implies that one of the Lyapunov exponents for a bounded continuous flow must be zero unless the flow attracts to a stable equilibrium.

The spectrum of Lyapunov exponents contains more information about the dynamics than does the largest exponent by itself. Considering that each exponent can be negative, zero, or positive, that their sum cannot be positive for a bounded system, and that at least one exponent must be zero except for a point attractor, there are five possible combinations for a three-dimensional state space as given in Table 1.1. In higher-dimensional state spaces, other cases are possible including attracting 3-toruses $(0, 0, 0, -, \ldots)$ and hyperchaos $(+, +, 0, -, \ldots)$.

What remains is the task of calculating the exponents other than λ_1. There are numerical procedures for doing this (Geist *et al.*, 1990), but they tend to be cumbersome and slow. Fortunately, there is a shortcut for three-dimensional chaotic flows. Since one exponent must be positive and one

must be zero, the third λ_3 must be negative. Therefore, if λ_1 and the sum of the exponents are known, then λ_3 can be determined from $\lambda_3 = \sum \lambda - \lambda_1$.

As mentioned earlier, the sum of the exponents is the rate of volume expansion (or contraction when negative as is usual) and is given by $\Sigma\lambda = \langle \partial\dot{x}/\partial x + \partial\dot{y}/\partial y + \partial\dot{z}/\partial z \rangle$ averaged along the trajectory. For example, the case in Eq. (1.5) is simply given by $\Sigma\lambda = \partial\dot{v}/\partial v = -b$, emphasizing that dissipation is the reason for state space contraction. Thus knowing that $\lambda_1 = 0.1400$ in Fig. 1.6 leads immediately to $\lambda = (0.1400, 0, -0.1900)$ since $b = 0.05$.

For four-dimensional conservative chaotic systems, the Lyapunov exponents occur in equal and opposite pairs, which means that there must be two zero exponents $\lambda_2 = \lambda_3 = 0$, and the lone negative exponent is then equal to the negative of the positive one $\lambda_4 = -\lambda_1$. For four-dimensional dissipative systems and for all systems with dimension greater than four, these methods fail, and one must resort to a more difficult numerical calculation of the spectrum of exponents. In those few cases where that is necessary in this book, the method described by Wolf *et al.* (1985) has been used.

1.11 Attractor Dimension

The spectrum of Lyapunov exponents allows one to calculate the dimension of the attractor. It is reasonable that when all the exponents are negative, corresponding to contraction in all directions, then the attractor must be a point with a dimension of zero. Think of an ellipsoid all of whose axes are shrinking. When the largest exponent is zero and all the others are negative, the only noncontracting direction is parallel to the flow, and hence the attractor must be a limit cycle with a dimension of one. When the largest two exponents are zero and all the rest are negative, the trajectory can move freely in one direction perpendicular to the flow, but contracts in the other perpendicular directions, in which case the attractor is a 2-torus, a surface with a dimension of two. The number of leading zero exponents represents the number of directions that the trajectory can fill without contracting or expanding and hence is the dimension of the attractor.

If λ_1 is positive and λ_2 is zero, then there is a surface defined by these two perpendicular directions in which a cluster of initial conditions expands without limit. Thus the attractor can have a dimension no less than two. However, if the sum of the first three exponents is negative, the volume

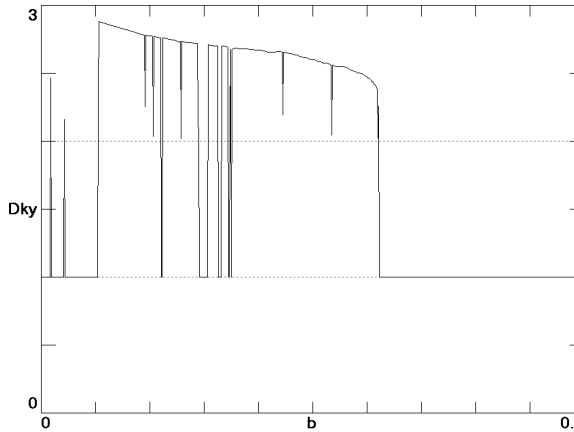

Fig. 1.19 Kaplan–Yorke dimension for the damped, forced pendulum in Eq. (1.5) with $\Omega = 0.8$ and initial conditions $(x_0, v_0, z_0) = (0, 1, 0)$ as a function of the damping b.

of a cluster of initial conditions must contract without limit, and so the attractor must have a dimension less than three. It is reasonable then to assign it a noninteger value between two and three in such a way that the cluster of initial conditions neither expands nor contracts in this fractional dimension.

