4

Models based on first order difference equations

4.1 General considerations

In this chapter we will consider some of the new dynamic features that occur in models based on first order difference equations

$$\Delta x \equiv x_{n+1} - x_n = G(x_n, c),$$

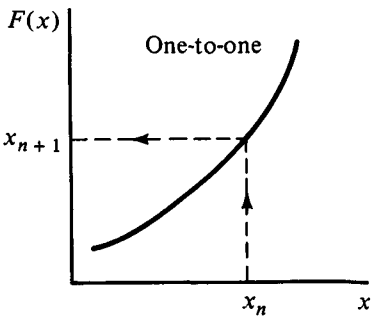
which can be put in the form

$$x_{n+1} = F(x_n, c) \tag{4.1.1}$$

where, again, c are control parameters ($c \in R^k$). In this form, the dynamics can be equally well viewed as a sequence of *mappings*, $F: x_n \to x_{n+1}$.

The function $F(x, c)$ in (4.1.1) is single-valued, which is required of all maps by definition, but may otherwise have a variety of properties all of which represent quite different dynamical behavior, and are related to very different physical systems. Most commonly we take $F(x, c)$ to be a continuous function of x, based on the physical assumption that 'neighboring states' (represented by contiguous points, x) should have essentially the same change in one 'step', and hence essentially the same value of $F(x, c)$. This is not always the case (consider what happens to physical states near catastrophes), but it is the most common case (Fig. 4.1).

Fig. 4.1

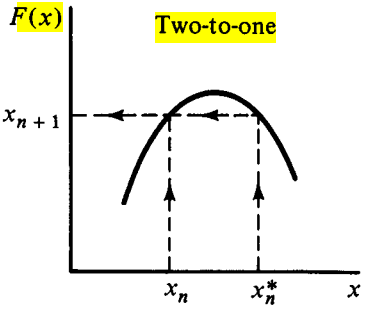


A more common difference which occurs between the dynamics of systems is whether or not they have a 'conservative' or 'dissipative' character to their motion. A typical

class of conservative differential systems are Hamiltonian systems, in which the volume elements in the phase space move in such a way as to 'conserve' their volume (the Liouville theorem). Many other systems can also be conservative, as discussed in Chapter 2, depending on the existence of other conserved (invariant) measures $d\mu = \rho(x)\,dx > 0$. The analogous systems in the case of maps involve functions $F(x, c)$, which 'keep track' of each point in the phase space, by mapping each point to a unique point – in other words, a one-to-one map together with the continuity requirements. Such a map is called a *homeomorphism* (appendix A). While conservative systems must be homeomorphisms, the converse is not always true, as will be illustrated shortly.

The second main category of maps involves those which map two or more phase points onto another phase point – that is, a *many-to-one-map* (Fig. 4.2). It is clear that

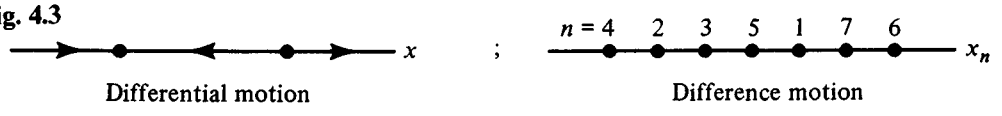
Fig. 4.2



this has a 'compressional' effect on volume elements in the phase space and is, in that sense, 'dissipative' in character. In such systems many states tend to 'settle down' to some select group of states. An example which immediately comes to mind is a pendulum with friction, so that nearly all initial states finally end up at rest, with the pendulum hanging down (except for the single inverted initial state, $\theta = \pi$!). More interestingly, this 'attracting' set of states in the phase space need not simply be at rest (fixed points), but may be a dynamic set. Thus 'dissipation' in the present sense does not necessarily imply that all states 'die', but rather that there is an attraction toward some final, generally dynamic, set of states, called an *attractor*.

The dynamics of equation (4.1.1) is much richer than a first order ODE because it is free from the continuity restrictions of differential equations. x_n can 'jump around' on the real axis, provided that the map is many-to-one, whereas $x(t)$ can only pass a point once, if $\dot{x} = F(x)$ (Fig. 4.3). This freedom makes it possible for (4.1.1) to exhibit several

Fig. 4.3

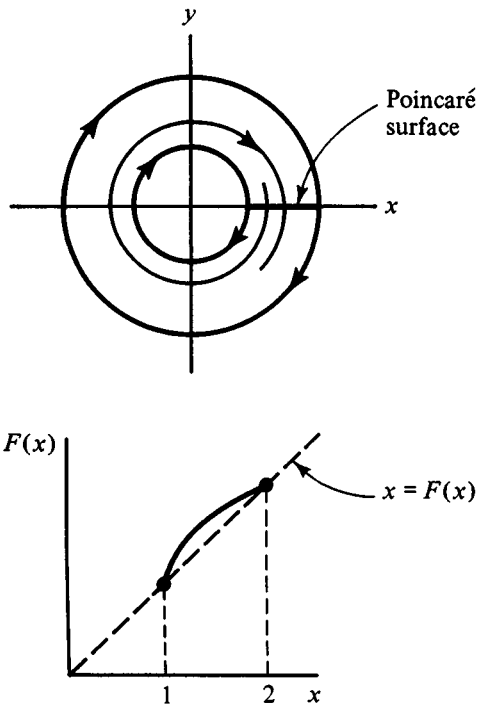


interesting types of bifurcation sequences, leading to various forms of coherent and 'chaotic' behavior, depending on the magnitude of a control parameter c. To obtain a

comparable degree of dynamic richness from physically motivated differential equations requires at least a third order ODE system (see Section 10).

One origin of one-dimensional maps, but certainly not the only possibility (see the logistic map below), is that they are generated by a Poincaré map of a flow in a limited region of a two-dimensional manifold. For example, consider a flow on an annulus, $1 \leqslant x^2 + y^2 \leqslant 4$, with the inside points rotating outward, as illustrated. If we take the surface ($y = 0$, $\dot{y} < 0$) in the annulus as a Poincaré surface, the map for successive x_n points might look as illustrated in Fig. 4.4. The points $x = 1, 2$ are fixed points of the

Fig. 4.4



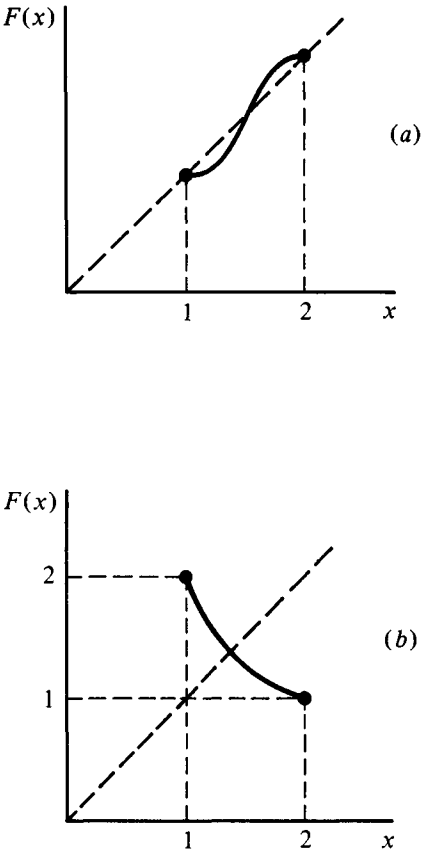
map, corresponding to periodic orbits bounding the annulus. Note that $F(x) > x$, except at the boundaries, so that $x_{n+1} > x_n$, which is the case for outward motion. Also note that, although this map is a homeomorphism, the only 'conserved' (invariant) set of points are the points $x = 1$ or 2. Thus there is no function $\rho(x) > 0$, for all $1 \geqslant x \geqslant 2$, which remains constant under this map, so the system is not conservative in this sense.

Exercise 4.1.

(a) What flow on the above annulus could yield the Poincaré map shown in Fig. 4.5(*a*)?

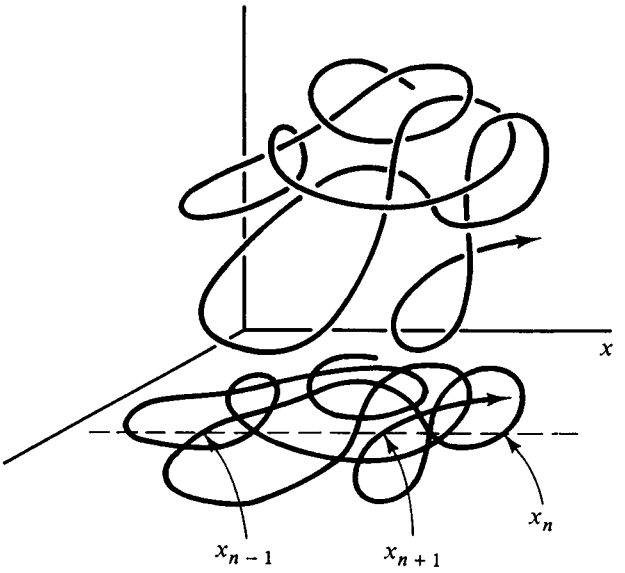
(b) What flow, and on what two-dimensional manifold, could yield the Poincaré map in Fig. 4.5(*b*)?

Fig. 4.5



However, the Poincaré map of autonomous second order systems is of limited interest because such maps must be one-to-one, due to the uniqueness of flows in the R^2 phase space. To the contrary, as we will see, a series of bifurcations which leads to 'chaotic' behavior depends on the nonuniqueness of the inverse of the map (4.1.1). In other words the occurrence of chaotic motion depends on the fact that at least two different initial points x_n, x_n' can map to the same point x_{n+1}. Such many-to-one maps can only occur from some '*projection*' process (see Chapter 1) which reduces a higher order system with unique solutions to a lower order model with nonunique dynamics. This is schematically illustrated in Fig. 4.6, where the unique dynamics in three dimensions is projected onto nonunique dynamics in the (x, y) plane. However, the 'map' of this dynamics along the dashed line need not be of the form (4.1.1). In other words, there is no basic reason for there to be any causal relationship between the values x_n and x_{n+1}, so the function $F(x, c)$ in (4.1.1) need not exist. Put another way, there is generally no reason for a variable, x, to be exactly *self-deterministc*. To take a familiar example: the change in the area of the shadows of tree leaves on the ground is neither unique nor self-deterministic, even for deterministic motion of the leaves. Therefore, usually such maps

Fig. 4.6

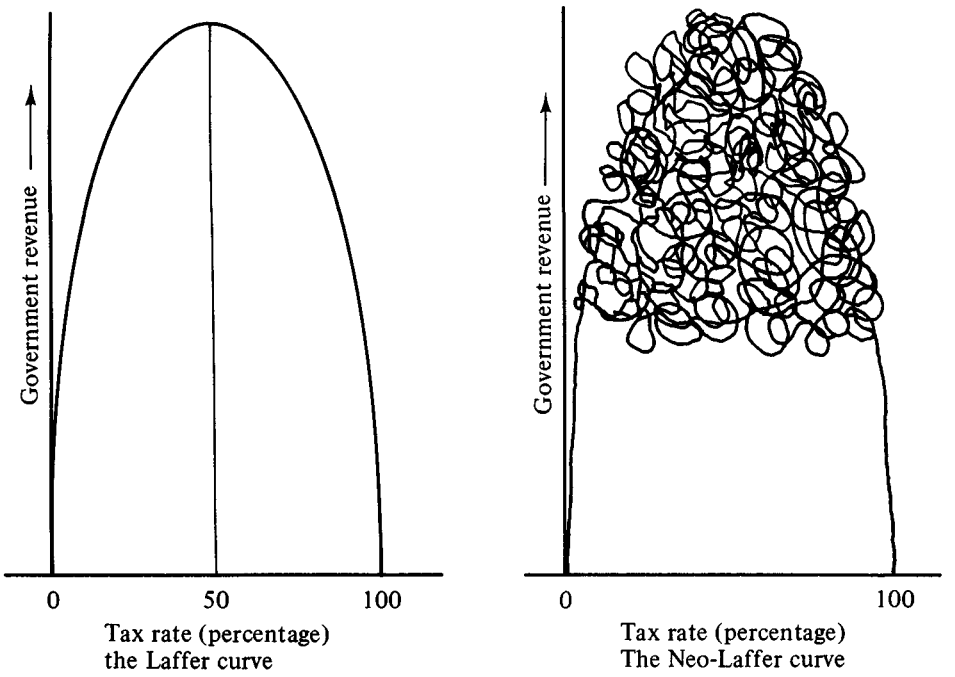


(4.1.1) are rather crude approximations of projections, based on loose arguments and optimistic expectations. This is particularly obvious when the low order maps are intended to represent such complicated systems as population dynamics in biological systems, various economic market dynamics, neurological networks, or even such relatively simple systems as periodically driven nonlinear oscillators.

The fact that complicated systems can not always be reduced to a few interrelated variables is well-illustrated by Fig. 4.7 (after Gardner, 1981, 1986) from the field of economics. The curve on the left (The Laffer Curve) purports to relate the taxation rate to the government's revenue. If there is too much taxation, people will not work as hard for a salary (they will look for other non-cash benefits), so the revenue will drop. Some supply-side economists believe that this curve accurately represents a relationship between taxation and revenue. However, the 'technosnarl' which occurs in the 'neo-Laffer' curve on the right appears to represent much more closely the best available data for the U.S. economy over the past 50 years. Of course, even the neo-Laffer curve is metaphorical. The point here is that there is no map Tax Rate → Government Revenue (except near the trivial end points), because there is no causal connection. Human nature is too complicated for such a connection! What is indeed fortunate, and often amazing, is that some complicated systems appear to have dynamics which can be approximated by simple first order models. While various 'generic' arguments are frequently proposed, we often do not have any profound physical understanding of this good fortune. These points will be discussed in more detail later.

We will find in the following sections that, in addition to the occurrence of various

Fig. 4.7



forms of coherent and chaotic motions, other very important results have been discovered about the dynamics of (4.1.1). One result involves asymptotic ($n \to \infty$) quantitative results (numbers!) which depend only on certain general qualitative features of the function $F(x, c)$. Put another way, quantitative values can also have a form of *structural stability*, since these values remain unchanged when $F(x, c)$ is smoothly varied. One group of asymptotic quantitative features, which has been identified for a large class of functions, $F(x, c)$, has been called '*quantitative universality*' by the discoverer, Feigenbaum (1978, 1979).

This discovery was historically preceded by the observation of *qualitative* '*universal sequences*' generated by an even larger class of functions, $F(x, c)$, and noted by Metropolis, Stein and Stein (1973). They observed that a large class of maps, $F(x, c)$, generate similar qualitative dynamic patterns. An example of such a qualitative pattern is obtained by following the map of the special point x_m, where max $F(x, c) = F(x_m, c)$. A pattern of Ls and Rs is then generated by the mapping sequence of x_m to the L (left) and R (right) of the maximum of $F(x, c)$. Moreover, these qualitative (L, R) patterns changed (bifurcated) in a common 'universal sequence' for a wide class of $F(x, c)$, as the control parameter c is increased. These 'universal' features', which have also been observed in a number of experimental situations, will be discussed in Section 3.

4.2 Two-to-one maps: the logistic map

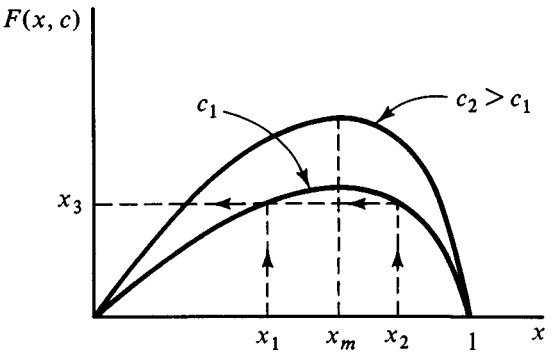
A classic example of (4.1.1) is the so-called *logistic equation*, the discrete form of the relatively unexciting logistic differential equation noted before. The logistic map is

$$x_{n+1} = cx_n(1 - x_n) \quad (0 \leqslant x_n \leqslant 1, 0 < c \leqslant 4). \tag{4.2.1}$$

In biological examples of such a map, x_n represents the population of some species in the n'th generation, normalized by the maximum population which the species can achieve due to mutual toxicity, or to competition for nutrients.

The essential feature of the present class of models, of which (4.2.1) is an example, is that $F(x, c)$ has the (x, c) dependence indicated in Fig. 4.8. More specifically, $F(x)$ has

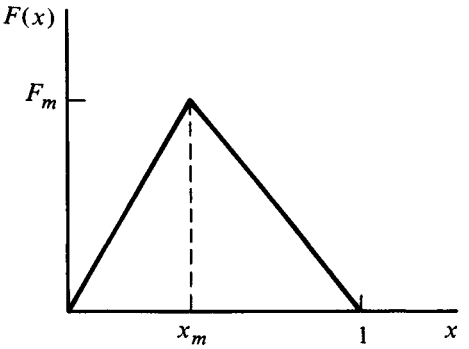
Fig. 4.8



only one maximum $F(x_m) \leqslant 1$ and vanishes at the end points $F(0) = F(1) = 0$. Thus this generally maps the interval [0, 1] into a smaller region, unless $F(x_m) = 1$. $F(x)$ also maps two points into one point (e.g., x_1 and x_2 map into x_3), so it is a two-to-one map. If $F(x)$ has certain properties of smoothness (differentiability, etc.), then the variety of dynamics is correspondingly limited.

