Chapter 1
Introduction and overview

1.1 Some history

Chaotic dynamics may be said to have started with the work of the French mathematician Henri Poincaré at about the turn of the century. Poincaré's motivation was partly provided by the problem of the orbits of three celestial bodies experiencing mutual gravational attraction (e.g., a star and two planets). By considering the behavior of orbits arising from *sets* of initial points (rather than focusing on *individual* orbits), Poincaré was able to show that very complicated (now called chaotic) orbits were possible. Subsequent noteworthy early mathematical work on chaotic dynamics includes that of G. Birkhoff in the 1920s, M. L. Cartwright and J. E. Littlewood in the 1940s, S. Smale in the 1960s, and Soviet mathematicians, notably A. N. Kolmogorov and his coworkers. In spite of this work, however, the possibility of chaos in real physical systems was not widely appreciated until relatively recently. The reasons for this were first that the mathematical papers are difficult to read for workers in other fields, and second that the theorems proven were often not strong enough to convince researchers in these other fields that this type of behavior would be important in their systems. The situation has now changed drastically, and much of the credit for this can be ascribed to the extensive numerical solution of dynamical systems on digital computers. Using such solutions, the chaotic character of the time evolutions in situations of practical importance has become dramatically clear. Furthermore, the complexity of the dynamics cannot be blamed on unknown extraneous experimental effects, as might be the case when dealing with an actual physical system.

In this chapter, we shall provide some of the phenomenology of chaos and will introduce some of the more basic concepts. The aim is to provide a motivating overview[1] in preparation for the more detailed treatments to be pursued in the rest of this book.

1.2 Examples of chaotic behavior

Most students of science or engineering have seen examples of dynamical behavior which can be fully analyzed mathematically and in which the system eventually (after some transient period) settles either into periodic motion (a limit cycle) or into a steady state (i.e., a situation in which the system ceases its motion). When one relies on being able to specify an orbit analytically, these two cases will typically (and falsely) appear to be the only important motions. The point is that chaotic orbits are also very common but cannot be represented using standard analytical functions. Chaotic motions are neither steady nor periodic. Indeed, they appear to be very complex, and, when viewing such motions, adjectives like wild, turbulent, and random come to mind. In spite of the complexity of these motions, they commonly occur in systems which themselves are not complex and are even surprisingly simple. (In addition to steady state, periodic and chaotic motion, there is a fourth common type of motion, namely quasiperiodic motion. We defer our discussion of quasiperiodicity to Chapter 6.)

Before giving a definition of chaos we first present some examples and background material. As a first example of chaotic motion, we consider an experiment of Moon and Holmes (1979). The apparatus is shown in Figure 1.1. When the apparatus is at rest, the steel beam has two stable steady-state equilibria: either the tip of the beam is deflected toward the left magnet or toward the right magnet. In the experiment, the horizontal position of the apparatus was oscillated sinusoidally with time. Under certain conditions, when this was done, the tip of the steel beam was observed to oscillate in a very irregular manner. As an indication of this very irregular behavior, Figure 1.2(a) shows the output signal of a strain gauge attached to the beam (Figure 1.1). Although the apparatus appears to be very simple, one might attribute the observed complicated motion to complexities in the physical situation, such as the excitation of higher order vibrational modes in the beam, possible noise in the sinusoidal shaking device, etc. To show that it is not necessary to invoke such effects, Moon and Holmes considered a simple model for their experiment, namely, the forced Duffing equation in the following form,