The simplest way to do this is with a linear interpolation as suggested by Kaplan and Yorke (1979) and given by $D_{KY} = 2 - \lambda_1/\lambda_3$ in a three-dimensional state space. It is this argument that led to the estimates of the dimensions of the Poincaré sections in Fig. 1.9. As an example, Fig. 1.19 shows the Kaplan–Yorke dimension for the case in Eq. (1.5) with $\Omega = 0.8$ as a function of the damping b. The method easily generalizes to state spaces with dimensions greater than three. We will not generally calculate the Kaplan–Yorke dimensions for the cases presented in this book, but it is a simple matter to determine them with a pocket calculator from the spectrum of Lyapunov exponents, which is usually given.

Note that we now have three different dimensions to characterize a dynamical system. We have the number of degrees of freedom (1 for the pendulum), the dimension of the state space (2 for the free pendulum and 3 for the forced pendulum), and the dimension of the attractor (between 2 and 3 for the chaotic forced pendulum, depending on the damping). In addition, we could add the dimension of the parameter space (2 corresponding to b and Ω for the damped, forced pendulum).

1.12 Chaotic Transients

The alert reader may have noticed a slight discrepancy between Fig. 1.14 and Fig. 1.15 at small values of b where the former shows a periodic window with $\lambda = 0$ that is absent in the latter at $\Omega = 0.8$. The main difference between the two calculations is the time over which the trajectories are followed and the precision of the estimated Lyapunov exponent. A two-dimensional plot such as Fig. 1.15 requires much more computation (a factor of 800 in this case) than Fig. 1.14 for equivalent accuracy. Thus one would be inclined to accept the more accurate result in Fig. 1.14.

A careful examination of the case at $b = 0.01$ and $\Omega = 0.8$ where the two disagree shows that the discrepancy is caused by a chaotic transient. One way to verify and exhibit this effect is to make a plot similar to the Poincaré section in Fig. 1.7 but with v plotted versus the logarithm of time rather than versus x at those times when $z \pmod{2\pi}$ has a particular value such as $\pi/2$. This amounts to projecting the Poincaré section onto the v-axis so that a set of points projects onto another set of points. Any section with a dimension greater than or equal to one will project onto a line. Thus, if the attractor is a limit cycle with a dimension of one, its Poincaré section has a dimension of zero, whereas a chaotic system will have a dimension greater than two with a Poincaré section that projects onto a line.

Figure 1.20 shows such a plot. The logarithm of time is used so that a wide range of times can be displayed on the same plot. It is evident that the trajectory is chaotic until $t = 3.2 \times 10^4$, whereupon it abruptly becomes periodic and thereafter remains so. This is an example of *transient chaos*, and it is not particularly rare, although such very long transients typically occur only in small regions of parameter space. Edward Spiegel refers to transient chaos as 'pandemonium' (Smith, 2007).

The implication of such chaotic transients is that one has to be very cautious in declaring that a system has a chaotic attractor based solely on a numerical computation of its trajectory, even if the Lyapunov exponent appears to have converged to a large positive value. Most of the examples of chaos in this book have been verified by calculating for 2×10^9 iterations, corresponding typically to a time of $t = 2 \times 10^8$. The exceptions are the PDEs, DDEs, and other very high-dimensional systems, which are very computationally intensive. Another way to avoid being fooled by a chaotic transient is to see if the chaos persists over a range of parameters and initial conditions.