One important class of maps involves maps of an interval into itself, F:[0, 1], where $F(x)$ has only one maximum, x_m, and has everywhere a negative '*Schwarzian derivative*' (see Appendix D). This requires that $|\partial F/\partial x|^{-1/2}$ has a positive second derivative on

Fig. 4.9



both sides of the maximum (not in intuitive condition!). We will call maps with everywhere negative Schwarzian derivatives *S-maps*. In this case, Singer (1978) has proved that, for each value of c, there is at most one stable periodic solution of (4.1.1) and, if it exists, the iterates of x_m tend toward this periodic solution (i.e., x_m is in the 'basin of attraction' of the periodic solution). The logistic map, (4.2.1), is an example of a S-map (Appendix D), and Singer's result will simplify our analysis of this dynamics.

There are other two-to-one maps (Fig. 4.9) which are frequently considered, and which are not S-maps. An important example is the so-called 'tent map' (sometimes a leaning tent!

$$F(x) = a(1 - x_m)x \quad (x \leqslant x_m),$$

$$F(x) = ax_m(1 - x) \quad (x_m \leqslant x \leqslant 1), \tag{4.2.2}$$

and $ax_m(1 - x_m) \leqslant 1$. This has a discontinuous derivative at $x = x_m$, and is not an S-map. It may have more than one stable periodic solution. Guckenheimer (1979) has obtained the conditions for which (4.2.1) is topologically conjugate to some tent map (4.2.2) (see Exercise 4.8).

Before considering any details it should be emphasized that (4.2.1) is a mapping of the interval $0 \leqslant x \leqslant 1$ into itself, which is two-to-one, and this gives rise to many of the interesting features noted in Section 1. As discussed in that section, while the mapping dynamics is deterministic, it is not unique because it is a two-to-one mapping. Hopefully the example of the biological origin of the logistic map makes it reasonable that both determinism and non-uniqueness may occur (approximately) in some observables, even if there is an underlying unique dynamics in all the variables. It is the self-determinism of this nonunique dynamics which is the most profound aspect of this 'projection'. Generally, of course, this simple one-dimensional determinism does not occur, as has been pointed out in Section 1.

We will now consider some of the properties of the general two-to-one maps frequently illustrating these properties with the logistic map (4.2.1). For a very nice (now classic) introductory article on this subject, see May (1976). We first introduce the notation for a sequence of mappings of some point x,

$$F^0(x) = x, \quad F^{n+1}(x) = F(F^n(x)). \tag{4.2.3}$$

A point is said to be a *fixed point* of a map if

$$F(x) = x \quad \text{(fixed point)}$$

and it is called a *period-n point* ($n = 0, 1, 2, \ldots$), if

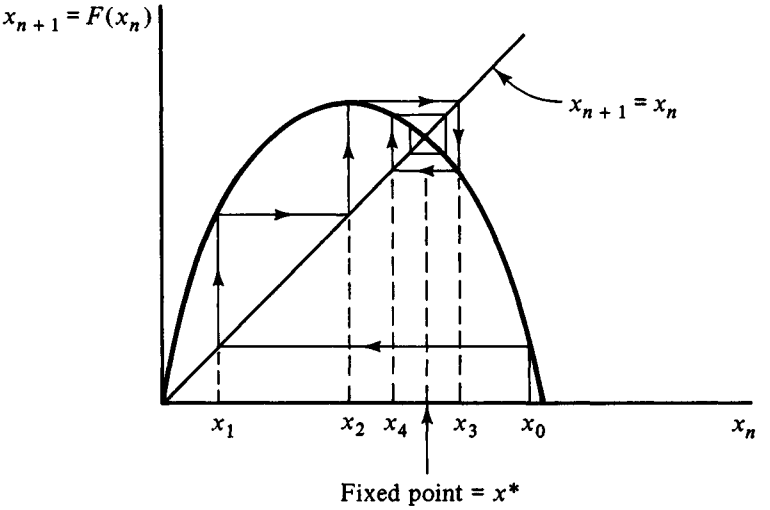
$$F^n(x) = x \quad \text{and} \quad F^k(x) \neq x \quad (\text{for } k < n). \tag{4.2.4}$$

A fixed point is simply a period-one point. If there is one period-n point, then there are n of them.

Exercise 4.2. Prove the last statement.

The sequence of values of x_n which are generated by the map can be obtained in a simple fashion by the following graphical construction. On a graph of $F(x)$ vs. x we also plot the straight line $x_{n+1} = x_n$. Starting at some point x_0, a vertical path is taken to obtain $F(x_0)$, and then a horizontal until it intersects the straight line. This intersection point occurs at x_1. The second map (of x_0) is then obtained by repeating this procedure, beginning at x_1, which then yields x_2. This is illustrated in Fig. 4.10.

Fig. 4.10



This figure also illustrates a sequence of maps converging to a fixed point, x^*. Note that all *fixed points* of a map are graphically represented by the intersection of the straight line with the function $F(x)$.

Exercise 4.3. This method can be used to establish a special case of *Brouwer's fixed point theorem*, which holds in R^n. Prove that, for any continuous map, $F(x)$, which takes the interval $[0, 1]$ into itself (so $0 \leqslant F(x) \leqslant 1$) there must be at least one fixed point $x^* = F(x^*)$.

In the above illustration the fixed point is stable, since nearby points converge towards it. A fixed point, x^*, is stable if

$$|dF(x)/dx|_{x^*} < 1 \quad \text{(stable)}. \tag{4.2.5}$$

This can be shown by taking an initial point nearby, $x_0 = x^* + \Delta$, and expanding $F(x_0)$ for small Δ,

$$F(x^* + \Delta) = F(x^*) + \Delta \quad (dF/dx)_{x^*} = x^* + \Delta \quad (dF/dx)_{x^*} = x_1,$$

where we have assumed that $(\mathrm{d}F/\mathrm{d}x)_{x^*} \neq 0$ (see superstable fixed points, below). Therefore the distance to the fixed point is decreased by the map, $|x_1 - x^*| < |x_0 - x^*|$, if (4.2.5) is satisfied. In the case of the logistic map (4.2.1), the fixed point is given by

$$x^* = cx^*(1 - x^*) \quad \text{or} \quad x^* = 1 - c^{-1} \quad (\text{if } c > 1), \tag{4.2.6}$$

ignoring the trivial fixed point $x^* = 0$, which is the only one if $c < 1$. Using the condition (4.2.5), we find that the fixed point (4.2.6) is stable provided that

$$|c(1 - 2x^*)| < 1 \quad \text{or} \quad 1 < c < 3 \equiv c_1. \tag{4.2.7}$$

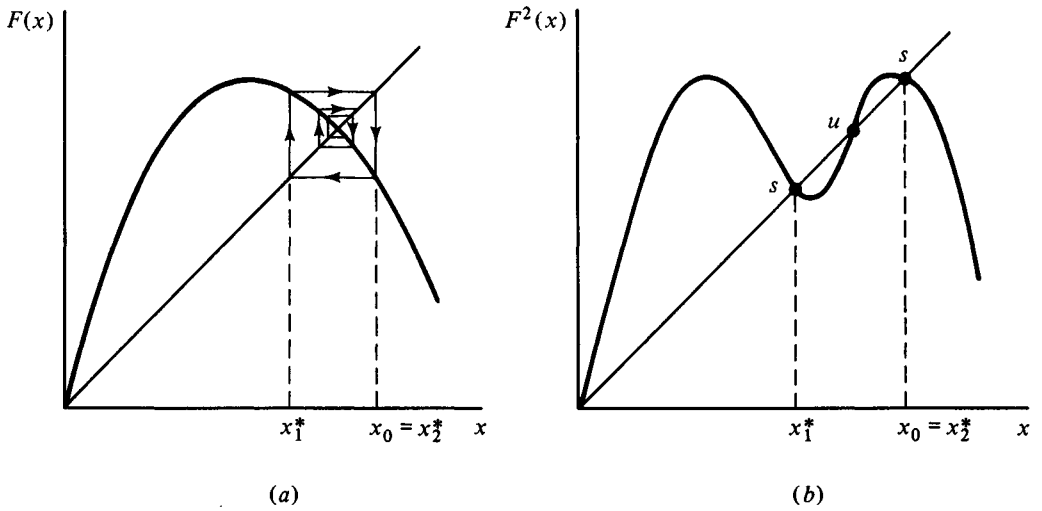
A curious feature of maps is that there can be points which are *eventually periodic*. The non-periodic point x_e is said to be eventually periodic if $x = F^m(x_e)$ is a periodic point for some finite m. These points can exist because $F(x)$ is not one-to-one. For example, for the logistic map, with $c = 3$, $x_e = \frac{1}{3}$ is eventually periodic when $m = 1$, and $x_e = \frac{1}{2} \pm (5^{1/2}/6)$ are both eventually periodic for $m = 2$.

Exercise 4.4. What are the eventually periodic points for the logistic map ($c = 3$) with $m = 3$? Beware of complex numbers!

Exercise 4.5. Another strange feature of these maps is that there are an infinite number of points that map into an unstable fixed point (try to show this using the inverse of the above graphical method).

As c is increased past the value $c_1 = 3$ the fixed point becomes unstable and points are then mapped towards two other points, x_1^* and x_2^*, as shown in Fig. 4.11(*a*). These two

Fig. 4.11



points are period-two points of $F(x)$, that is

$$F(x_1^*) = x_2^* \quad \text{and} \quad F(x_2^*) = x_1^*.$$

This means that after two maps their values are repeated, or

$$F^2(x_k^*) = F(F(x_k^*)) = x_k^* \quad (k = 1, 2). \tag{4.2.8}$$

Therefore, when c passes through the value c_1, the asymptotic limit ($n \to \infty$) of most points changes from a period-one state ($\lim_{n\to\infty} x_n = x^*$) to a period-two state ($\lim_{n\to\infty} x_n \in \{x_1^*, x_2^*\}$). This topological change in the asymptotic states therefore represents a *bifurcation* at $c = c_1$ (extending the concept of bifurcations to maps in a natural way). It can be seen from (4.2.8) that period-two points can be viewed as fixed points of the map $F^2(x)$. The graphical representation of this is shown in Fig. 4.11(*b*). All of the intersections of $F^2(x)$ with the straight line are fixed points, but the central point is unstable, whereas the outer fixed points are stable provided that

$$|\mathrm{d}F^2(x)/\mathrm{d}x|_{x_k^*} < 1 \quad (k = 1, 2). \tag{4.2.9}$$

This follows in the same way that (4.2.5) was established.

It is important now to establish the fact that if one of these fixed points goes unstable when c is increased, then the other fixed point also goes unstable. Indeed, what we will show is that the derivative of F^2 is the same at both fixed points,

$$(\mathrm{d}F^2(x)/\mathrm{d}x)_{x_1^*} = (\mathrm{d}F^2(x)/\mathrm{d}x)_{x_2^*} \tag{4.2.10}$$

To show this we simply need to use the chain rule of differentiation. Consider any point x_0. Then

$$F(x_0) = x_1 \quad \text{and} \quad F(x_1) = x_2, \quad \text{or } F^2(x_0) = x_2.$$

Therefore

$$\begin{aligned}(\mathrm{d}F^2/\mathrm{d}x_0) &= (\mathrm{d}x_2/\mathrm{d}x_0) = (\mathrm{d}F(x_1)/\mathrm{d}x_1)(\mathrm{d}x_1/\mathrm{d}x_0) \\ &= (\mathrm{d}F(x_1)/\mathrm{d}x_1)(\mathrm{d}F(x_0)/\mathrm{d}x_0)\end{aligned}$$

It is obvious that this can be generalized for any n to read

$$\mathrm{d}F^n(x_0)/\mathrm{d}x_0 = \prod_{k=0}^{n-1} \mathrm{d}F(x_k)/\mathrm{d}x_k. \tag{4.2.11}$$

Now we specialize this result to the case where the set of points $\{x_0, x_1, \ldots, x_{n-1}\}$ are period-n points of $F(x)$ or, in other words, fixed points of $F^n(x)$. A little thought makes it clear that, regardless of which fixed point is used in the left side of (4.2.11) (the one we called 'x_0'), the derivatives on the right are always taken at the same set of points, namely the period-n points of $F(x)$. Since the derivatives on the right are always taken at the same set of points, the values on the left are all equal. Thus we have proved that, for

any two period-n points of F^n, x_k^* and x_l^*,

$$(\mathrm{d}F^n/\mathrm{d}x)_{x_k^*} = (\mathrm{d}F^n/\mathrm{d}x)_{x_l^*} \tag{4.2.12}$$

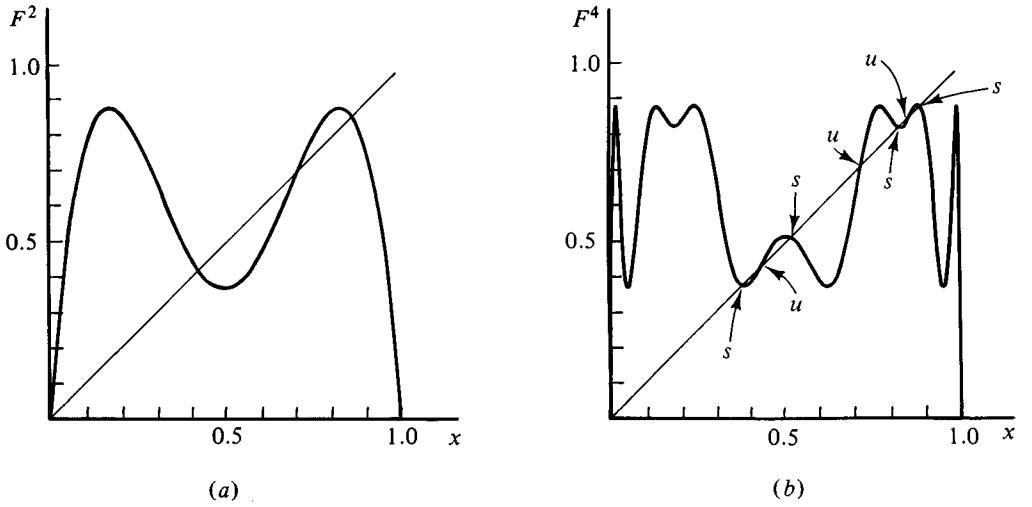
Since the fixed points of F^n are stable if and only if

$$|\mathrm{d}F^n/\mathrm{d}x|_{x^*} < 1 \quad \text{(stable fixed point)} \tag{4.2.13}$$

(by using the same reasoning as in (4.2.9)), the result (4.2.12) shows that this set of n fixed points are either all stable or all unstable. Note that F^n has other fixed points, but they do not belong to the period-n set of $F(x)$. This result is of basic importance in establishing the character of a sequence of bifurcations which are exhibited by such maps, as we will now discover.

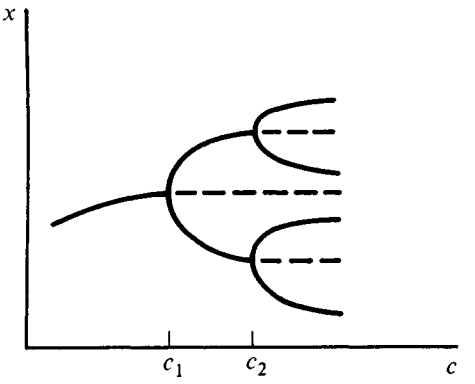
We now return to F^2 and consider what happens as c is further increased. The function $F^2(x)$ develops larger maxima and a smaller minimum, and its slope at the fixed points thereby increases. Finally a value of c is reached (call it c_2) where the condition (4.2.9) is no longer satisfied at the fixed points, and they become unstable. This is illustrated in Fig. 4.12(a). Now F^2 has three fixed points, but they are unstable. On the other hand, the two fixed points which become unstable at $c = c_2$ 'give birth' to a stable cycle of period $2^2 = 4$. This is made fairly clear by considering the graph of $F^4(x)$

Fig. 4.12



(Fig. 4.12(b)). The two outside unstable fixed points have each 'bifurcated' two stable fixed points of F^4 (that is, the condition (4.2.13) is satisfied at each of the points numbered on the x-axis). In the case of the logistic map, this bifurcation occurs at the value $c_2 = 1 + 6^{1/2} \simeq 3.45$. In the control-phase space these bifurcations look like Fig. 4.13, where the dashed portions are unstable fixed points of F^4. We recognize here

Fig. 4.13



the '*pitch-fork*' (*double point*) bifurcation structure and the *exchange of stability feature* of F^n, found in the last chapter in a different context. Here, however, we find a much more interesting, indeed exciting, bifurcation phenomena, which Thom (1969) viewed as an example of what he called a '*generalized catastrophe*'. As c increases, the above period-four set becomes unstable, giving birth to a stable period-eight, and this process of period doubling repeats indefinitely. That is, there is a set, $\{c_n\}$, such that

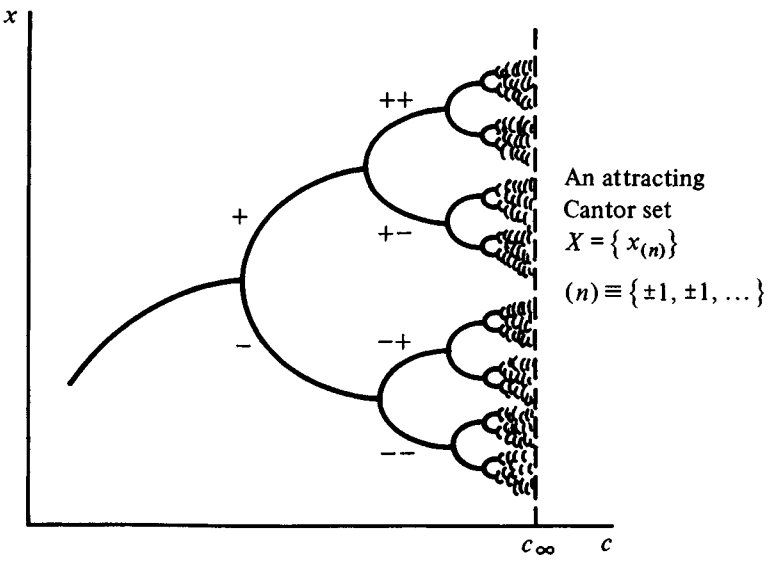
$$\text{if } c_{n+1} > c \geqslant c_n, \text{ there is a stable period-}2^n \text{ cycle} \tag{4.2.14}$$

and, moreover, it turns out that

$$\lim_{n\to\infty} c_n \equiv c_\infty \text{ (finite).} \tag{4.2.15}$$

In the control-phase space this *infinite* set of *period doubling bifurcations* looks something like that shown in Fig. 4.14. We have sequentially labeled the branches $\pm$,

Fig. 4.14



so that each bifurcation point, $x_{(n)}$ can be associated with its branch by a '$\pm$' subscript (e.g., $(n) = \{+, +, -, +, -, -\}$).