$$\frac{d^2y}{dt^2} + \nu\frac{dy}{dt} + (y^3 - y) = g\sin t. \tag{1.1}$$

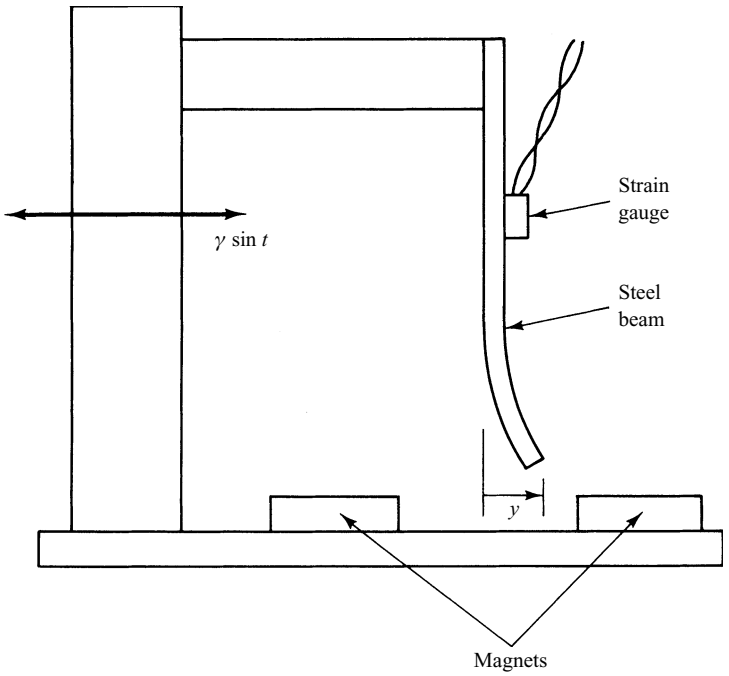

Figure 1.1 The apparatus of Moon and Holmes (1979).

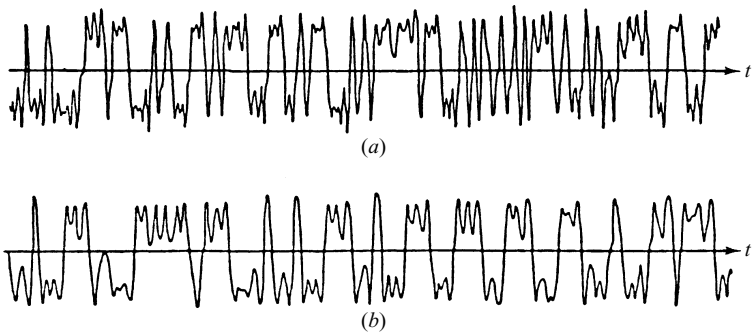

Figure 1.2 (*a*) Signal from the strain gauge. (*b*) Numerical solution of Eq. (1.1) (Moon and Holmes, 1979).

In Eq. (1.1), the first two terms represent the inertia of the beam and dissipative effects, while the third term represents the effects of the magnets and the elastic force. The sinusoidal term on the right-hand side represents the shaking of the apparatus. In the absence of shaking ($g = 0$), Eq. (1.1) possesses two stable steady states, $y = 1$ and $y = -1$, corresponding to the two previously mentioned stable steady states of the beam. (There is also an unstable steady state $y = 0$.) Figure 1.2(b) shows the results of a digital computed numerical solution of Eq. (1.1) for a particular choice of ν and g. We observe that the results of the physical experiment are qualitatively similar to those of the numerical solution.

Thus, it is unnecessary to invoke complicated physical processes to explain the observed complicated motion.

As a second example, we consider the experiment of Shaw (1984) illustrated schematically in Figure 1.3. In this experiment, a slow steady inflow of water to a 'faucet' was maintained. Water drops fall from the faucet, and the times at which successive drops pass a sensing device are recorded. Thus, the data consists of the discrete set of times $t_1, t_2, \ldots, t_n, \ldots$ at which drops were observed by the sensor. From these data, the time intervals between successive drops can be formed, $\Delta t_n \equiv t_{n+1} - t_n$. When the inflow rate to the faucet is sufficiently small, the time intervals Δt_n are all equal. As the inflow rate is increased, the time interval sequence becomes periodic with a short interval Δt_a followed by a longer interval Δt_b, so that the sequence of time intervals is of the form $\ldots, \Delta t_a, \Delta t_b, \Delta t_a, \Delta t_b, \Delta t_a, \ldots$. We call this a period two sequence since $\Delta t_n = \Delta t_{n+2}$. As the inflow rate is further increased, periodic sequences of longer and longer periods were observed, until, at sufficiently large inflow rate, the sequence $\Delta t_1, \Delta t_2, \Delta t_3, \ldots$ apparently has no regularity. This irregular sequence is argued to be due to chaotic dynamics.