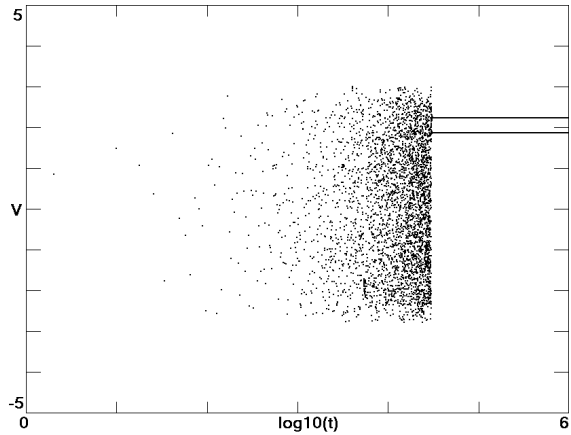

Fig. 1.20 Bifurcation diagram at $z \pmod{2\pi} = \pi/2$ versus time showing transient chaos for the damped, forced pendulum in Eq. (1.5) with $b = 0.01$ and $\Omega = 0.8$ for initial conditions $(x_0, v_0, z_0) = (0, 1, 0)$.

1.13 Intermittency

A phenomenon related to transient chaos is *intermittency* in which long periods of periodicity are interrupted by brief bursts of chaos, or perhaps the reverse (Pomeau and Manneville, 1980). Such systems are legitimately (though often only weakly) chaotic but require a long calculation for their Lyapunov exponent to converge. Intermittency is detected in the same way as transient chaos. Figure 1.21 shows an example of intermittency in the damped, forced pendulum with $b = 0.124304$ and $\Omega = 0.8$. The time scale here is linear rather than logarithmic. Fortunately, such behavior tends to occur only for very carefully chosen values of the parameters as is the case here, and so one is unlikely to stumble across it accidentally.

In searching for chaos, it is possible to have both false positives and false negatives, and one can never be completely confident that a system is chaotic based solely on numerical evidence, although considerable care has been taken to ensure that the chaotic examples in this book are not likely to be transients.

1.14 Basins of Attraction

The foregoing discussion assumed that a given dynamical system has only a single attractor for a given set of parameters, and thus by the ergodic

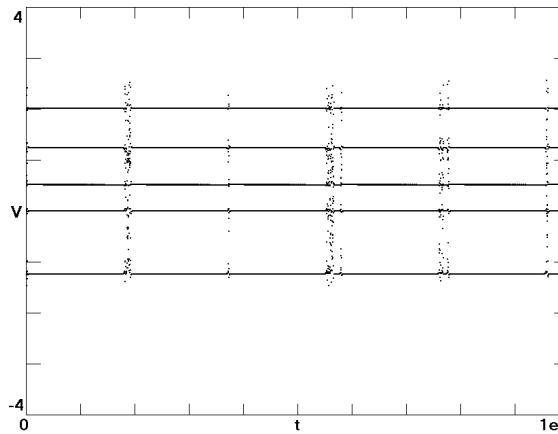

Fig. 1.21 Bifurcation diagram at $z \pmod{2\pi} = \pi/2$ versus time showing intermittency for the damped, forced pendulum in Eq. (1.5) with $b = 0.124304$ and $\Omega = 0.8$ for initial conditions $(x_0, v_0, z_0) = (0, 1, 0)$, $\lambda = (0.0037, 0, -0.1280)$.

hypothesis that the Lyapunov exponent is the same for all initial conditions. In fact, many dynamical systems have multiple coexisting attractors, sometimes infinitely many. Furthermore, even when a single attractor is present, some initial conditions may go to infinity rather than to the attractor, in which case infinity can be considered as an attractor. The various attractors can have quite different properties such as Lyapunov exponents and dimension, and can occur in all combinations — equilibrium points, limit cycles, toruses, and strange attractors. The number and type of attractors can change drastically as the parameters are varied.

Every attractor is contained within a *basin of attraction*, which is the region of state space over which initial conditions approach the attractor as $t \to \infty$. A useful analogy in two dimensions is a watershed representing those points on the Earth's surface where rainwater ends up in a given lake, or a wash basin that collects water from the faucet (tap in the UK) and directs it down the drain (or sink). Since chaos usually requires a three-dimensional state space, the basins for strange attractors are volumes (or hypervolumes in higher dimensions) inside of which the attractor resides. The basins can stretch to infinity, but more typically they have a boundary that may itself be a fractal (Aguirre *et al.*, 2009).