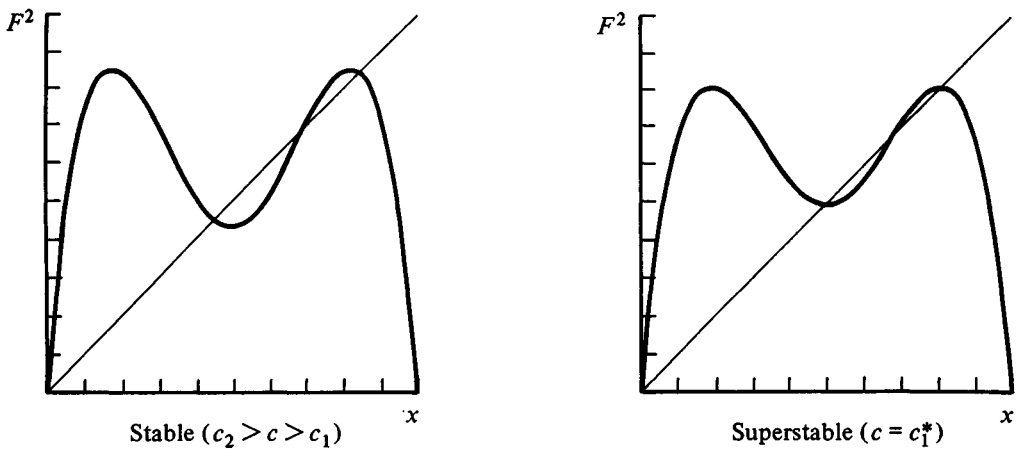
The infinite set, $X = \{x_{(n)}\}$, stable period-2^n ($n \to \infty$) points obtained as $c \to c_\infty$ is interesting to study, since it is an *attractor* of most of the points in the interval (0, 1). It, of course, does not attract the unstable periodic points. The set X is nowhere dense because between each two stable points must be an unstable point. Therefore X is nowhere a continuum, and in fact it is a Cantor set (see Chapter 2, Section 6, and Appendix A). We might ask if it has a dimension (capacity or information) greater than zero. A nice method of obtaining the dimension of such sets has been devised by Chang and McCowan (1984), and will be discussed in Section 7.

The values of c_n, where the period-2^n cycle first occurs (equation (4.2.14)) depend, of course, on the function $F(x, c)$, as does the limiting value c_∞ (4.2.15). The following values of c_n can be obtained for the logistic map.

n	c_n	n	c_n
1	3	5	3.568 759
2	3.449 499	6	3.569 692
3	3.544 090	7	3.569 891
4	3.564 407	8	3.569 934
		∞	3.569 946

These bifurcation values of c are not easy to obtain by any means except numerical methods. There is, however, another set of c values, which we will call $\{c_k^*\}$, that are more accessible and contain the same topological information about the map $x_{n+1} = F(x_n, c)$ as the set of bifurcation values $\{c_n\}$.

Fig. 4.15



We need first to introduce the concept of a *superstable* 2^n-cycle. Fig. 4.15 illustrates the distinction between a stable 2-cycle and a superstable 2-cycle. Here we will assume that $F(x)$ is smooth (see Section 3 for some generalizations). In the case of a superstable 2-cycle the stable fixed points of F^2 occur at the maxima, and minima of F^2. It should also be obvious that $c_2 > c_2^* > c_1$. We know from the general result (4.2.12) that if one of these fixed points (corresponding to period-2 points of F) occurs at an extrema of F^2, then so must the other fixed point (i.e., $(dF^2/dx)_{x_1^*} = 0 = (dF^2/dx)_{x_2^*}$). The difference in the sequence of points generated by $F^2(x)$ in the two cases is illustrated above. In the superstable case a point near the fixed point, $x_0 = x^* + \Delta$, maps to $x_1 = x^* + (1/2)(\Delta)^2(d^2F^2/dx^2)_{x^*}$, which is now *quadratically* close to x^*, since it differs by Δ^2. This is a much more rapid approach to x^* than a linear dependence, Δ. The fact that the superstable fixed points occur at the extrema of F^2 (or, more generally, F^{2^n}) makes it easier to determine their associated values of c. We therefore define these values, c_n^*, by this condition

$$c = c_n^* \quad \text{if} \quad (dF^{2^n}/dx)_{x^*} = 0 \quad (x^*\text{: fixed point of } F^{2^n}). \tag{4.2.16}$$

Let us see how this works using the logistic map,

$$F(x) = cx(1 - x).$$

It follows from (4.2.11) that x_m is also an extremum of $F^{2^n}(x)$ by taking $x_0 \equiv x_m$. If c is such that x_m is also a fixed point of $F^{2^n}(x)$ then, according to the definition (4.2.16), this value of $c = c_n^*$ corresponds to a superstable period-2^n cycle. We conclude that if x_m, (4.2.17), is a period-2^n point

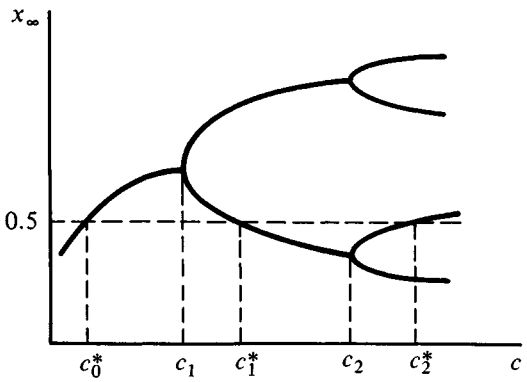
$$F^{2^n}(x_m, c) = x_m, \tag{4.2.18}$$

then $c = c_n^*$, so this is a superstable period-2^n.

Let us see how this works using the logistic map,

$$F(x) = cx(1 - x),$$

Fig. 4.16



for which $x_m = \frac{1}{2}$. This is also a fixed point of F provided that

$$c\tfrac{1}{2}(1-\tfrac{1}{2}) = \tfrac{1}{2} \quad \text{or} \quad c_0^* = 2.$$

Next, we have

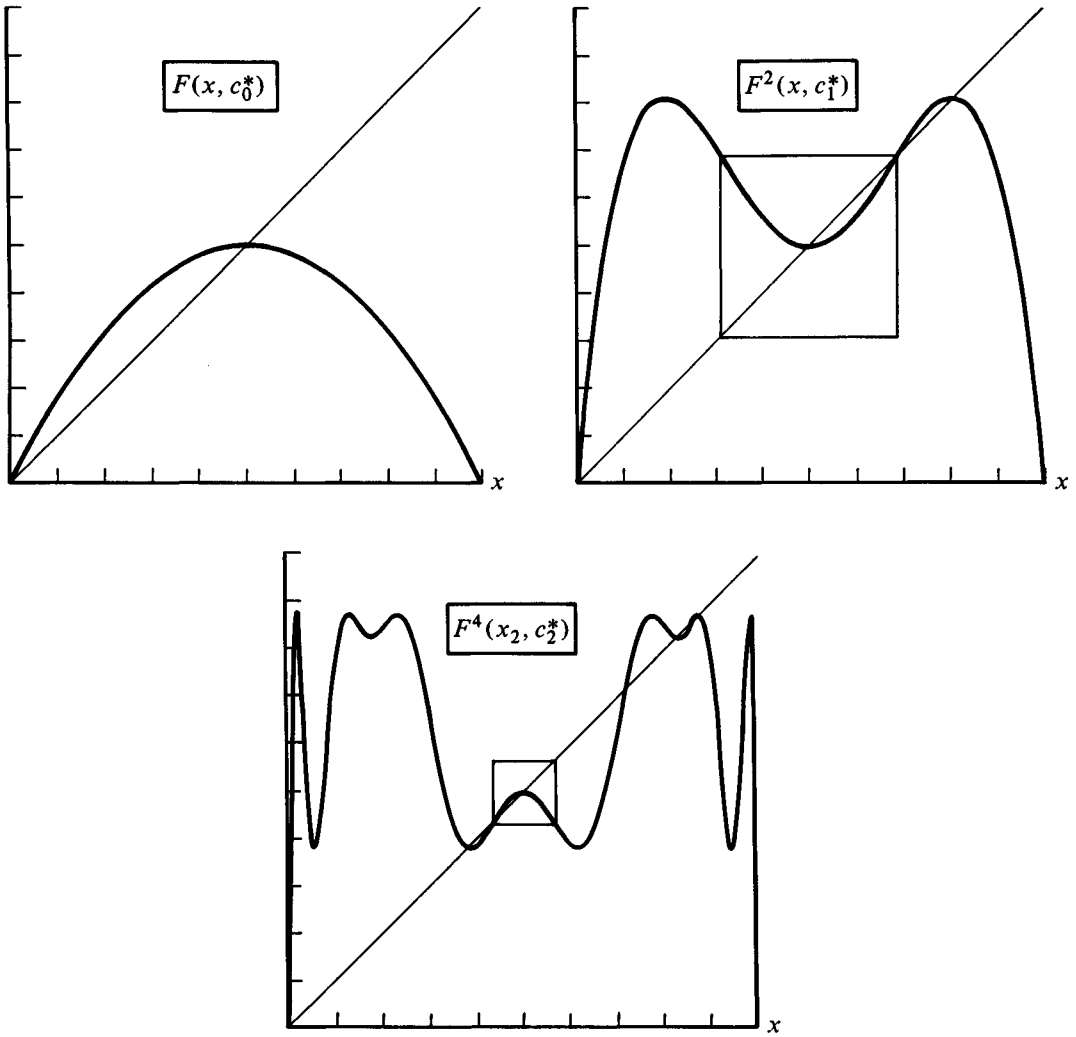
$$F^2(x) = c[cx(1-x)](1-[cx(1-x)]), \tag{4.2.19}$$

and $F^2(x_m) = x_m$ if

$$c^2(4-c) = 8$$

or $c_1^* \simeq 3.2361$ (Fig. 4.16). One can proceed in this fashion and obtain the following values for the period 2^n superstable control parameters.

Fig. 4.17



n	c_n^*	n	c_n^*
0	2.00	5	3.569 244
1	3.236 068	6	3.569 793
2	3.498 562	7	3.569 913
3	3.554 641	8	3.569 939
4	3.566 667	∞	3.569 946

The first few functions $F^{2^n}(x, c_n^*)$ are shown in Fig. 4.17 (for $n = 0, 1, 2$). We note that the functions $F^{2^n}(x_1, c_n^*)$ all look similar in the regions around $x = 0.5$, except for a change in scales (and inversions). This is illustrated by the boxes in Fig. 4.17. Moreover, the asymptotic features ($n \to \infty$) in these boxes would not be changed even if $F(x, c)$ looked quite different away from the point $x = 0.5$. This leads to the consideration of possible 'universal features' and scaling properties.

4.3 Universal sequences and scalings

Metropolis, Stein and Stein (1973) were the first to consider superstable c values and discovered that a wide class of two-to-one maps of an interval onto itself all had a *qualitative feature* in common, which they called *U-sequences* (*u*niversal). They considered maps

$$x_{n+1} = c f(x_n) \equiv F(x_n, c) \quad (0 \leqslant c \leqslant c_m) \tag{4.3.1}$$

where $f(x)$ is continuous and continuously differentiable, with a single maximum at $x = x_m$. Note that the control parameter is multiplicative in (4.3.1) with some bound c_m. It has been known for a long time that, if $f(z)$ is analytic (in the complex plane), then any periodic solution of this map contains the point x_m within its region of attraction. Metropolis, Stein and Stein (MSS) found this to also be true for much more general functions $f(x)$ and, in particular, more general than the S-maps discussed above. They considered those values of c such that the point x_m is a periodic point of $F^p(x, c)$. In other words, they considered the superstable $c = \bar{c}_p$ defined by

$$x_m = F^p(x_m, \bar{c}_p), \tag{4.3.2}$$

where now p need not be 2^n. Note that c_n^* of (4.2.18) equals $\bar{c}_{2^n}$. The *number of* $\bar{c}_p$ values, for different values of p, they found to be 1 (for $p = 2$), 1(3), 2(4), 3(5), 5(6), 9(7), 16(8), 28(9), etc.(!), for a total of 2370 $\bar{c}_p$ values for $p \leqslant 15$. These results were obtained from extensive computer studies. They considered the set of points generated by one of these ($\bar{c}_p$) maps, $x_1 = F(x_m, \bar{c}_p)$, $x_2 = F^2(x_m, \bar{c}_p), \ldots$, starting from x_m, and considered whether $x_k > x_m$ (*a* '*R*' point) or $x_k < x_m$ (a '*L*' point). Then any periodic set is characterized by

some pattern, such as

$$x_m \to R \to L \to L \to L \to R \to R \to x_m \quad (p = 7) \tag{4.3.3}$$

which contains $(p-1)$ terms (R, L). They represented such a pattern with the obvious notation RL^3R^2. What they discovered was that the sequence of these patterns (and hence the ordering of the periods p), as c is increased through the values contained in the set $\{\bar{c}_p\}$, is independent of the details of the function $f(x)$, within a large class of functions. The values of the $\bar{c}_p$ of course do depend on the function $f(x)$, but not the sequence of these patterns. This sequence they called the *U-sequence*. For periods of seven or less, the first half of this sequence is

$$\begin{aligned} &R, RLR, RLR^3, RLR^4, RLR^2, RLR^2LR, \\ &\quad RL, RL^2RL, RL^2RLR, RL^2R, \ldots (+\text{eleven more}), \end{aligned} \tag{4.3.4}$$

where $p = (1 + \text{sum of the exponents}) \leqslant 7$. For the logistic equation $(x_m = \frac{1}{2})$, the corresponding $\bar{c}_p$ values for the patterns (4.3.4) are

$$\bar{c}_p = 3.236, 3.499, 3.628, 3.702, 3.739, 3.774, 3.832, 3.845, 3.886, 3.906, \ldots. \tag{4.3.5}$$

For the first three patterns above, the actual ordering of the iterates along the x-axis for the logistic map (taking $x_1 = 0.5$) is

$$x_1 < x_2; \quad x_3 < x_1 < x_4 < x_2; \quad x_3 < x_1 < x_5 < x_4 < x_6 < x_2.$$

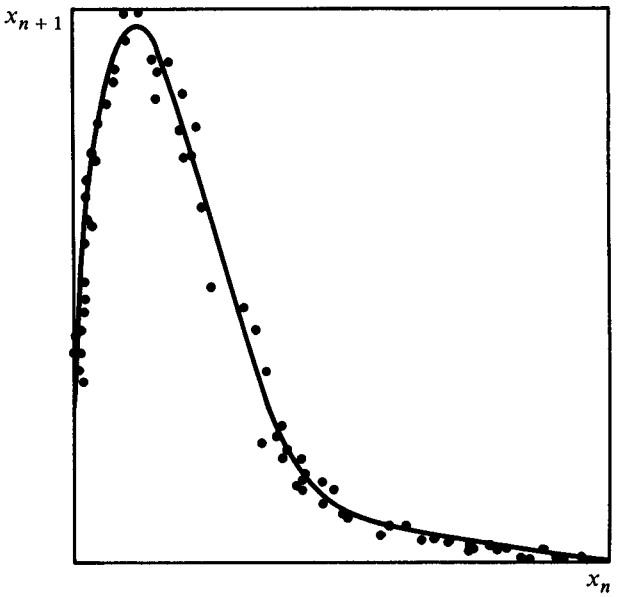
A given period can occur with different patterns, at different values of c. Thus, for $p = 5$, the only possible patterns are RLR^2, RL^2R, and RL^3. For a more exhaustive (and exhausting) list of all the 209 patterns through period eleven, see MSS.