As a third example, we consider the problem of chaotic Rayleigh–Benard convection, originally studied theoretically and computationally in the seminal paper of Lorenz (1963) and experimentally by, for example, Ahlers and Behringer (1978), Gollub and Benson (1980), Bergé *et al.* (1980) and Libchaber and Maurer (1980). In Rayleigh–Benard convection, one considers a fluid contained between two rigid plates and subjected to gravity, as shown in Figure 1.4. The bottom plate is maintained at a higher temperature $T_0 + \Delta T$ than the temperature T_0 of the top plate. As a result, the fluid near the warmer lower plate expands, and buoyancy creates a tendency for this fluid to rise. Similarly, the cooler

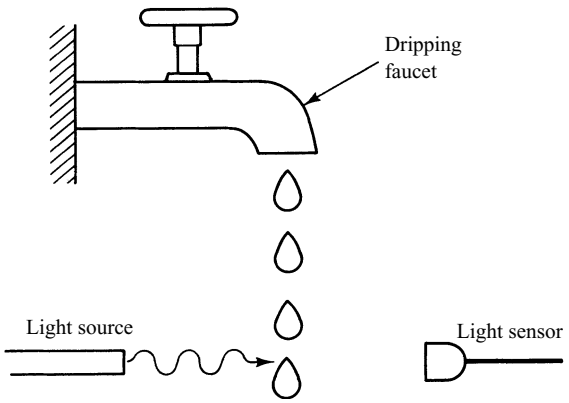

Figure 1.3 Schematic illustration of the experiment of Shaw (1984).

more dense fluid near the top plate has a tendency to fall. While Lorenz's equations are too idealized a model to describe the experiments accurately, in the case where the experiments were done with vertical bounding sidewalls situated at a spacing of two to three times the distance between the horizontal walls, there was a degree of qualitative correspondence between the model and the experiments. In particular, in this case, for some range of values of the temperature difference ΔT, the experiments show that the fluid will execute a *steady* convective cellular flow, as shown in the figure. At a somewhat larger value of the temperature difference, the flow becomes time-dependent, and this time dependence is chaotic. This general behavior is also predicted by Lorenz's paper.

From these simple examples, it is clear that chaos should be expected to be a very common basic dynamical state in a wide variety of systems. Indeed, chaotic dynamics has by now been shown to be of potential importance in many different fields including fluids,[2] plasmas,[3] solid state devices,[4] circuits,[5] lasers,[6] mechanical devices,[7] biology,[8] chemistry,[9] acoustics,[10] celestial mechanics,[11] etc.

In both the dripping faucet example and the Rayleigh–Benard convection example, our discussions indicated a situation as shown schematically in Figure 1.5. Namely, there was a system parameter, labeled p in Figure 1.5, such that, at a value $p = p_1$, the motion is observed to be nonchaotic, and at another value $p = p_2$, the motion is chaotic. (For the faucet example, p is the inflow rate, while for the example of Rayleigh–Benard convection, p is the temperature difference ΔT.) The natural question raised by Figure 1.5 is *how does chaos come about as the parameter p is varied continuously from* p_1 *to* p_2? That is, how do the dynamical motions of the system evolve with continuous variation of p from p_1 and p_2? This question of the *routes to chaos*[12] will be considered in detail in Chapter 8.