The damped, forced pendulum has limit cycles that coexist with the strange attractor for certain values of the parameters such as $0.0210 < b < 0.0275$ with $\Omega = 0.8$. Just as it is difficult to display an attractor

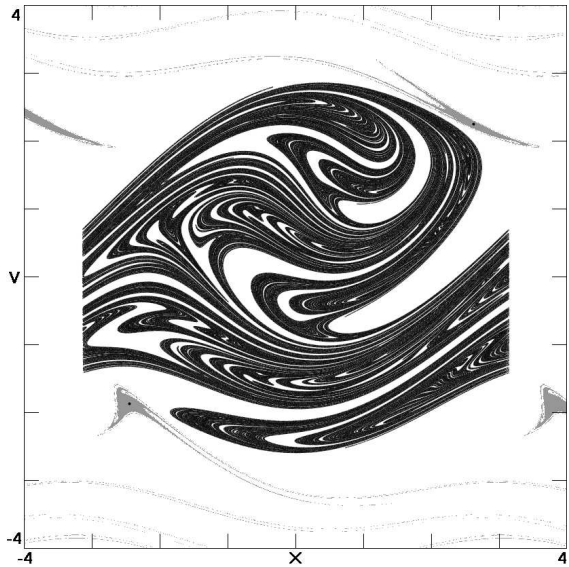

Fig. 1.22 Poincaré section for the damped, forced pendulum in Eq. (1.5) with $b = 0.024$ and $\Omega = 0.8$ at $z \pmod{2\pi} = \pi/2$ showing in gray the basin of attraction for the limit cycles that coexist with the strange attractor and shown in cross section by the two small black dots.

whose dimension is greater than 2.0 in a two-dimensional plot, it is even more difficult to display its basin of attraction, whose dimension is three or greater. However, one can plot a cross section of the basin in the same way the Poincaré section was plotted in Fig. 1.7. In fact, the plots can be overlaid since the variables are the same but evaluated at different times. Figure 1.22 shows such a plot for the damped, forced pendulum in Eq. (1.5) with $b = 0.024$ and $\Omega = 0.8$ at $z \pmod{2\pi} = \pi/2$. The small regions in gray with long, thin tails represent those initial conditions that go to each of a pair of limit cycles (represented by two small black dots, slightly enlarged for enhanced clarity) rather than to the strange attractor whose Poincaré section dominates the central part of the figure.

Figure 1.23 shows the two limit cycles superimposed on a Poincaré section of the strange attractor. The limit cycles correspond to unidirectional rotations of the pendulum in opposite directions, and they each repeat after three transits around the cylinder (a spatial periodicity in x of 6π).

Multiple coexisting attractors are more likely to occur when the damping is small, and when they do exist, it is likely that most if not all of

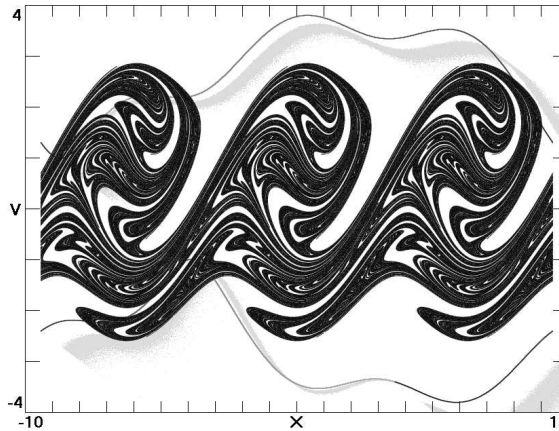

Fig. 1.23 Poincaré section for the damped, forced pendulum in Eq. (1.5) with $b = 0.024$ and $\Omega = 0.8$ at $z \pmod{2\pi} = \pi/2$ showing also projections of the two limit cycles onto the xv-plane.

them will be limit cycles (Feudel and Grebogi, 2003). Coexisting strange attractors are less common (Battelino *et al.*, 1988). There can be hundreds of coexisting attractors with fractal basin boundaries that come very close to almost every point in the state space (Feudel *et al.*, 1996). Systems with multiple attractors are said to be *multistable*.