Such patterns, and short sequences of such patterns, have now been observed in a number of experimental investigations. Many of these observations are related to aperiodic dynamics in chemical and solid state systems, and the two states (R and L above) may refer, for example, to large and small amplitude outputs from the systems. Then, as some control parameter is varied (e.g., concentrations, flow rates, voltages, etc.) a limited sequence of such patterns is observed.

As an example, Fig. 4.18 shows a 'map' which is experimentally related to the dynamics of the bromide concentration in a well-stirred Belousov–Zhabotinskii reaction (Simoyi, Wolf and Swinney 1982). This is a discussed further in Chapter 8. We simply note here that, as they varied the flow rate of the chemical reactants, the sequence of periodic states which they observed was found to be:

Observed:	R;	RLR;	RLR^3LR;	RLR^3LRLR;	RLR^3;	RLR^2
MSS *index*:	1;	2;	3;	4;	5;	16,
Observed:	RL;	RL^2RL;	RL^2RLR^2L;	RL^2R;	RL^2;	RL^3RL
MSS *index*:	29;	30;	31;	51;	94;	103

Fig. 4.18



where R and L are right and left of the maximum of the above 'map'. Here the MSS 'index' refers to the placement of the periodic patterns in the MSS listing, given in an appendix of their study. Note that, while many members of the U-sequence were not detected (for a variety of experimental reasons), none of the observed patterns occurred out of the order given by MSS.

This 'universal' *qualitative* discovery by MSS was followed by a 'universal' *quantitative* discovery by Feigenbaum (1978), again with the aid of computer calculations. He considered various subsets of the set $\{\bar{c}_p\}$ introduced by MSS, one of which is (4.2.18), that is $p = 2^n$. What is particularly useful about the set $\{c_n^*\}$ is that it makes a scaling property quite obvious. As can be seen from Fig. 4.19, if one concentrates on the region around x_m (here $x_m = \frac{1}{2}$), the sequence of functions $F(x, c_0^*)$, $F^2(x, c_1^*)$, $F^4(x, c_2^*)$ (etc.) all look similar except for an inversion process and a *scale reduction.* The scale reduction occurs both in the range of x and also in the magnitude of the variation of $F^{2^n}(x, c_n^*)$ over that (boxed) range of x. To represent this scaling, it is useful to center the origin of these figures at $(x = \frac{1}{2}, F = \frac{1}{2})$, so we set

$$y = x - 1/2 \tag{4.3.6}$$

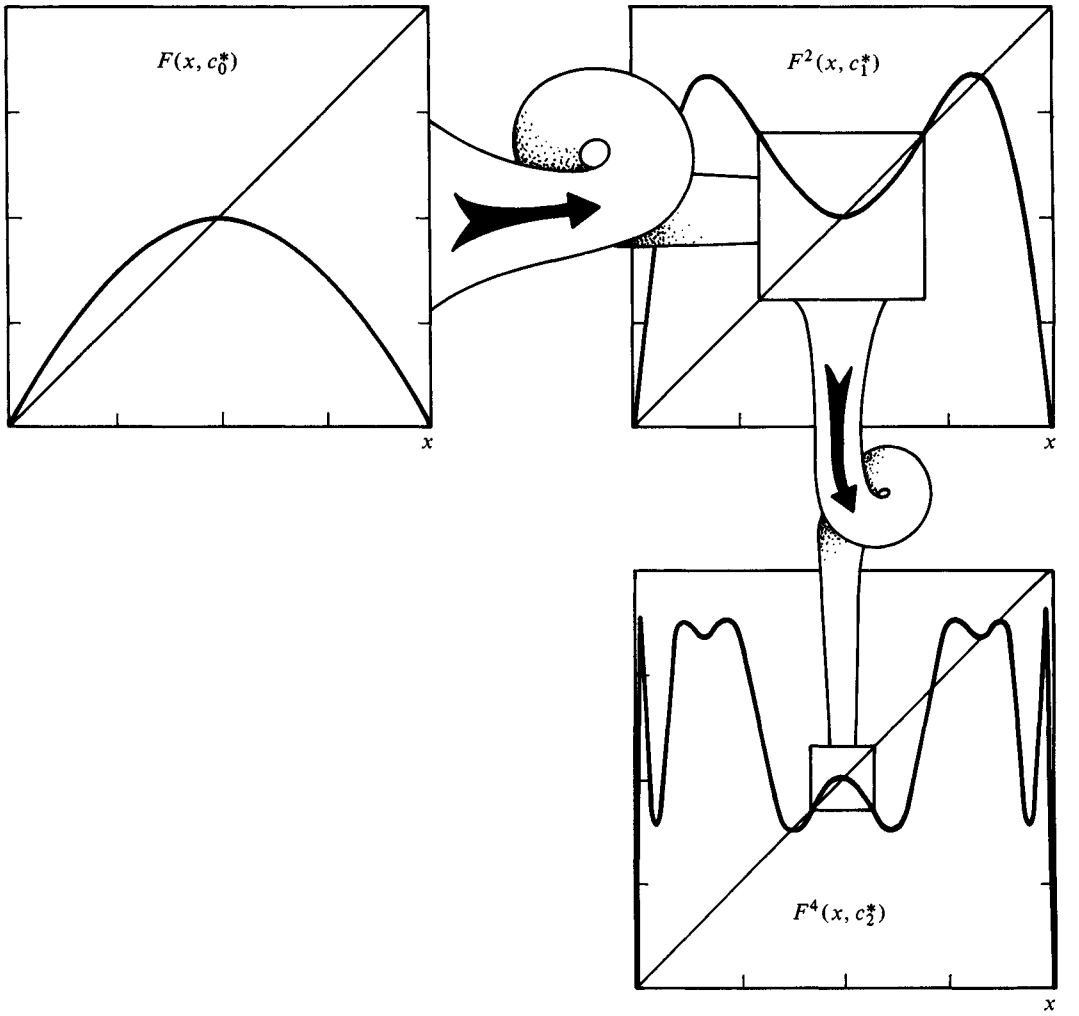
$$G(y, c) \equiv F(y + 1/2, c) - \tfrac{1}{2}$$

We can see from (Fig. 4.20) that (roughly, at first)

$$G(y, c_0^*) \simeq -2.5\, G^2\left(-\frac{y}{2.5}, c_1^*\right)$$

$$G^2(y, c_1^*) \simeq -(2.5)^2 G^4\left(-\frac{y}{(2.5)^2}, c_2^*\right)$$

Fig. 4.19



or, more generally, this suggests that there is some scaling parameter (call it α) such that

$$(-\alpha)^n G^{2^n}(y(-\alpha)^{-n}, c_n^*)$$

goes to some universal function of y as $n \to \infty$, call it $g_0(y)$, if α is selected correctly. Feigenbaum found that, for

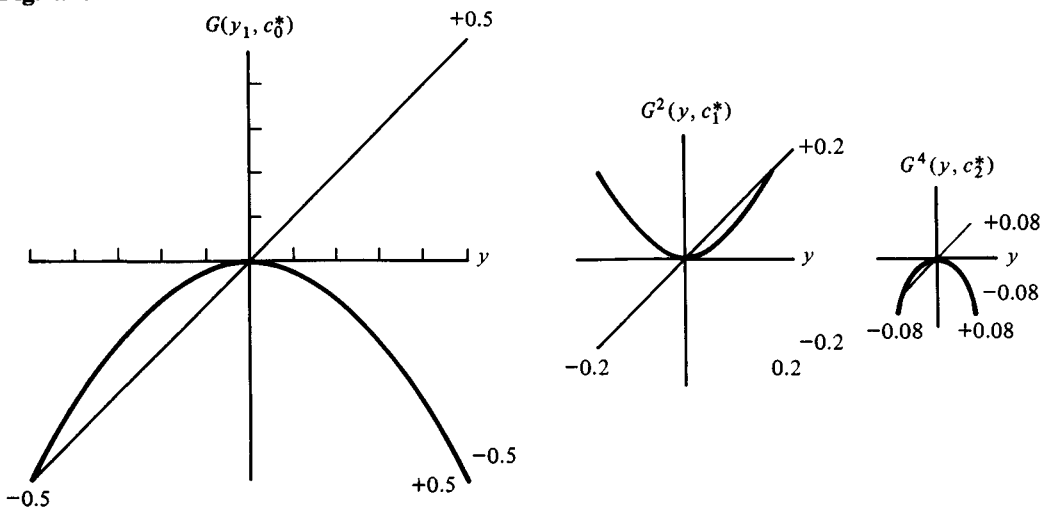
$$\alpha = 2.502\,907\,875\,095\,892\,848\,5\ldots \tag{4.3.7}$$

this indeed happens,

$$\lim_{n\to\infty} (-\alpha)^n G^{2^n}(y(-\alpha)^{-n}, c_n^*) \equiv g_0(y). \tag{4.3.8}$$

More important than all the significant figures in (4.3.7), Feigenbaum found that both α

Fig. 4.20



and the function $g_0(y)$ are independent of the precise form of $G(y)$, provided only that it has a single quadratic maximum at $y = 0$. For example, $F(x, c) = c \sin(\pi x)$, $(0 \leqslant x \leqslant 1)$, or $G(y, c) = c \sin(\pi(y + \frac{1}{2})) - \frac{1}{2}$ has the same asymptotic scaling. In other words α *is a universal scaling*, and $g_0(y)$ *is a universal function.* [See Feigenbaum for a discussion of a family of universal functions.]

Feigenbaum also found from his computer studies that

$$\lim_{n \to \infty} \frac{c_{n+1}^* - c_n^*}{c_{n+2}^* - c_{n+1}^*} = \delta \tag{4.3.9}$$

where

$$\delta = 4.669\,201\,609\,102\,9\ldots \text{(etc!)} \tag{4.3.10}$$

is again a *universal constant*. Equation (4.3.9) can also be written

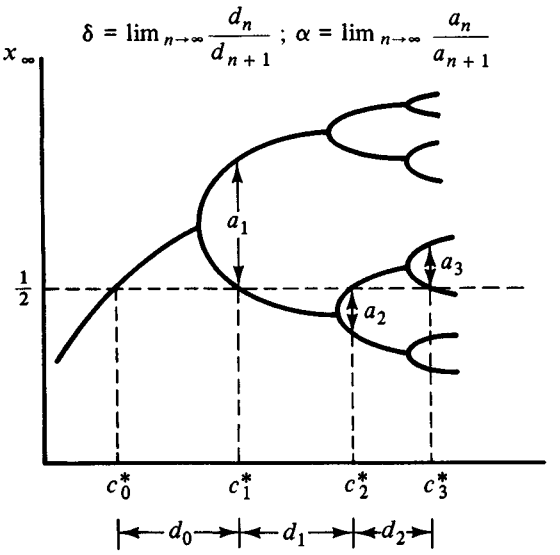
$$c_n^* = c_\infty^* - A^* \delta^{-n} \quad (n \to \infty) \tag{4.3.11}$$

The constants (c_∞^*, A^*) do depend on $G(y)$, and hence are not universal constants (Fig. 4.21). The result (4.3.11) is not limited to the superstable values of c, but also holds for example, for the *period-doubling bifurcations* (at c_n)

$$c_n = c_\infty - A\delta^{-n} \quad (n \to \infty). \tag{4.3.12}$$

An important lesson that can be learned from this discovery is that, in addition to qualitatively similar dynamic systems (e.g., topologically equivalent) which have some form of structural stability, there are also quantitatively similar bifurcation sequences which presumably also have some form of structural stability as indicated by the universal numbers (α, δ). It is known that there are other 'universal numbers' (α_k, δ_k)

Fig. 4.21

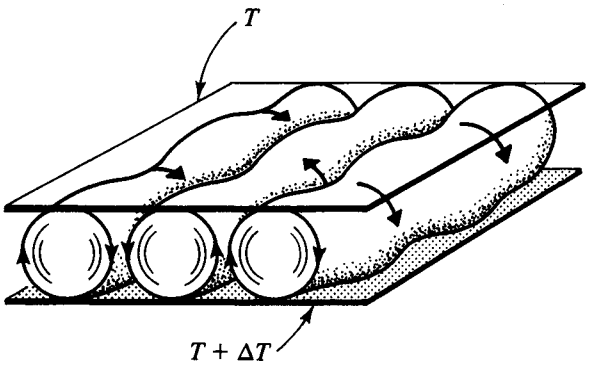


associated with more complicated maps which have other bifurcation sequences. This suggests (in a truly nebulous fashion) that there may exist a form of 'generalized' catastrophe theory which can relate these quantitative families, the universal set of 'universal numbers' $\{\alpha_k, \delta_k\}$. Whether all of these 'universal numbers' have physical significance, remains to be studied.

Feigenbaum has suggested that these universal constants should be observable in bifurcation sequences of much more complex systems than the present 'simple' maps. These issues, however, are not well understood at present (e.g., for a cautious view, see Wilson, 1983). Feigenbaum (1980) found that such period doubling bifurcations occur in higher order differential equations (e.g., the forced Duffing equation), and that they appear to yield the same bifurcation sequence, (4.3.9), with the numerical constant (4.3.10). This will be discussed further in Chapter 5, Section 14.

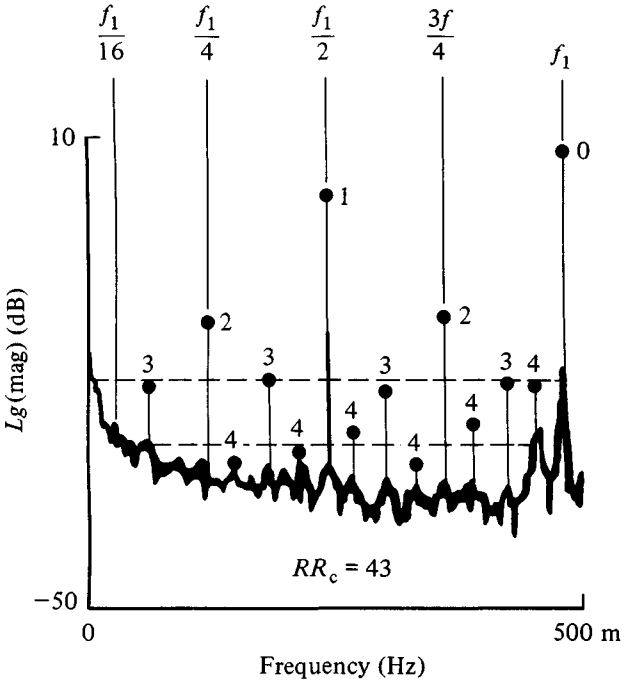
Other more complex systems also have at least several period-two bifurcations,

Fig. 4.22



whose successive amplitudes appear to be related to (4.3.7). This is illustrated by Libschaber and Maurer's (1979) study of the Rayleigh–Bénard convection of fluid between two plates held at different temperatures (see Fig. 4.22). As the temperature difference ΔT is increased, the convective vortical rolls of fluid become oscillatory (see the review by Busse, 1985). Feigenbaum's analysis (1979) of Libschaber and Maurer's data is illustrated, indicating four period-doublings of the initial frequency, f_1. He estimated the amplitude of the temperature frequency spectrum, as illustrated in Fig. 4.23 by the horizontal lines, and related their values to α, (4.3.7). That theory

Fig. 4.23



predicts an asymptotic scaling factor of 8.18 db between successive amplitudes, whereas the experimental ratios were 8.4 ± 0.5 between $n = 3$ and $n = 2$ and 8.3 ± 0.4 between $n = 3$ and $n = 4$ (after discarding the highest frequency amplitude). The ratio between $n = 1$ and 2 was not sufficiently 'asymptotic' to fit the theory.

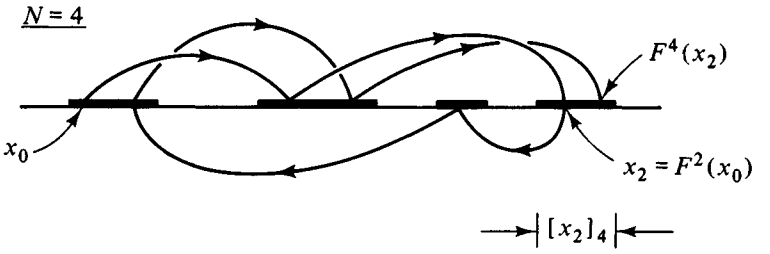
For $c > c_\infty$, the logistic map can, in addition to the infinite number of unstable 2^n period solutions, generate both new stable periodic solutions (e.g., 3×2^n, 5×2^n, etc.) and also aperiodic (nonrepeating) solutions depending on the value of c. We will consider first these *aperiodic solutions.*

An aperiodic solution involves a point which is mapped sequentially through 2^n 'bands' $\{l_k \leqslant x \leqslant u_k;\ k = 1, \ldots, 2^n\}$, instead of onto 2^n points, as in the case of periodic solutions. In other words, the points map periodically through 2^n disjoint bands of x values, but are not periodic (or eventually periodic) points. This is illustrated in the

figure below. For this class of aperiodic sequences Lorenz (1980) has introduced the terminology '*semiperiodic*' *sequences*, which he also found to occur in Poincaré maps of solutions of his third order differential model, which will be discussed in another chapter.