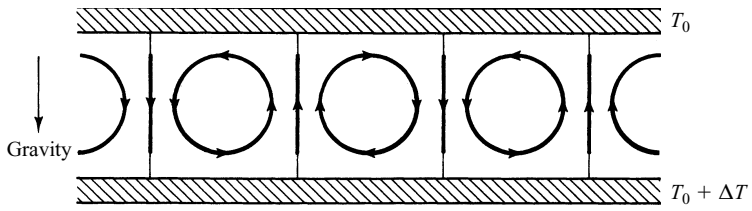

Figure 1.4 Rayleigh–Benard convection.

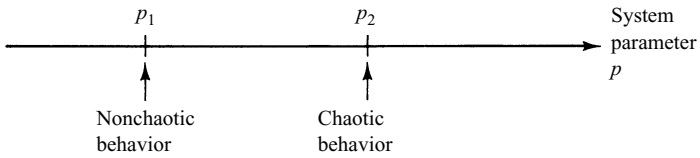

Figure 1.5 Schematic illustration of the question of the transition to chaos with variation of a system parameter.

1.3 Dynamical systems

A *dynamical system* may be defined as a deterministic mathematical prescription for evolving the state of a system forward in time. Time here either may be a continuous variable, or else it may be a discrete integer-valued variable. An example of a dynamical system in which time (denoted t) is a continuous variable is a system of N first-order, autonomous, ordinary differential equations,

$$\left.\begin{aligned} dx^{(1)}/dt &= F_1(x^{(1)}, x^{(2)}, \ldots, x^{(N)}), \\ dx^{(2)}/dt &= F_2(x^{(1)}, x^{(2)}, \ldots, x^{(N)}), \\ &\vdots \\ dx^{(N)}/dt &= F_N(x^{(1)}, x^{(2)}, \ldots, x^{(N)}), \end{aligned}\right\} \tag{1.2}$$

which we shall often write in vector form as

$$d\mathbf{x}(t)/dt = \mathbf{F}[\mathbf{x}(t)], \tag{1.3}$$

where $\mathbf{x}$ is an N-dimensional vector. This is a dynamical system because, for any initial state of the system $\mathbf{x}(0)$, we can in principle solve the equations to obtain the future system state $\mathbf{x}(t)$ for $t > 0$. Figure 1.6 shows the path followed by the system state as it evolves with time in a case where $N = 3$. The space $(x^{(1)}, x^{(2)}, x^{(3)})$ in the figure is referred to as *phase space*, and the path in phase space followed by the system as it evolves with time is referred to as an *orbit* or *trajectory*. Also, it is common to refer to a continous time dynamical system as a *flow*. (This latter terminology is apparently motivated by considering the trajectories generated by *all* the initial conditions in the phase space as roughly analogous to the paths followed by the particles of a flowing fluid.)

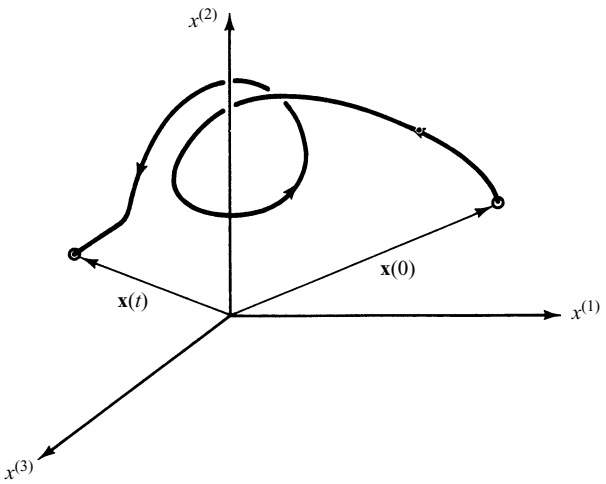

Figure 1.6 An orbit in a three-dimensional ($N = 3$) phase space.