The possible existence of multiple attractors means that it is necessary to search different initial conditions as well as different parameters when determining whether a given dynamical system is capable of exhibiting chaos. Furthermore, it means that plots such as Figs. 1.15 and 1.19 that show the regions of parameter space over which different types of dynamics occur are not generally unique, but rather depend on the chosen initial conditions. For this reason, initial conditions are given for most of the figures in this book even though they are usually not very critical.

In some cases, the parameters are slowly changed in the plots without reinitializing the variables with each change so as to remain in the chaotic or bounded region over the widest range of parameters. Such plots will often exhibit *hysteresis* in which case a second branch of the curve will be evident when the parameter is scanned in the opposite direction. Multiple attractors can often be identified in this way, but a more general method is just to try many different initial conditions and calculate for each case some value such as the Lyapunov exponent or more simply the mean value of a quantity such as $\langle v^2 \rangle$ that is not likely to be the same for the different

attractors (Sprott, 2006). This method allows the search for coexisting attractors to be automated, and it is the way Fig. 1.22 was produced.

Transient chaos can be viewed as a situation in which an attractor touches its basin of attraction but only at places that are rarely visited by the trajectory. The trajectory is initially drawn to the attractor and wanders around on it for a long time before eventually coming to a place outside the basin of attraction, whereupon it escapes. Think of a fly buzzing around in a box for a long time before discovering a small hole in the wall that leads to the outside world. Of course the hole could also be a small patch of flypaper that would bring the fly to a permanent halt, just as a stable equilibrium might for a transiently chaotic trajectory.

1.15 Numerical Methods

One of the reasons so much attention has been paid to discrete-time systems is that they can be solved with great precision using digital computers. On the other hand, the systems in this book are governed by differential equations whose solutions advance forward in continuous time and are much more poorly approximated by the numerical methods that are used to solve them, all of which essentially consist of reducing them to some nearly equivalent discrete-time system, with a finite but small time step. Analytic solutions are available for linear systems of ODEs and for a few nonlinear ones, but essentially all chaotic systems require numerical solutions. Many books have been written about numerical methods for solving ODEs, for example, Gear (1971) and Shampine and Gordon (1975).

Nearly all the systems in this book were solved using a fourth-order Runge–Kutta algorithm with adaptive step size similar to that described by Press *et al.* (2007). The fourth-order Runge–Kutta method is given by

$$\begin{aligned}
\mathbf{k}_1 &= \mathbf{F}(\mathbf{x}_t)\delta t \\
\mathbf{k}_2 &= \mathbf{F}(\mathbf{x}_t + \mathbf{k}_1/2)\delta t \\
\mathbf{k}_3 &= \mathbf{F}(\mathbf{x}_t + \mathbf{k}_2/2)\delta t \\
\mathbf{k}_4 &= \mathbf{F}(\mathbf{x}_t + \mathbf{k}_3)\delta t \\
\mathbf{x}_{t+\delta t} &= \mathbf{x}_t + \mathbf{k}_1/6 + \mathbf{k}_2/3 + \mathbf{k}_3/3 + \mathbf{k}_4/6,
\end{aligned} \tag{1.8}$$

where $\mathbf{x}$ is a vector, each component of which is one of the dynamical variables: $\mathbf{x} = (x, y, z, \ldots)$, and $\mathbf{F}$ is a vector function whose components are $\mathbf{F} = (\dot{x}, \dot{y}, \dot{z}, \ldots)$. Note that time does not appear explicitly on the

right-hand side of Eq. (1.8) because it can always be removed using the method illustrated in Eq. (1.5).

The maximum step size is typically taken as $\delta t = 0.1$, but for each step, the system is solved twice, once using the entire step and once using two half steps. If the two results differ by more than 1×10^{-8} in any of the variables, then the step is broken into as many as 1000 smaller time steps as required to achieve an accuracy of 1×10^{-8} over the whole step. Especially problematic cases have been checked by further reducing the step size or by putting a smaller error tolerance (typically 1×10^{-12}) on each step.

The maximum step size of $\delta t = 0.1$ is more than adequate for most cases because the natural period of oscillation is typically 1 radian per second when the parameters are of order unity, and thus there is the order of $2\pi/\delta t \approx 63$ iterations per cycle. For Eq. (1.1), the frequency of small-amplitude oscillations is exactly 1 radian per second, and the frequency of oscillation for the van der Pol oscillator is also very close to that value. Similarly, for the forced cases, the forcing frequency is 1 radian per second whenever $\Omega = 1$.