To be more precise, let $[x_k]_N$ be the range of the infinite set of point, $(F^N)^j$ $(j = 1, 2, \ldots)$, generated by F^N, beginning with $x_k = F^k(x_0)$. In other words $[x_k]_N = \max(x_{k+jN}) - \min(x_{k+jN})$, for all j (see Fig. 4.24). The mapping is called *semi-*

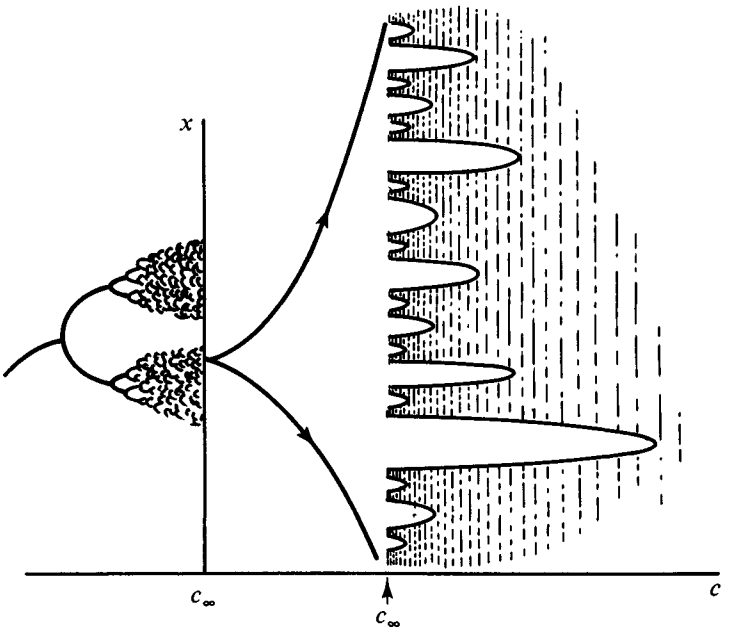
Fig. 4.24



periodic if the ranges $[x_k]_N$ are disjoint for $0 \leqslant k \leqslant N$, but the ranges $[x_k]_M$ overlap for $0 \leqslant k < M$, if $M > N$.

When c is above c_∞ it is frequently found that points do indeed map periodically through such a collection of 2^n 'bands', or disjoint x-regions. The number of such bands goes to infinity as c *decreases* toward c_∞. In other words, the number of bands, N, decreases as c increases and this is referred to as *reverse bifurcation.* This is loosely illustrated in Fig. 4.25, showing a magnification of a small region of x near $c_\infty (c > c_\infty)$.

Fig. 4.25



As the shaded region illustrates, these bands widen and then overlap (in pairs) as c is increased. Chang and Wright (1981) found that this series of semiperiodic bifurcations also satisfy Feigenbaum's asymptotic relationship, now with

$$\hat{c}_m \sim 3.569\,945\,671\ldots + \hat{A}\hat{\delta}^{-m} \quad (m \to \infty)$$

where $\hat{A} = 0.494\,454\ldots$ and $\hat{\delta} = 4.669\,201\ldots$. The band $N = 0$ is at $c = 3.678\,57\ldots$, beyond which odd periodic solutions occur. The shaded region in the above figure is meant to convey the fact that the dynamics of x_n, while periodic in the bands, 'skips around' in an erratic fashion each time it returns to a band. It's motion, while deterministic, is quite different for the motion of another initially nearby point. We will call such motion '*chaotic*'. One of our objectives will be to obtain various methods of characterizing *chaos*, both quantitatively and qualitatively.

On the larger scale of the following figure, many of these details are lost, and we can see at most only four of the semiperiodic bands. The density of points is not uniform, as the shading illustrates, and this leads to the rather enticing 'overlapping veils' appearance in this increasingly complex region. Contained in this region are an infinite number of '*windows*', so called because the chaotic 'veil' disappears in this region, and we can 'see through a window'. The veil is replaced by a simple periodic structure at some value of c, and then the map goes through a series of bifurcations much the same as occurs for c below c_∞, until another 'chaotic' region is reached. Only three of these windows are illustrated in the following figure (namely those that begin with period six, with period five, and with period three). The remaining windows have been illustrated in some detail by Collet and Eckmann (1980). We will examine them further in the following sections.

The nature of the periodic solutions (and only one, at most, can be stable for S-maps) is given by the following important and remarkable theorem:

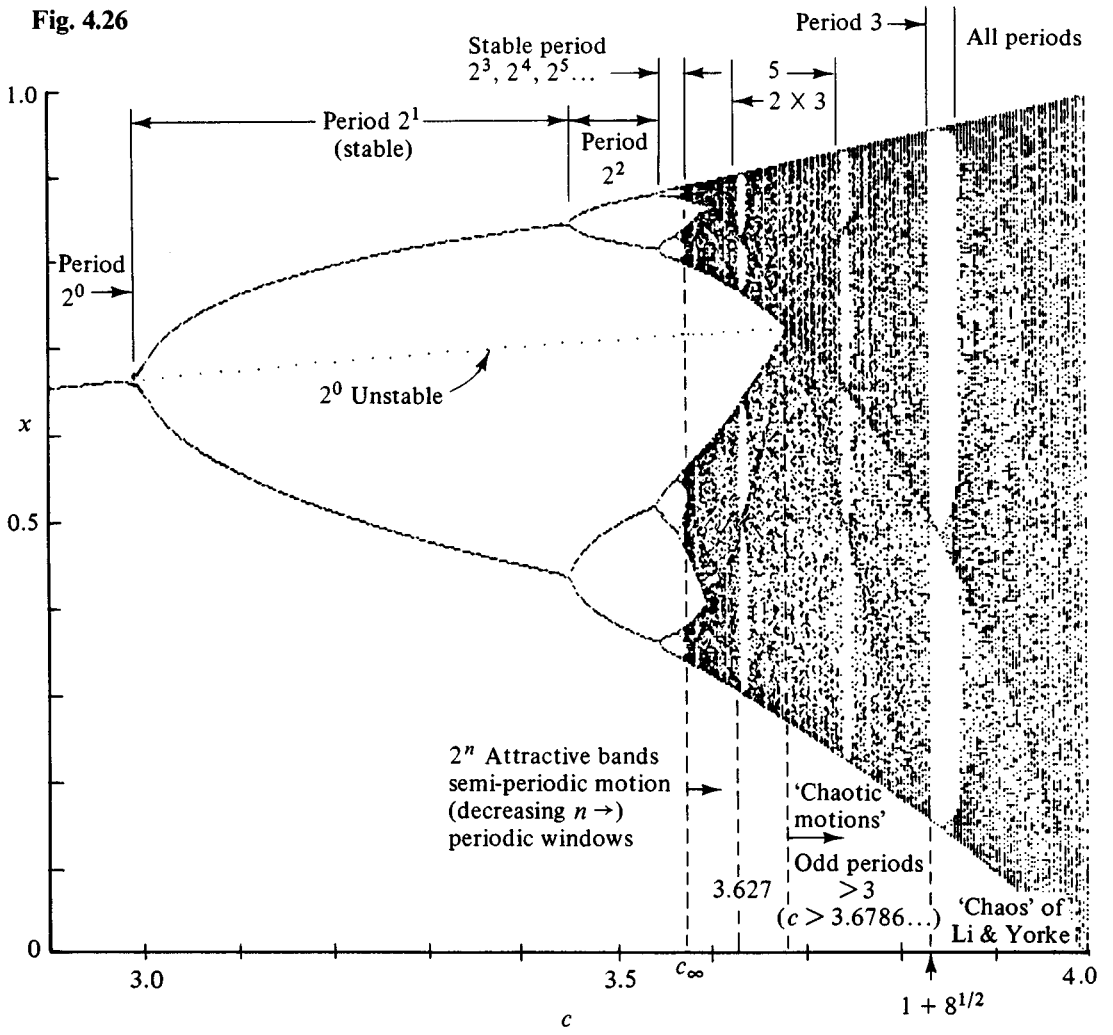
Theorem (Sharkovsky, A.N., *Ukranian Math. J.*, **16**(1), 61 (1964))

> *Let T be the ordered set $\{3 < 5 < 7 < \cdots < 2 \times 3 < 2 \times 5 < 2 \times 7 < \cdots < 2^2 \times 3 < 2^2 \times 5 < 2^2 \times 7 < \cdots < 8 < 4 < 2 < 1\}$. Let $f: I \to I = [0, 1]$, be a smooth map such that $f(0) = f(1) = 0$ which has a single critical point. If $m < n$ relative to the order in the set T, and f has a periodic point with prime (i.e., shortest) period m, then f has a periodic point of period n.*[†]

Thus for the logistic example, in Fig. 4.26 where $c = 3.627\ldots$ there is a period six solution, the beginning of a 'window'. According to the above theorem there are periodic (unstable) solutions with periods 2×5, 2×7, and so forth, following to the

[†]For a further discussion of this theorem, see Guckenheimer (1977), and the simplified proof by Osikawa and Oono (1981).

Fig. 4.26



right through the set T. Note that the odd periods have not yet occurred (they begin at $c = 3.6786\ldots$, with a new group of infinite periods!). As c is increased the periodic solutions have periods moving to the left through the T set. When $c = 1 + 8^{1/2}$, the period-3 window is reached, and we are at the first member of the T set (all periods exist).

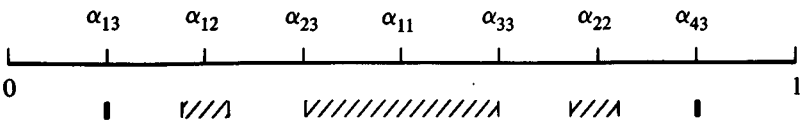
Some appreciation of why periodic solutions of a general continuous map must obey some ordering, as given by Sharkovsky's T set, can be obtained without much difficulty. The maps produced by a continuous $f(x)$, which has a periodic point $p = f^k(p)(p \neq f^i(p)$ if $i < k)$, can be analyzed quite readily by using a *directed graph* (or '*digraph*') method suggested by Straffin (1978), which is discussed in an Appendix E.

For $c > c_\infty$ the difficulty in representing the periodic vs. aperiodic regions of c is

related to the following features. The values of c for which the logistic map has stable periodic solutions fall into continuous bands. Two values (c_1, c_2) which have no stable periodic solutions are apparently separated by such a continuous band of c values, associated with stable periodic solutions. Thus the set of nonperiodic c values is apparently nowhere dense. This does *not* imply, however, that the set of nonperiodic c values has zero measure. These features were first conjectured by Lorenz (1964), who gave an example of a nowhere dense set of points which nonetheless has a finite measure, thereby bearing some resemblance to the present nonperiodic c set. We have already seen one example of such sets, the fat Cantor sets discussed in Chapter 2. We next consider Lorenz's example.

Consider the points $\alpha_{mn} = (2m-1)/2^n$, restricted to the interval $0 < \alpha < 1$, which form an enumerable set. A few points are illustrated in Fig. 4.27. Around each of these points

Fig. 4.27



construct an interval of length $(1/2)^{2^n}$, which rapidly decrease in length as n increases. For $n > 3$ the intervals are essentially points on the above scale (as illustrated). Since there are 2^{n-1} intervals for each n, the total length contained in all intervals (counting their overlaps) is

$$\sum_{n=1}^{\infty} 2^{n-1} 2^{-2^n} = 0.390\,747\ldots$$

Moreover, since some of these intervals overlap, the intervals clearly have a measure less than 0.4 (say). Hence the set of points outside of these continuum intervals have a measure greater than 0.6. However, since the α_{mn} are a dense set of points, the points outside their enclosing intervals are nowhere dense. Thus the set of points outside these intervals is nowhere dense but has a finite measure (>0.6!). The largeness of this measure is a result of the rapidly decreasing length of the intervals as n increases – a feature which is also true of the periodic c intervals, as the period increases. Whether this decrease is rapid enough to leave a finite measure for the nonperiodic c set appears to be unknown.

As fascinating as much of this 'fine structure' may be from a mathematical point of view, it is important to keep in mind that many of these details cannot be observed in experiments. Thus, for example, unstable periodic solutions are experimentally unobservable, there is no observational distinction between semiperiodic motion and periodic motion with very long periods, or with semiperiodic motion with large periods which is observationally aperiodic. Similarly, while Cantor sets are great fun mathematically, they do not exist in physical observations. The same is true of fractals.

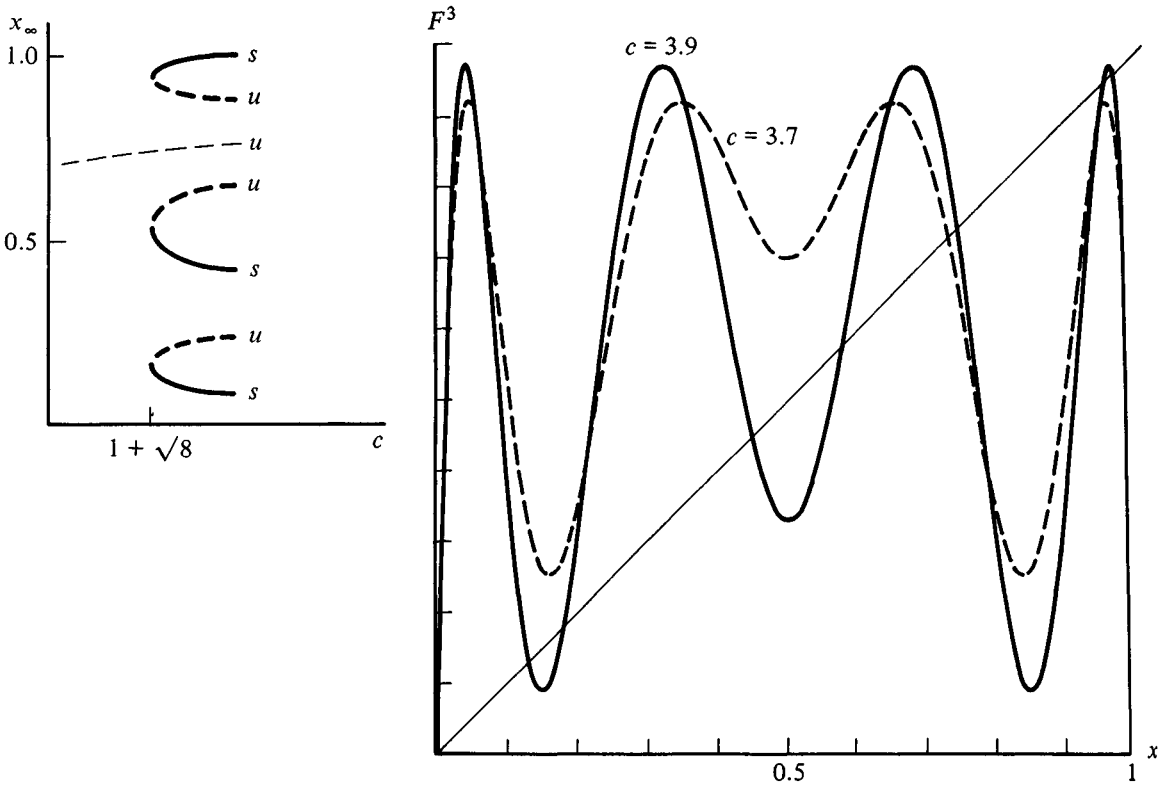
Only finite approximations can be observed. However, aside from quantum mechanical restrictions (e.g., the uncertainty principle), the ultimate observational restriction can vary greatly in different experiments, which perhaps justifies considering various asymptotic temporal features, and the entire Cantor set. In any case, we will continue to consider these features that are based on continua. This distinction between mathematical and physical continuity was discussed long ago by Poincaré, for example in his *Mathematics and Science: Last Essays*, Chapter III (Dover Pub., 1963). More recent considerations of this basic issue can be found in Section 10 of this chapter (also see Chapter 10, Section 16).

4.4 Tangent bifurcations, intermittencies

The appearance of a period-three solution was believed at one time to represent a type of 'benchmark' in this bifurcation process. It is on the one hand the first member of Sharkovsky's T set, so that all periodic solutions are now present. Secondly, the type of bifurcation which generates these odd periodic solutions are quite different from the 'pitchfork' period 2^n bifurcations. The nature of these bifurcations can be most readily appreciated by considering the graph of $F^3(x, c)$ for the two values of c shown (Fig. 4.28). Note, however, that this is the last such odd bifurcation as c is increased, which then completes Sharkovsky's T set. As c increases from 3.7 to 3.9 the second, fourth, and seventh extrema of F^3 become tangent, and then pass through the 45° line, all at the same value of c. Thus, in addition to a continuing unstable fixed point, three new fixed points of F^3 are 'born out of the blue', and then bifurcate into six fixed points. Only one x-value of each pair satisfies $|dF^3/dx| < 1$, and hence is stable. This *tangent bifurcation* contrasts with the pitchfork type, in which an existing fixed point splits into two stable and one unstable fixed point (recall double point bifurcations, and their exchange of stability). In the logistic case the period three bifurcation occurs at $c = 1 + 8^{1/2}$, and is illustrated in Fig. 4.28 in the control-phase space, showing the stable and unstable branches. Note that the symmetry of the stable–unstable 'horseshoes' is broken by the continuing unstable fixed point. Also note that unstable fixed points of F^3 can be next to each other (contiguous) for the same value of c, whereas this cannot occur for flows (Chapter 3).