In the case of discrete, integer-valued time (with n denoting the time variable, $n = 0, 1, 2, \ldots$), an example of a dynamical system is a map, which we write in vector form as

$$\mathbf{x}_{n+1} = \mathbf{M}(\mathbf{x}_n), \tag{1.4}$$

where $\mathbf{x}_n$ is N-dimensional, $\mathbf{x}_n = (x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)})$. Given an initial state $\mathbf{x}_0$, we obtain the state at time $n = 1$ by $\mathbf{x}_1 = \mathbf{M}(\mathbf{x}_0)$. Having determined $\mathbf{x}_1$, we can then determine the state at $n = 2$ by $\mathbf{x}_2 = \mathbf{M}(\mathbf{x}_1)$, and so on. Thus, given an initial condition $\mathbf{x}_0$, we generate an orbit (or trajectory) of the discrete time system: $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$. As we shall see, a continous time system of dimensionality N can often profitably be reduced to a discrete time map of dimensionality $N - 1$ via the Poincaré surface of section technique.

It is reasonable to conjecture that the complexity of the possible structure of orbits can be greater for larger system dimensionality. Thus, a natural question is *how large does N have to be in order for chaos to be possible*? For the case of N first-order autonomous ordinary differential equations, the answer is that

$$N \geqslant 3 \tag{1.5}$$

is sufficient.[13] Thus, if one is given an autonomous first-order system with $N = 2$, chaos can be ruled out immediately.

Example: Consider the forced damped pendulum equation (cf. Figure 1.7)

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \nu\frac{\mathrm{d}\theta}{\mathrm{d}t} + \sin\theta = T\sin(2\pi ft), \tag{1.6a}$$

where the first term represents inertia, the second, friction at the pivot, the third, gravity, and the term on the right-hand side represents a sinusoidal torque applied at the pivot. (This equation also describes the behavior of a simple Josephson junction circuit.) We ask: is chaos ruled out for the driven damped pendulum equation? To answer this question, we put the equation (which is second-order and nonautonomous) into first-order autonomous form by the substitution

$$x^{(1)} = \mathrm{d}\theta/\mathrm{d}t,$$
$$x^{(2)} = \theta,$$
$$x^{(3)} = 2\pi ft.$$

(Note that, since both $x^{(2)}$ and $x^{(3)}$ appear in Eq. (1.6a) as the argument of a sine function, they can be regarded as angles and may, if desired, be defined to lie between 0 and 2π.) The driven damped pendulum equation then yields the following first-order autonomous system.

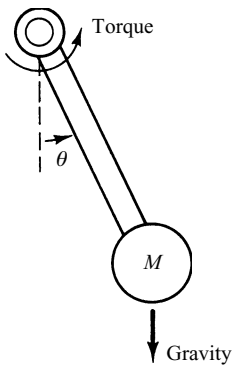

Figure 1.7 Forced, damped pendulum.

$$\left.\begin{aligned} dx^{(1)}/dt &= T\sin x^{(3)} - \sin x^{(2)} - \nu x^{(1)}, \\ dx^{(2)}/dt &= x^{(1)}, \\ dx^{(3)}/dt &= 2\pi f. \end{aligned}\right\} \tag{1.6b}$$

Since $N = 3$, chaos is not ruled out. Indeed, numerical solutions show that both chaotic and periodic solutions of the driven damped pendulum equation are possible depending on the particular choice of system parameters ν, T and f.

We now consider the question of the required dimensionality for chaos for the case of maps. In this case, we must distinguish between invertible and noninvertible maps. We say the map $\mathbf{M}$ is invertible if, given $\mathbf{x}_{n+1}$, we can solve $\mathbf{x}_{n+1} = \mathbf{M}(\mathbf{x}_n)$ uniquely for $\mathbf{x}_n$. If this is so, we denote the solution for $\mathbf{x}_n$ as

$$\mathbf{x}_n = \mathbf{M}^{-1}(\mathbf{x}_{n+1}), \tag{1.7}$$

and we call $\mathbf{M}^{-1}$ the inverse of $\mathbf{M}$. For example, consider the one-dimensional ($N = 1$) map[14],