Even so, it is important to have independent checks on the reliability of the computations. One check is to monitor the energy versus time for a frictionless case such as Eq. (1.1) where it should be conserved. Another is to see that periodic solutions actually return to their starting points. Yet another is to see that the Lyapunov exponents sum to zero for those cases that have no dissipation. Such tests provide confidence that the chaotic cases presented here are not numerical artifacts resulting from a flawed integration method as can otherwise easily happen.

1.16 Elegance

Unlike the material in the previous sections, *elegance* is not a mathematical term, but rather, like beauty, is in the eye of the beholder. The concept is hard to quantify, but as Supreme Court Justice Potter Stewart said of pornography, 'I know it when I see it.' Certainly Eq. (1.6) is elegant, whereas a system such as

$$\ddot{x} - (0.28 - 0.32x^2 + 0.01\dot{x}^2)\dot{x} + 0.63\sin(-1.24x - 0.45) + 1.55x^3 = 3.42\sin 0.74t, \tag{1.9}$$

which is also chaotic and possesses many of the same properties, is much less elegant. Its state space plot is shown in Fig. 1.24.

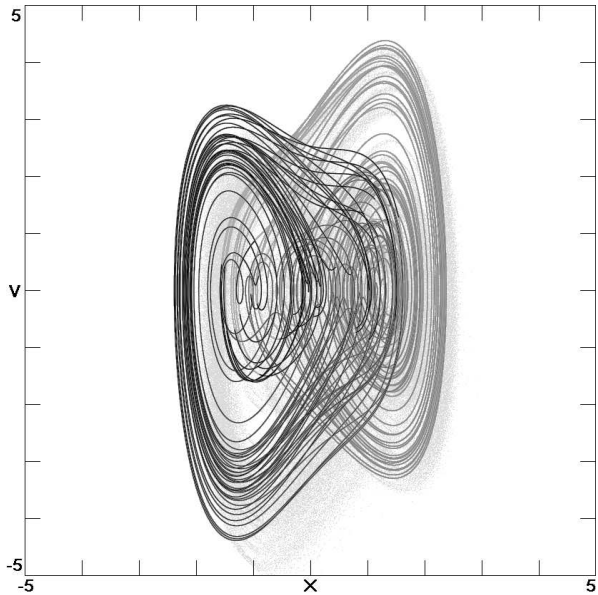

Fig. 1.24 State space plot showing a strange attractor for the very inelegant, damped, forced pendulum in Eq. (1.9) with an initial condition of $(x_0, v_0, z_0) = (1, 1, 0)$, $\lambda = (0.0978, 0, -0.1391)$.

While the term 'elegance' throughout this book refers to the form of the equation, many of the figures are elegant in their own right, but that is not what the book is about. The engaging plots are merely a delightful byproduct of the elegant equations, and there is no reason to believe that the elegance of the figure has any relation to the elegance of the equation that produced it. In fact, contrary to intuition, some of the most complicated dynamics arise from the simplest equations, while complicated equations often produce very simple and uninteresting dynamics. It is nearly impossible to look at a nonlinear equation and predict whether the solution will be chaotic or otherwise complicated. Small variations of a parameter can change a chaotic system into a periodic one, and vice versa.

A system of equations is deemed most elegant if it contains no unnecessary terms or parameters and if the parameters that remain have a minimum of digits. This notion can be quantified by writing an equation such as Eq. (1.9) in its most general form such as

$$\ddot{x} - (a_1 - a_2 x^2 - a_3 \dot{x}^2)\dot{x} + a_4 \sin(a_5 x + a_6) + a_7 x^3 = a_8 \sin a_9 t \qquad (1.10)$$

and adjusting the parameters a_1 through a_9 to achieve this end. Ideally, we want as many of the parameters of be zero as possible while preserving the chaos, and the greatest number of those that remain should be ± 1. Note that it is generally possible to linearly rescale the variables (x and t in this case) so that a corresponding number of the parameters are ± 1.