Another feature of tangency bifurcations is that, just before the bifurcation, the dynamics of the map produces an effect called *intermittency* which is responsible for the 'folded veil' appearance in Fig. 4.26. Fig. 4.29 shows F^3 for $c = 3.828$, which is slightly lower than the period-three bifurcation point ($c = 1 + 8^{1/2} = 3.828\,427\ldots$). The slight opening between F^3 and the straight line can not be seen on this scale. When the mapped points come into the vicinity of this near-tangency region, the subsequent maps are 'held up' (delayed) in this region (with period three, of course). In other words the maps of $F(x)$ stay sequentially close to the three near-tangency regions, behaving

Fig. 4.28



much like a semiperiodic motion for a large number of maps. Fig. 4.30 illustrates this behavior in the present case, for the logistic map. It will be noted that it takes 42 iterations to 'break out' of this semiperiodic behavior. Moreover, this map returns to $x = 0.5065$ after 165 iterations, and hence returns to the semiperiodic behavior for another forty or so iterations. In other words the dynamics spends one quarter of its time in this semiperiodic state. This percentage increases dramatically as c is increased. If $c = 3.8284$ it takes 200 iterates of $F(x)$ before this semiperiodic behavior ends, and after 85 more iterates it returns to the semiperiodic dynamics, which now represents 70% of its behavior.

Intermittencies of various types are rather common features of 'turbulent' states. Thus in steady state turbulence in fluids it is not uncommon to see orderly ('semiperiodic') motion over a period of time in various regions of space, which then breaks up into much more chaotic behavior, only to become orderly at another space-time region. The temptation, of course, is to draw some parallels between these different forms of intermittency, but a solid connection (if it exists) awaits further research.

As noted above, these tangent bifurcations give birth to the 'windows' in the chaotic dynamics. The bifurcation structure within a window is essentially a microcosm of the

Fig. 4.29

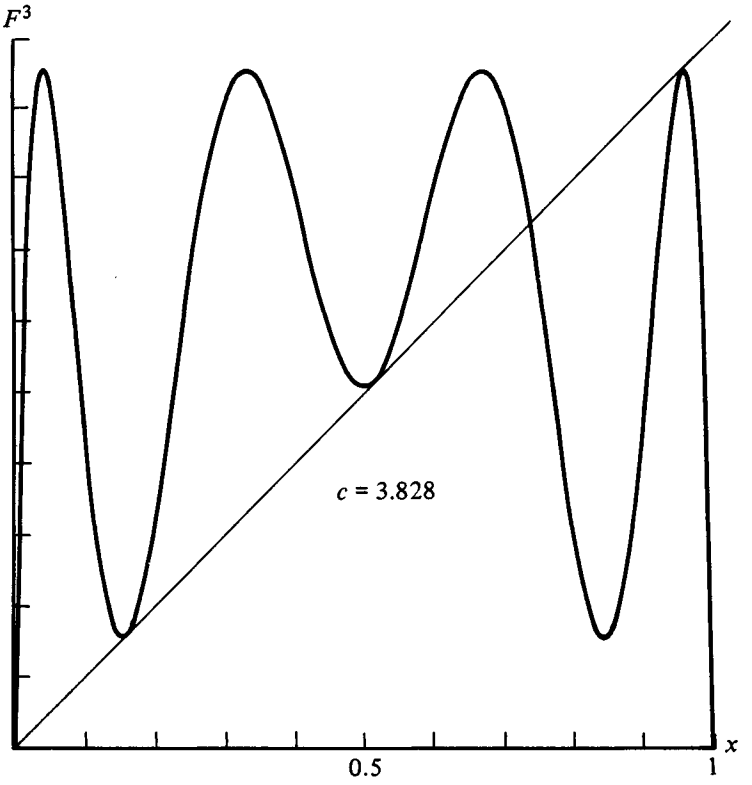
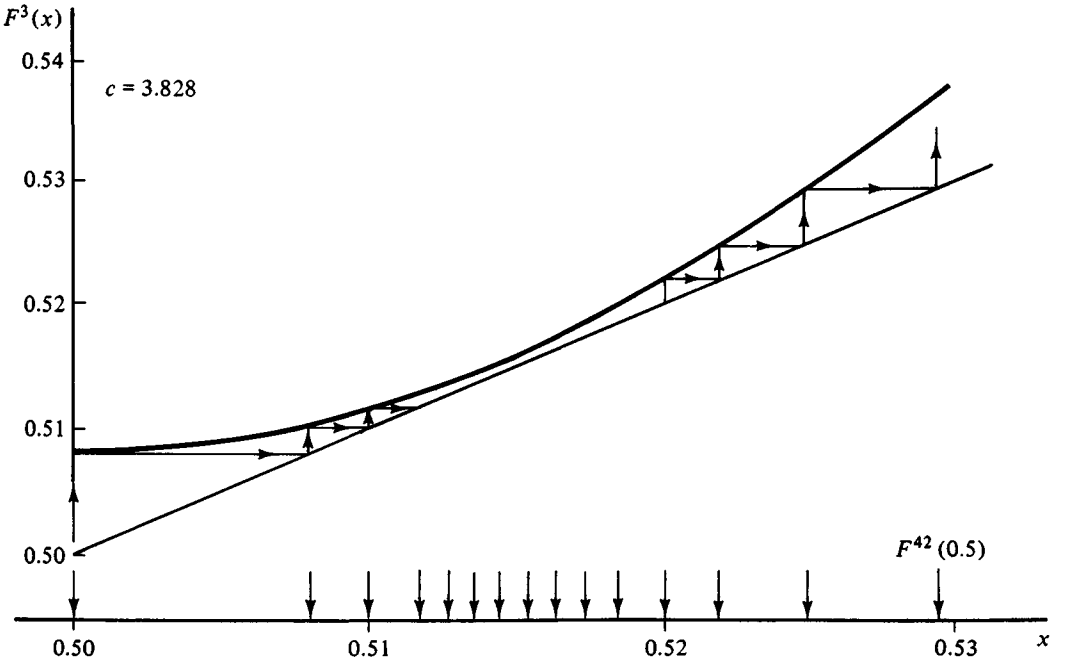
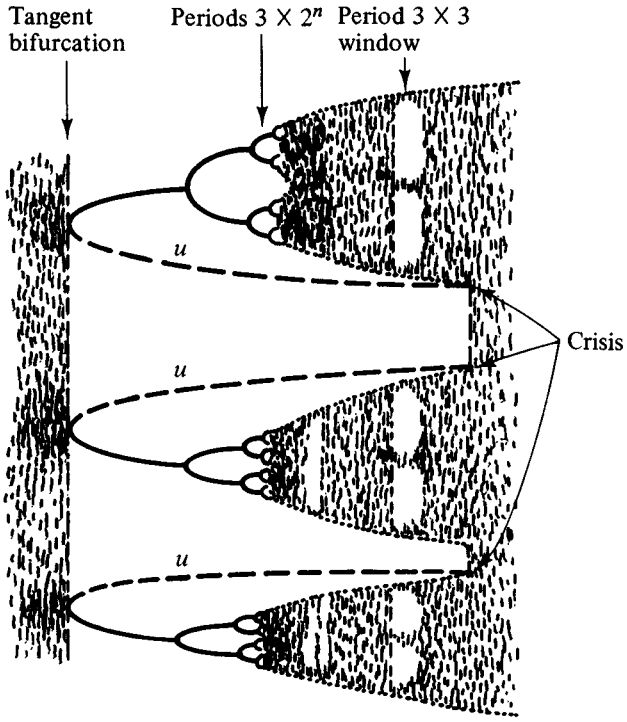


Fig. 4.30



scheme which occurs for $1 \leqslant c \leqslant 4$, but multiplied by the periodicity of the window. This is schematically illustrated in Fig. 4.31 for the period-three window, where there

Fig. 4.31



are three microcosms. Thus the former period-2^n bifurcations now become period-3×2^n cycles. This is followed by three chaotic attracting regions, which in turn have their periodic windows. A period-nine (3×3) window is illustrated (and, of course, in that window is ...!). The end of a window is characterized by a bifurcation in which the chaotic attractor discontinuously increases in size to form a continuous attractor (rather than three regions). This occurs at the value of c where the unstable branches of the tangent bifurcation intersects the chaotic attractors. This intersection bifurcation has been termed an interior *crisis* (Grebogi *et al.*, 1982), and represents a distinctly different type of 'bifurcation' (note that the term 'bifurcation' has become generalized to include modification of chaotic sets).

4.5 Characterizing 'deterministic chaos'

The occurrence of windows, and in particular the period-three window which is easiest to observe, presents the first opportunity to give some detailed meaning to the term

'*deterministic chaos*'. The adjective 'deterministic' will be implied in all our discussions, but will frequently be omitted for brevity.

Quite generally, we will see that chaos has many 'faces', describing different features of this multifaceted concept. Broadly speaking, we can divide chaotic aspects into:

Qualitative features	Quantitative features
Correspondence with coin tossing (Bernoulli sequences); Topological entropy	(uses measure/distance concepts) Dynamic limiting properties; Lyapunov exponents; Mixing; ergodicity; Kolmogorov–Sinai entropy. Power spectrum; Correlation functions.

We will discuss several of these features in this chapter, beginning with a qualitative description of the dynamics.

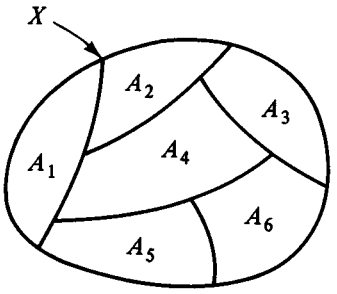
To characterize discrete dynamics, such as maps, in a qualitative fashion, it is useful to replace the continuum of numbers (i.e., R^1) by some finite set of symbols. The letters of the alphabet, or any other finite set of symbols, is acceptable. Such a characterization of the dynamics might be referred to as a dynamics of symbols, or more concisely, *symbolic dynamics*. Unfortunately the latter description has been used in a more restrictive sense by mathematicians (see Section 4.11). In any case, the patterns involving R and L, used by Metropolis, Stein and Stein in this U-sequences (Section 4.3), is an example of a qualitative description of periodic dynamics in terms of symbols. Their description is a special case of a much more general method of qualitatively describing dynamics symbolically.

To do this, we first divide the phase space into cells of our choosing. We assume that either the phase space is finite (say, the interval [0, 1]), or the dynamics remain bounded (e.g., on some bounded integral manifold). Then we can divide the space into a finite number of cells. This is referred to as *partitioning the phase space*. A partition, $\{A_i\}$, is any finite set of regions, which are disjoint and cover the dynamic region of interest; call it X.

$$\text{Partition} = \left\{ A_i | i = 1, \ldots, n;\ A_i \cap A_j = \phi, \text{ if } i \neq j;\ \bigcup_{i=1}^{n} A_i = X \right\}. \tag{4.5.1}$$

Such a partitioning of X is illustrated in Fig. 4.32. Metropolis, Stein and Stein used the partition $\{R, L\}$ of the space [0, 1]. More generally, we can describe the dynamics of a mapping by the sequence of symbols corresponding to the partitions visited by a

Fig. 4.32

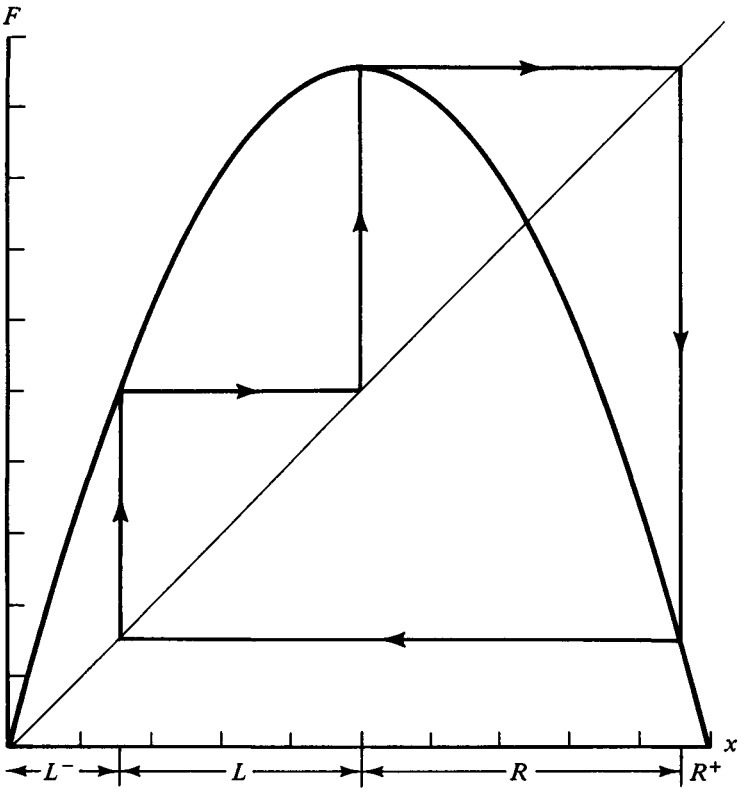


solution,

$$A_2, A_6, A_3, A_3, A_3, A_1, A_5, A_1, A_4, \ldots. \tag{4.5.2}$$

We will now use this partitioning method to describe the dynamics in the period-three window. Specifically, we will take $c = 3.83$, which has a stable period-three orbit, as illustrated. We choose to introduce the partition, involving four regions, L^-, L, R, R^+ (Fig. 4.33), and note that the mapping of the region L^- covers both L^- and L. We

Fig. 4.33

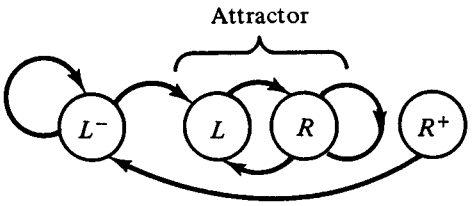


write this as $FL^- = L^- \cup L$. Similarly, for the other regions,

$$FL = R; \quad FR = R \cup L; \quad FR^+ = L^-.$$

That is, all points in R^+ map into L^-, and in fact FR^+ covers L^-. Similarly points in L^- either map to L^- or L. After enough iterations, any initial point of L^- enters L (Fig. 4.34). In other words the region $L \cup R$ is an attractor. All points in $L \cup R$ stay in this region.