$$\mathbf{M}(x) = rx(1 - x), \tag{1.8}$$

which is commonly called the 'logistic map.' As shown in Figure 1.8, this map is not invertible because for a given x_{n+1} there are two possible values of x_n from which it could have come. On the other hand, consider the two-dimensional map,

$$\begin{aligned} x_{n+1}^{(1)} &= f(x_n^{(1)}) - Jx_n^{(2)}, \\ x_{n+1}^{(2)} &= x_n^{(1)}. \end{aligned} \tag{1.9}$$

This map is clearly invertible as long as $J \neq 0$,

$$\begin{aligned} x_n^{(1)} &= x_{n+1}^{(2)}, \\ x_n^{(2)} &= J^{-1}[f(x_{n+1}^{(2)}) - x_{n+1}^{(1)}]. \end{aligned} \tag{1.10}$$

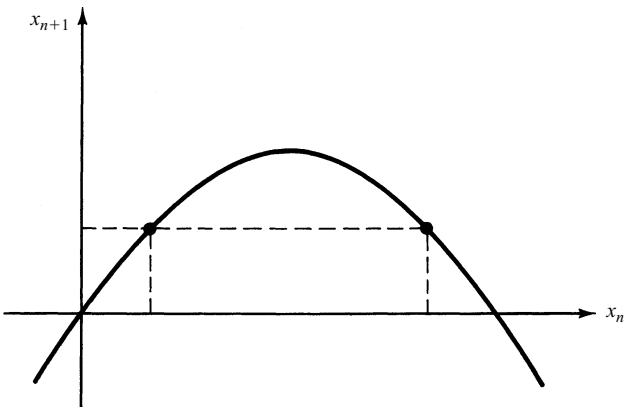

Figure 1.8 Noninvertibility of the logistic map.

We can now state the dimensionality requirements on maps. If the map is invertible, then there can be no chaos unless

$$N \geqslant 2. \tag{1.11}$$

If the map is noninvertible, chaos is possible even in one-dimensional maps. Indeed, the logistic map Eq. (1.8) exhibits chaos for large enough r.

It is often useful to reduce a continuous time system (or 'flow') to a discrete time map by a technique called the Poincaré surface of section method. We consider N first-order autonomous ordinary differential equations (Eq. (1.2)). The 'Poincaré map' represents a reduction of the N-dimensional flow to an $(N-1)$-dimensional map. For illustrative purposes, we take $N = 3$ and illustrate the construction in Figure 1.9. Consider a solution of (1.2). Now, choose some appropriate $(N-1)$-dimensional surface (the 'surface of section') in the N-dimensional phase space, and observe the intersections of the orbit with the surface. In Figure 1.9, the surface of section is the plane $x^{(3)} = K$, but we emphasize that in general the choice of the surface can be tailored in a convenient way to the particular problem. Points A and B represent two successive crossings of the surface of section. Point A uniquely determines point B, because A can be used as an initial condition in (1.2) to determine B. Likewise, B uniquely determines A by reversing time in (1.2) and using B as the initial condition. Thus, the Poincaré map in this illustration represents an invertible two-dimensional map transforming the coordinates $(x_n^{(1)}, x_n^{(2)})$ of the nth piercing of the surface of section to the coordinates $(x_{n+1}^{(1)}, x_{n+1}^{(2)})$ at