One way to quantify the elegance of Eq. (1.10) is thus to count the number of nonzero parameters and then add to that count the total number of digits including the decimal point but excluding leading and trailing zeros for any parameters that are not ± 1 so that integer parameters are preferred. The resulting number, which perhaps should be called the *inelegance*, is the quantity to be minimized. By this criterion, Eq. (1.9) has an inelegance of 39, whereas Eq. (1.6), when viewed as a special case of Eq. (1.10) with $a_1 = a_2 = a_3 = a_6 = a_7 = 0$ and $a_4 = a_5 = a_8 = a_9 = 1$, has an inelegance of 4.

With such a quantitative measure of elegance, one can automate the optimization process, starting either from a known chaotic system or from one that is potentially chaotic. In this way, the method can be used to find new chaotic systems and then to simplify them. Unfortunately, the optimization is computationally intensive since there is little alternative to trial and error. In practice, the method starts from a known chaotic case or from an arbitrary place in parameter space and searches a random Gaussian neighborhood of the starting point, typically with a variance of 1.0, seeking a case with a Lyapunov exponent greater than 0.001. When a candidate is found, the search moves to that place in parameter space and continues searching from there. Each time a chaotic case is found, the search thereafter considers only cases whose elegance is at least as great as the best case found so far.

A system such as Eq. (1.10) typically converges to an optimum after a few thousand trials and about an hour of computation on a personal computer as Fig. 1.25 shows. One can achieve faster convergence at the expense of more false positives by relaxing the precision of the calculation. It also helps to reduce the number of parameters so that the search space has a lower dimension or by starting with an otherwise more elegant case. Sometimes it is useful to put constraints on the search space such as limiting the search to only positive values of the parameters or by imposing some special desired symmetry (Gilmore and Letellier, 2007). The cases in this book represent about ten CPU-years of searching and a similar amount verifying the results and calculating the Lyapunov exponents.

Of course, there may not be a single optimum since more than one

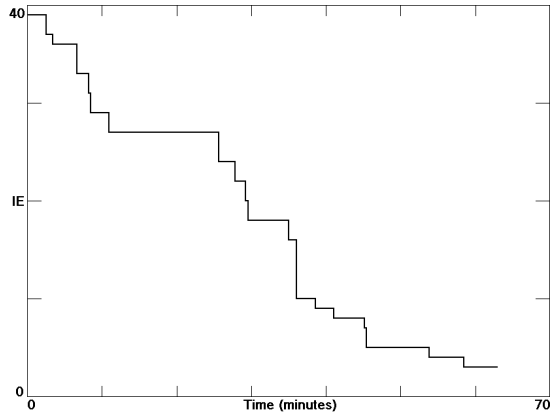

Fig. 1.25 Time history by the wall clock of the inelegance during a typical search for the maximally elegant form of Eq. (1.10) as given by Eq. (1.6) starting from the very inelegant form in Eq. (1.9) using a vintage 2008 personal computer.

combination of parameters may have the same elegance. In that event, the case with the largest Lyapunov exponent is usually chosen, or the case with coefficients closest to unity (2 is better than 3, and 0.9 is better than 0.8), or perhaps a case is chosen that has some nice symmetry such as equal values of two of the parameters, or simply one whose state space plot is more 'interesting.'

Just as one can find the most elegant set of parameters for a given system, it is possible to find the most elegant set of initial conditions within the basin of attraction or chaotic sea. However, it is usually more useful to have initial conditions that are close to the attractor to reduce the transients that would otherwise occur. Thus most of the initial conditions cited in this book are a compromise chosen to be close to the attractor (typically within a distance of 0.1) but using values of the initial conditions rounded to a few digits, while avoiding cases for which all the time derivatives are zero, which would imply an equilibrium point.

Finally, it is important to verify any chaotic system found in this way with a much longer, independent calculation of higher precision to avoid programming errors, transients, unboundedness, or other anomalies. Even so, there is no claim, much less guarantee, that the cases in this book are maximally elegant by the chosen criterion. Finding such cases is a goal, but failure to do so is not a serious flaw since the criterion is arbitrary. There is ample room for the interested reader to find cases yet more elegant than those presented here.