Fig. 4.34



Therefore we will focus our attention on the dynamics between L and R (only). We note, however, that this dynamics is not symmetric. Points in L always map to R, whereas those in R may map to R or L. Because we want to compare the dynamics with the process of flipping a coin, which is symmetric, we will consider the map $G \equiv F^2$, rather than F. This is illustrated in Fig. 4.35, and now we use the symbols (H, T) for the two regions of interest, obviously to relate them to a coin. We note that the dynamics is now symmetric, since

$$GH = H \cup T \quad \text{and} \quad GT = H \cup T.$$

Now we return to deterministic chaos. What is one characterization of this concept? It seems reasonable to say that a system behaves chaotically if there are dynamic solutions which behave like the flipping of a coin (Fig. 4.36). A series of coin flips is characterized by some infinite sequence of symbols

$$H, H, T, H, T, T, T, H, T, \ldots. \tag{4.5.3}$$

Since the probability of H or T turning up in the next flip is 1/2, this is called a *Bernoulli sequence*. In our deterministic system, there is, of course, no probability involved – it is simply a question of comparing sequences. This leads to one definition of deterministic chaos:

> *A deterministic system is* **chaotic** *if we can find, for any prescribed Bernoulli sequence, an initial state which will dynamically produce the same sequence of symbols, as it moves through some fixed partition of the phase space. In that case, we will say that initial state corresponds to the prescribed Bernoulli sequence.*

Fig. 4.35

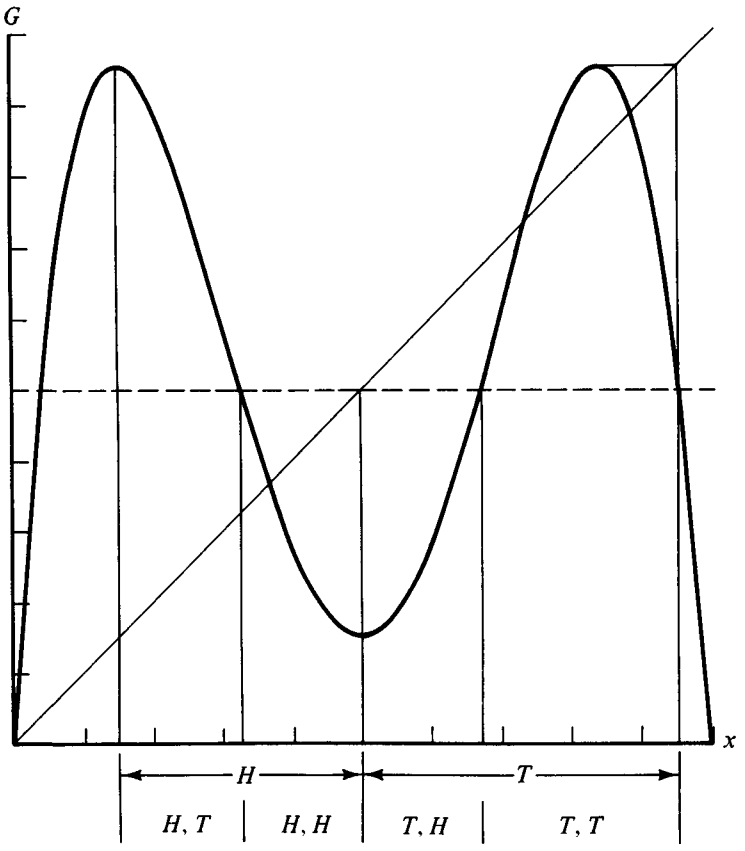
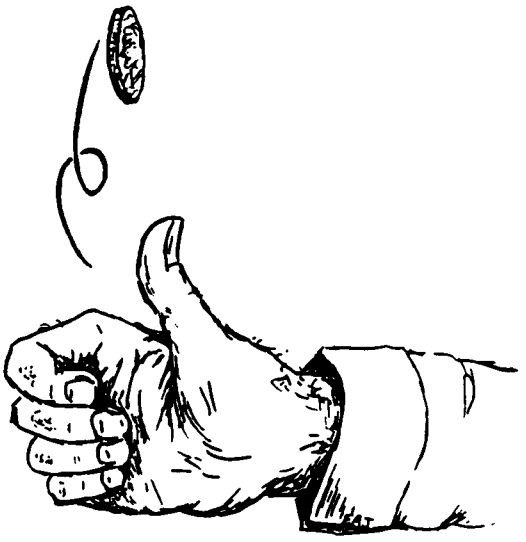


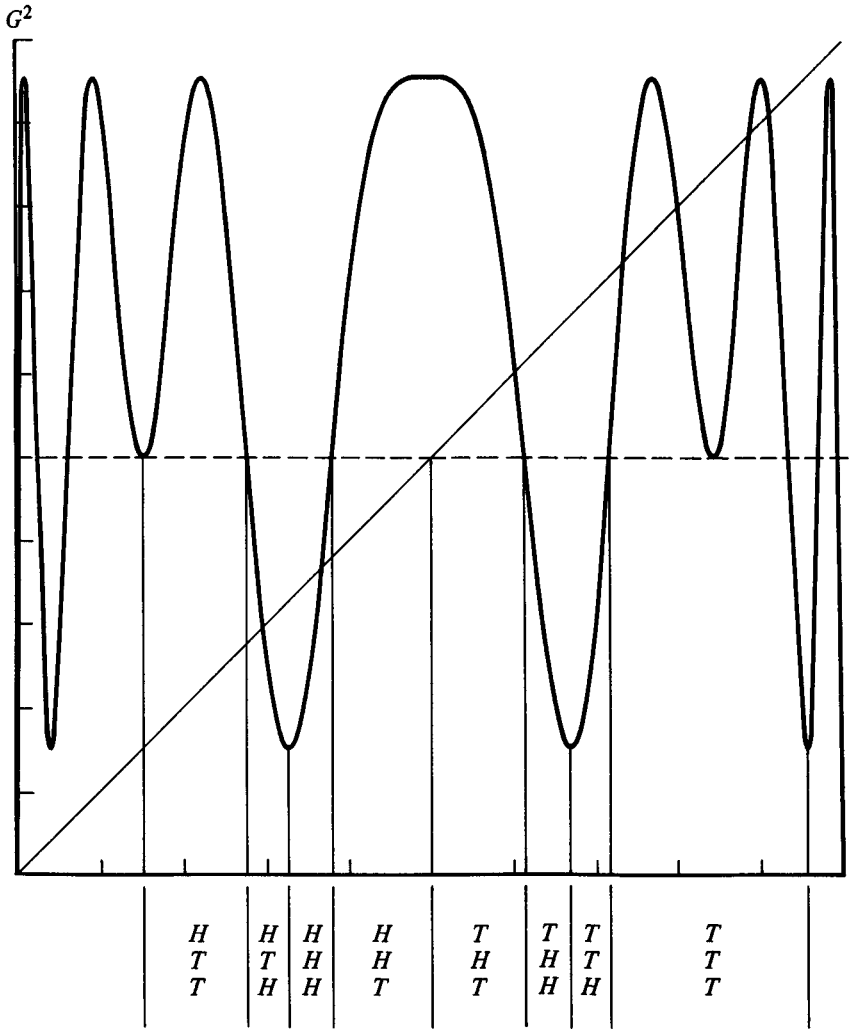
Fig. 4.36



To see that this correspondence can be made in the period-three region, consider the figure for the map G. If the initial state is in the x-region labeled (H, T), then this mapping will take it from the H region to the T-region. In other words, an initial state in the (H, T)-region produces the sequence H, T in its first step. Similarly, beginning in (T, H) will produce the sequence T, H in one step.

For two steps, consider G^2. Now, for example, any initial state in the region HTT will begin in H, map to T, then again map to T. We see from Fig. 4.37 that there are initial

Fig. 4.37



conditions which will produce any of the 2^3 sequences of length 3,

$$H,T,T;\quad H,T,H;\quad H,H,H;\quad H,H,T;\quad T,H,T;\quad T,H,H;\quad T,T,H;\quad T,T,T.$$

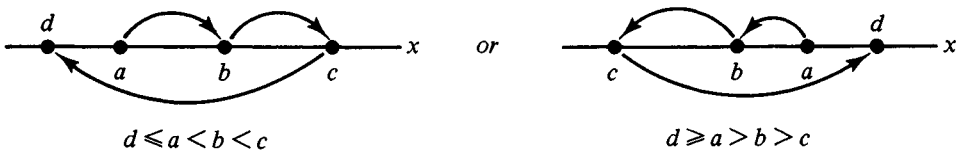
Similarly, by examining G^3, you can see that there are initial conditions which will generate any of the 2^4 sequences of length 4. We conclude that the dynamics of the system can produce an orbit which corresponds to any Bernoulli sequence – so the system is chaotic, in that sense.

The fact that the appearance of a period-three mapping is sufficient (but it turns out not to be necessary) to produce one form of chaos (to be defined shortly) was recognized first by Li and Yorke (1975). The theorem of Li and Yorke deals with the general equation

$$x_{n+1} = F(x_n)$$

where $F(x)$ gives a continuous mapping of a closed interval (say $[0, 1]$) into itself. It says that if F maps some point a into a point $b\,(=F(a))$, and b into $c\,(=F(b))$, and c into d $(=F(c))$, all satisfying the inequality shown in Fig. 4.38, then the points in the interval

Fig. 4.38



$[0, 1]$ will necessarily map in a well-defined chaotic fashion. In particular the above condition is satisfied if there is a point with period three (i.e., $d = a$), hence their title, '*Period Three Implies Chaos.*'; this only holds in one dimension (see Section 6.4).

One of the chaotic features in this case is that there are points which have any desired period n (Sharkovsky's theorem), and there is an uncountable number of points which are not even asymptotically periodic. We will call this set S. If q is a periodic point, p is called an asymptotically periodic point if

$$\lim_{n\to\infty} |F^n(p) - F^n(q)| \to 0. \tag{4.5.5}$$

Hence this result says that there is a uncountable set of points, S, which do not even tend toward a periodic solution (nor are any points in S periodic).

Perhaps what is of even greater interest are two other results due to Li and Yorke, which give us another (non-Bernoullian) way to characterize '*chaotic motion*'. The

formal statements are that, for every $p, q \in S$ $(p \neq q)$

$$\limsup_{n\to\infty} |F^n(p) - F^n(q)| > 0, \tag{I}$$

and

$$\liminf_{n\to\infty} |F^n(p) - F^n(q)| = 0. \tag{II}$$

We now need to define and interpret these results.

First of all recall that any infinite sequence of points, $\{x_n\}$, on a closed (compact) interval has at least one subsequence $\{x_{n_j}\} \in \{x_n\}$ which possesses a limit point (the Bolzano–Weierstrass theorem); that is $\lim_{n_j\to\infty} x_{n_j} \equiv l_{\{n_j\}}$ (dependent, of course, on the subsequence). The definition of lim sup in (I) is

$$\limsup_{n\to\infty} x_n = \max_{\{n_j\}} l_{\{n_j\}} \equiv l_{\max}. \tag{4.5.6}$$

that is, it is the maximum of all limits obtained from all subsequences of $\{x_n\}$. Similarly, in (II),

$$\liminf_{n\to\infty} x_n = \min_{\{n_j\}} l_{\{n_j\}} = l_{\min}, \tag{4.5.7}$$

the minimum of all limits obtained from all subsequences. Applied to (II) this says that the sequence of values, $|F^n(p) - F^n(q)|$, obtained from any two initial points (p, q) in S, has a subsequence of mappings such that

$$\lim_{n_j\to\infty} |F^{n_j}(p) - F^{n_j}(q)| = 0$$

In other words we can find maps F^{n_j} of p and q which are arbitrarily close to each other as $n_j \to \infty$.

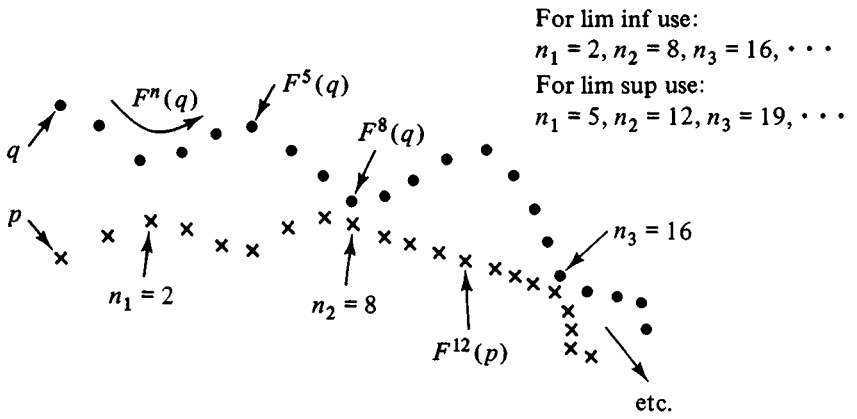
On the other hand (I) says that we can also find a sequence of mappings such that

$$\lim_{n_j\to\infty} F^{n_j}(p) = p_0, \quad \lim_{n_j\to\infty} F^{n_j}(q) = q_0 \quad \text{and} \quad |p_0 - q_0| = l_{\max} > 0.$$

Therefore what (I) and (II) state is that the mappings, F^n, of p and q remain erratic as $n \to \infty$ in the sense that they both become arbitrarily close for some steps, n_j, and remain a finite distance apart ($l_{\max}$) for another subset of these maps. This is schematically illustrated in Fig. 4.39.

This is an interesting characterization of a chaotic mapping, but it has the disadvantage that it yields no quantitative measure of this chaotic motion. In particular it gives no measure of how rapidly mixing occurs for the mapping of adjacent points. This is a characteristic which is measured by the Kolmogorov–Sinai entropy, to be discussed later, and the Lyapunov exponents (Section 6). Another feature not clear from this theorem is that, while the set S is uncountable, it may have (Lebesgue) measure zero. It should always be kept in mind, however, that 'measure' comparisons

Fig. 4.39

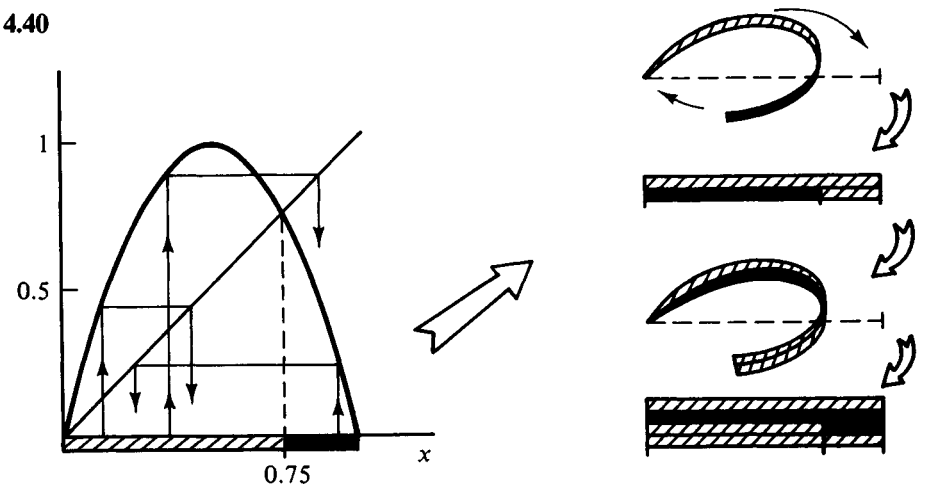


are strictly mathematical concepts, and need not represent the physical importance of the quantities.

The chaos defined by Li and Yorke is by no means the only possible definition of chaos, and the condition of period three is not necessary even for their form of chaos. Various definitions of chaos and sufficient conditions have been discussed by Oono and Osikawa (1980).

Other forms of chaos will arise in the study of higher order systems, but at the heart of all chaotic dynamics is the *fold-over* ('*kneading*') nature of the flows or maps. This is illustrated in Fig. 4.40, where it is shown how a map (e.g., the logistic map, with $c = 4$)

Fig. 4.40



can take the interval [0, 1] and fold it double onto the same interval. Note that the dark and stripped regions are separated by the fixed point.

It should be noted that the logistic map, with $c = 4$ (i.e., $F(x) = 4x(1 - x)$), has no attracting periodic orbits, as illustrated in the exercise below. At the same time, the periodic orbits are dense in the interval (0, 1) (i.e., any region in the interval contains a

point which is a periodic orbit). Moreover, the dynamics is *ergodic* over [0, 1] for this value, $c = 4$. This means that the only invariant subregions have either Lebesgue measure zero or one. This will be discussed further in Section 8.

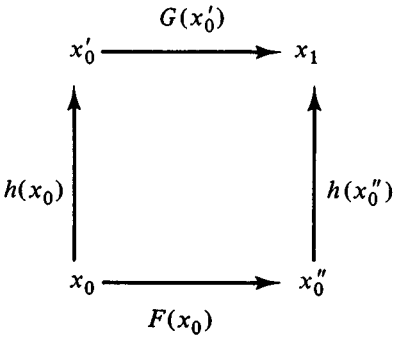
Exercise 4.6. (a) If period three implies chaos, why doesn't the period-three window of the logistic map appear chaotic? Think about what is required to observe a chaotic set. (b) Using ideas from the above fold-over nature of the $c = 4$ logistic map, together with *Brouwer's fixed point theorem*, can you speculate on the distribution of periodic points in [0, 1]?

Exercise 4.7. Show that $x_n = \sin^2(2^n \pi\theta)$ is a solution of $x_{n+1} = 4x_n(1 - x_n)$ for all θ, and comment on the distribution of both the periodic and the nonperiodic solutions (also see S.M. Ulam, *A Collection of Mathematical Problems* (Interscience Pub., 1960) pp. 150).

In Chapter 4 we discussed the topological similarities of the continuous flows produced by ODE. This led to the concept of TOE. An analogous property can be defined for two maps, $F(x)$ and $G(x)$, if their respective dynamics can be related by any homeomorphism, $h(x)$. If we can find $h(x)$ such that $F(x)$ and $G(x)$ can be related by $G \cdot h = h \cdot F$ (as operators on the points of R^n), then F and G are said to be *topologically conjugate* (TC). These sequential operations are illustrated schematically in Fig. 4.41. Written in terms of functions, rather than as operators, $F(x)$ and $G(x)$ are TC if

$$G(h(x)) = h(F(x)) \tag{4.5.8}$$

Fig. 4.41



for all x, and some homeomorphism $h(x)$. In particular if P_1 and P_2 are two *Poincaré maps* of TOE flows, then P_1 and P_2 are *topologically conjugate.*

Exercise 4.8. Show that the (differentiable) map $F(x) = 4x(1 - x)$ $(1 \geqslant x \geqslant 0)$ and the *tent map* $G(x) = 2x$, $(\frac{1}{2} \geqslant x \geqslant 0)$, $G(x) = 2(1 - x)$, $(1 \geqslant x \geqslant \frac{1}{2})$ are related by the homeomorphism $h(x) = (2/\pi)\arcsin x^{1/2}$ (noted first by S.M. Ulam and J. von

Neumann; see P. Stein and S. Ulam, 'Nonlinear transformation studies on electronic computers', *Rozprawy Metamatyczne*, **39**, 401–84 (1964)). In other words, $G \cdot h = h \cdot F$ (as operators). Determine the stability of the periodic orbits of F, by determining the stability of the periodic orbits of G.