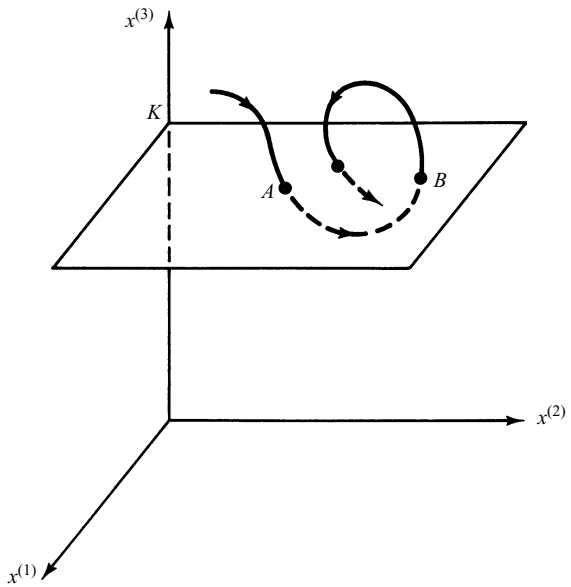

Figure 1.9 A Poincaré surface of section.

piercing $n+1$. This equivalence of an N-dimensional flow with an $(N-1)$-dimensional invertible map shows that the requirements Eq. (1.11) for chaos in a map follows from Eq. (1.5) for chaos in a flow.

Another way to create a map from the flow generated by the system of autonomous differential equations (1.3) is to sample the flow at discrete times $t_n = t_0 + nT$ ($n = 0, 1, 2, \ldots$), where the sampling interval T can be chosen on the basis of convenience. Thus, a continuous time trajectory $\mathbf{x}(t)$ yields a discrete time trajectory $\mathbf{x}_n \equiv \mathbf{x}(t_n)$. The quantity $\mathbf{x}_{n+1}$ is uniquely determined from $\mathbf{x}_n$ since we can use $\mathbf{x}_n$ as an initial condition in Eqs. (1.3) and integrate the equations forward for an amount of time T to determine $\mathbf{x}_{n+1}$. Thus, in principle, we have a map $\mathbf{x}_{n+1} = \mathbf{M}(\mathbf{x}_n)$. We call this map the time T map. The time T map is invertible (like the Poincaré map), since the differential equations (1.3) can be integrated backward in time. Unlike the Poincaré map, the dimensionality of the time T map is the same as that of the flow.

1.4 Attractors

In Hamiltonian systems (cf. Chapter 7) such as arise in Newton's equations for the motion of particles without friction, there are choices of the phase space variables (e.g., the canonically conjugate position and momentum variables) such that phase space volumes are preserved under the time evolution. That is, if we choose an initial ($t = 0$) closed $(N-1)$-dimensional surface S_0 in the N-dimensional $\mathbf{x}$-phase space, and then evolve each point on the surface S_0 forward in time by using them as initial conditions in Eq. (1.3), then the closed surface S_0 evolves to a closed surface S_t at some later time t, and the N-dimensional volumes $V(0)$ of the region enclosed by S_0 and $V(t)$ of the region enclosed by S_t are the same, $V(t) = V(0)$. We call such a volume preserving system *conservative*. On the other hand, if the flow does not preserve volumes, and cannot be made to do so by a change of variables, then we say that the system is *nonconservative*. By the divergence theorem, we have that

$$\mathrm{d}V(t)/\mathrm{d}t = \int_{S_t} \nabla \cdot \mathbf{F} d^N x, \tag{1.12}$$

where $\int_{S_t}$ signifies the integral over the volume interior to the surface S_t, and $\nabla \cdot \mathbf{F} \equiv \sum_{i=1}^{N} \partial F_i(x^{(1)}, \ldots, x^{(N)})/\partial x^{(i)}$. For example, for the forced damped pendulum equation written in first-order autonomous form, Eq. (1.6b), we have that $\nabla \cdot \mathbf{F} = -\nu$, which is independent of the phase space position $\mathbf{x}$ and is negative. From (1.12), we have $\mathrm{d}V(t)/\mathrm{d}t = -\nu V(t)$ so that V decreases exponentially with time, $V(t) = \exp(-\nu t)V(0)$. In general, $\nabla \cdot \mathbf{F}$ will be a function of phase space position $\mathbf{x}$. If $\nabla \cdot \mathbf{F} < 0$ in some region of phase space (signifying volume contraction in that region),