4.6 Lyapunov exponents

One measure of 'chaotic' motion can be related to the sensitivity of the dynamical behavior of a system when its initial state is changed by a small amount. If the system's behavior is very *sensitive to initial conditions*, so that nearby points in the phase space separate 'fast' (say exponentially with time, or iterates of a map) over most of the phase space, then the system (not just a particular solution) can reasonably be described as being dynamically unstable. This may not be easy to establish because we may not be able to study most orbits of a system, but have to be satisfied with (hopefully) some 'representative' orbits. We are then effectively investigating the Lyapunov stability of an orbit, γ, but limiting our concern to exponential instabilities.

The same question can obviously also be raised about the properties of a map

$$x_{n+1} = f(x_n).$$

Let (x_0, y_0) be two nearby initial points in the phase space (R^1) and

$$x_n = f^n(x_0), \quad y_n = f^n(y_0).$$

If these points separate exponentially with n,

$$|y_n - x_n| = A \exp(\lambda n) \quad (\lambda > 0)$$

where $A = |y_0 - x_0|$. For large n.

$$(1/n) \ln |y_n - x_n| \to \lambda \quad (\text{large } n).$$

However, in the case of motion in a bounded region such exponential separation cannot occur for very large n, unless the initial points (x_0, y_0) are very close. Therefore, before taking the limit $n \to \infty$, we must take the limit $|x_0 - y_0| \to 0$. This then defines a constant

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \lim_{|x_0 - y_0| \to 0} \ln \left| \frac{x_n - y_n}{x_0 - y_0} \right|$$

or

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \lim_{|x_0 - y_0| \to 0} \ln \left| \frac{f^n(x_0) - f^n(y_0)}{x_0 - y_0} \right|$$

$$= \lim_{n\to\infty} \frac{1}{n} \ln \left| \frac{\mathrm{d} f^n(x_0)}{\mathrm{d}x} \right|$$

Now, using (4.2.11), we obtain

$$\lambda_\gamma = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \ln \left| \frac{\mathrm{d}f(x_k)}{\mathrm{d}x_k} \right| \tag{4.6.1}$$

which is called the *Lyapunov exponent*, for the orbit $\gamma \equiv \{x_n = f^n(x_0); n = 0, 1, \ldots\}$. If a group of solutions (i.e., with different initial x_0) are asymptotically attracted to the same ergodic subset of the phase space, then their Lyapunov exponents will all be the same, and the subscript γ can be dropped. This is true because the limiting set of points, large k, determine the value of λ_γ in (4.6.1).

More generally, if we know the probability distribution function, $P(x)$, of the asymptotic ($n \to \infty$) set of points for all equally probably initial conditions, then an average Lyapunov exponent (for all initial states) can be obtained by replacing the sum ('time average') in (4.5.1) by the phase average for all initial states

$$\bar{\lambda} = \int \ln |\mathrm{d}f/\mathrm{d}x| \, P(x) \, \mathrm{d}x. \tag{4.6.2}$$

In practice, it frequently occurs that most initial states (in a measure sense) are attracted to some ergodic subset and therefore we can drop both the subscript on λ_γ and the bar on $\bar{\lambda}$, and discuss only the unique Lyapunov exponent of the system. However, the above provisions should be kept in mind for more general situations.

Since $P(x)\,\mathrm{d}x$ is the probability of finding a mapped point, x_n, in the region $\mathrm{d}x$ of x as $n \to \infty$, this distribution must be an invariant function of this map. That is

$$P(y) = \int \delta(y - f(x)) \, P(x) \, \mathrm{d}x. \tag{4.6.3}$$

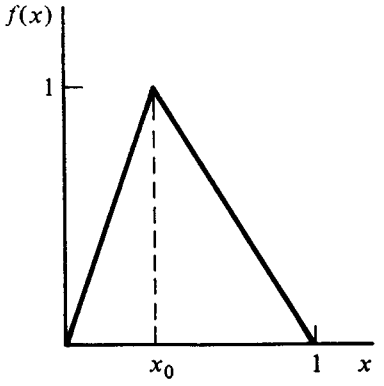
Exercise 4.9. A useful example of these ideas is the case of the logistic map with $c = 4$, $f(x) = 4x(1-x)$, which was originally studied by M. Kac (*Ann. Math.*, **47**, 33–49 (1946)). Show first that $P(x) = [\pi(x(1-x))^{1/2}]^{-1}$ is a normalized probability distribution on $[0, 1]$. Next, show that it satisfies the invariance property (4.6.3) for this map. Finally, show that the Lyapunov exponent for this map is $\lambda = \ln 2$.

Exercise 4.10. Using (4.6.2), show that any stable (unstable) periodic orbit has a negative (positive) Lyapunov exponent.

In some simple cases the invariant distribution can be obtained explicitly, and the Lyapunov exponent can be computed, using (4.6.2). For example:

Exercise 4.11. Show that for the 'tent map' (see Fig. 4.42) $f(x) = x/x_0$ (for $0 \leqslant x \leqslant x_0$), $f(x) = (1-x)/(1-x_0)$ (for $x_0 \leqslant x \leqslant 1$) the invariant distribution is $P(x) = 1$, and obtain $\bar{\lambda}$ as a function of x_0. What's the most unstable map in this family, $\{0 < x_0 < 1\}$?

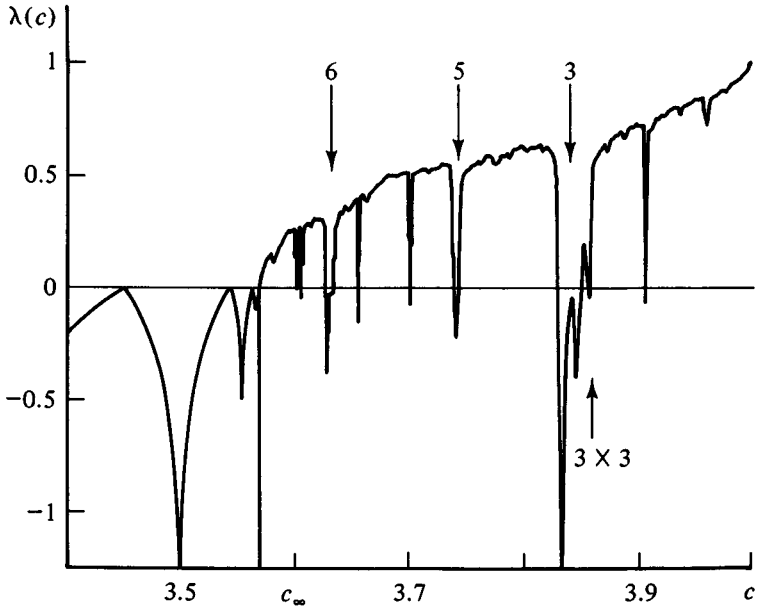
Fig. 4.42



On very rare occasions, we can determine $\bar{\lambda}$ if $P(x)$ is not known. Thus, if $f(x)$ consists of any number of linear sections with slopes of equal magnitude, $\mathrm{d}f/\mathrm{d}x = \pm n$, then $\bar{\lambda} = \int P(x) \ln n \, \mathrm{d}x = \ln n$, regardless of the details concerning $P(x)$. This is clearly an exceptional situation!

The Lyapunov exponent can, of course, be obtained directly by evaluating (4.6.1), without obtaining $P(x)$. Shaw (1981) used $\log_2$ to obtain Fig. 4.43 for λ vs c, for

Fig. 4.43



the logistic map, using 300 values of c, separated by 0.002. Each value of λ was obtained by using 100 000 maps in (4.6.1). The negative regions correspond to attractive periodic orbits which have the relatively wide parameter windows. The windows corresponding to periods 6, 5, and 3, and even the 'microcosm' 3 × 3 window (Section 4), are labeled in

Fig. 4.43. The infinite number of other attractive orbits cannot be seen on this scale. We should also emphasize several facts which are included in the above figure:

c = 4

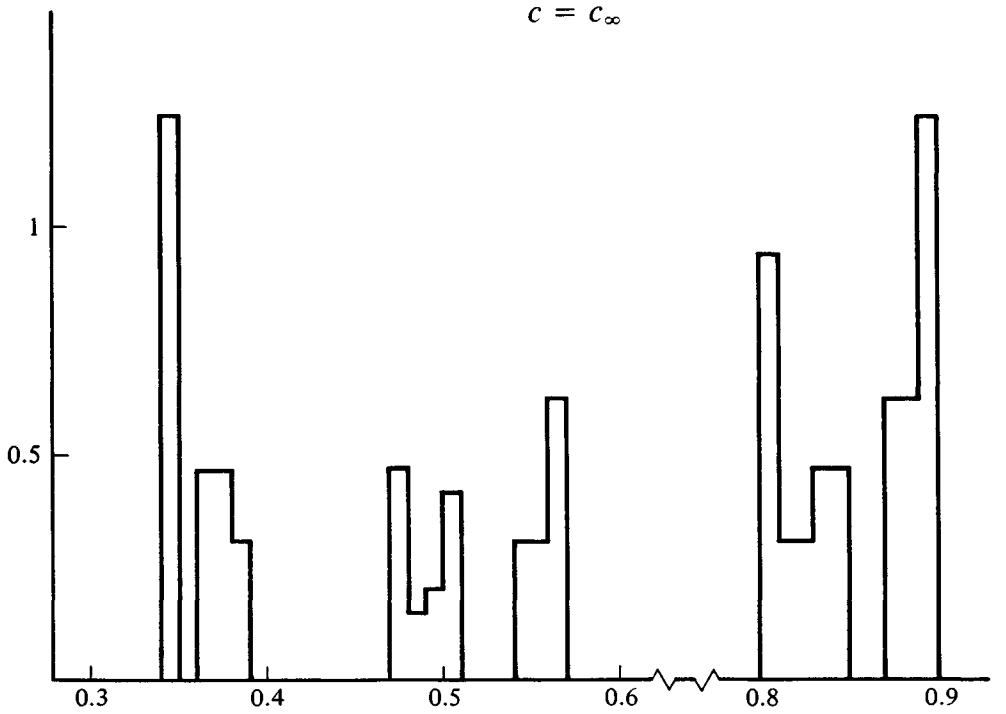
This is the most chaotic situation in the sense that λ is the largest, so the set of points have the greatest sensitivity to initial conditions. On the other hand, essentially all initial states move continually over the entire interval (0, 1). Therefore this set cannot be regarded as an attractor.

$c = c_\infty^{(k)}$

These are the Cantor sets which occur at the end of the bifurcation sequences $k2^n$, in the period k window. Our previous c_∞ is $c_\infty^{(1)}$ in the present notation. Some of these are illustrated in the above figure at the upward crossing of the c axis, so that $\lambda = 0$. Therefore these sets are not sensitive to initial conditions. On the other hand they are attractors of most initial states.

We will see later that there are dynamic systems which combine the features of these two types of sets. Namely, there are sets which on the one hand are *attractors* and yet

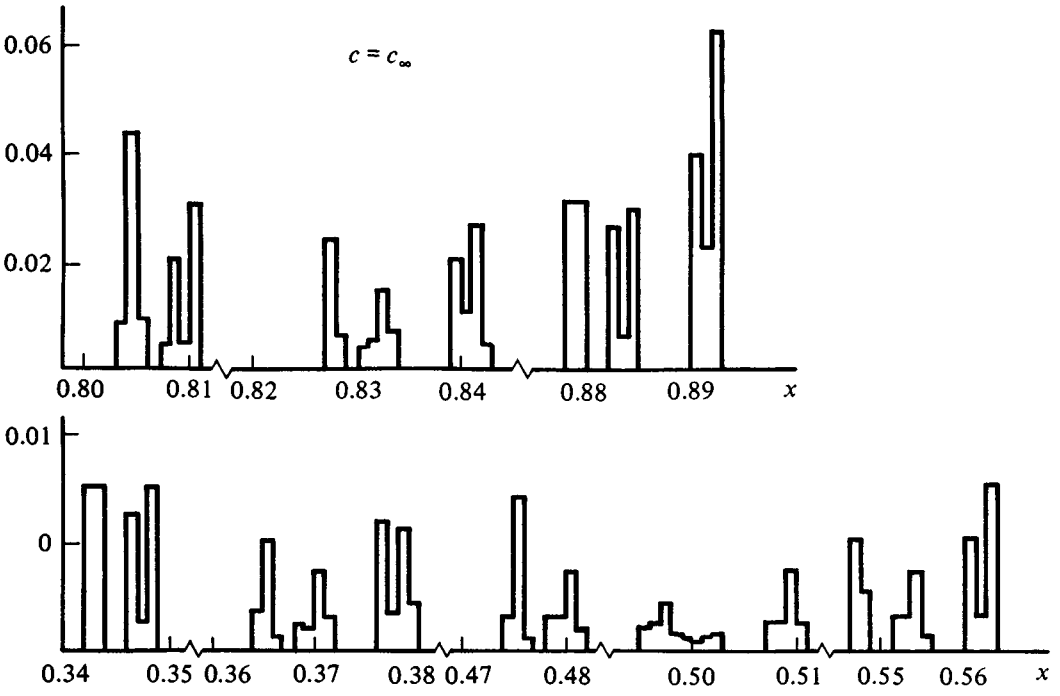
Fig. 4.44



have positive Lyapunov exponents, $\lambda > 0$, so that they are *sensitive to initial conditions.* Such a dynamic set is sometimes referred to as a *strange attractor*, although there is presently no generally accepted precise definition of a strange attractor. We will encounter such sets in Chapter 6.

The determination of the invariant distribution, $P(x)$, for general maps can be obtained to whatever accuracy that one is willing to invest computer time (and money!). The situation is simplest if all initial points (except a set of measure zero) go asymptotically to some ergodic region, as discussed above. In that case, $P(x)$ can be approximated by a histogram generated by the mappings of any representative initial state, provided that enough iterates are taken to approximate the asymptotic (invariant) set. As an example, a coarse histogram (bin size $\Delta x = 0.01$) for the logistic map at c_∞ is shown in Fig. 4.44. We see that all points of the nonenumerable set, X (see Fig. 4.14), are contained within six disjoint regions on this scale (note the broken x scale). If the bin size is reduced to $\Delta x = 10^{-3}$, then the histogram looks like the following. There are now twenty disjoint regions which are discernible (Fig. 4.45). More

Fig. 4.45



generally, one can begin with a uniform distribution, $P_0(x) = 1$, and map it for a large number of x values, yielding $P_1(x) = F(P_0(x))$. Continuing this process will asymptotically yield the invariant $P(x)$ (see, e.g., Hoppensteadt and Hyman, 1977, Shaw, 1981, and Chang and Wright, 1981).

4.7 The dimensions of 'near self-similar' cantor sets

The limit set of stable 2^n periodic points of the logistic map, obtained as $c \to c_\infty$, has the approximate distribution illustrated in Section 4.6. From this information we can obtain an aproximate value for either the capacity or information dimension of this set. These dimensions are discussed in Section 2.6, which may need to be reviewed.

Exercise 4.12. From Fig. 4.45, estimate d_c. Note that $\varepsilon = \Delta x = 10^{-3}$ in this figure. The answer is not very accurate, as we will see below. Note also that the estimate of d_1 involves a considerable amount of additional labor.

While, in principle, the values of d_c and d_1 can be obtained in this fashion by decreasing ε, a large number of cells, $N(\varepsilon)$, are required for significant accuracy. Chang has proposed a much more rapidly converging method of obtaining the dimensions of a class of such Cantor sets, which is based on their near self-similar structure (recall the general definition of Cantor sets, Appendix A, does not imply self-similarity).

We begin with the set, X (see (4.2.15)ff), obtained in the limit of the bifurcations of stable 2^n cycles, as $c \to c_\infty$,

$$x_n = F^n(1/2, c_\infty) \quad (n = 1, 2, \ldots). \tag{4.7.1}$$

Recall that, for an S-map, the maximum of $F(x)$ tends to any stable periodic orbit. Next, we use the iterates to make a geometrical construction of the attracting set, X, based on the sequential removal of open regions of x which clearly do not contain the attractor. This sequential removal of open regions is, of course, precisely the way Cantor sets are constructed, as discussed in Chapter 2. The clever aspect of this method is to recognize and use the near self-similarity which occurs in these sequential removals, in order to terminate the iterates (4.7.1) at a low order, and still obtain an accurate estimate of d_c.

Consider first the points x_1 and x_2, obtained from (4.7.1), which are illustrated in Fig. 4.46. It is not difficult to see that all $0 < x < 1$ will eventually map into the region $x_2 < x \leqslant x_1$, or (2, 1) for brevity. Therefore the attractor X is contained in (2, 1). Next consider the iterates x_3 and x_4, also illustrated in the figure. We can readily see that all points will eventually map into the union of (2, 4) and (3, 1). That is, we have removed the open interval, $x_4 < x < x_3$, from (2, 1) in the process of constructing the attractor. Now we can proceed in two ways, one being exact, the other an approximation based on a self-similar approximation. The exact method involves obtaining the next two iterates of (4.7.1) in each of the regions (2, 4) and (3, 1). This involves the points x_5, x_6, x_7, and x_8. These points are illustrated in Fig. 4.47. We conclude, as before, that the attractor must lie in the union of (2, 6), (8, 4), (3, 7), and (5, 1). In other words, we have removed the open intervals $x_6 < x < x_8$ and $x_7 < x < x_5$ from (2, 4) and (3, 1) respectively. The result is exact, but we are still a long way from obtaining d_c if we follows this procedure.

Fig. 4.46

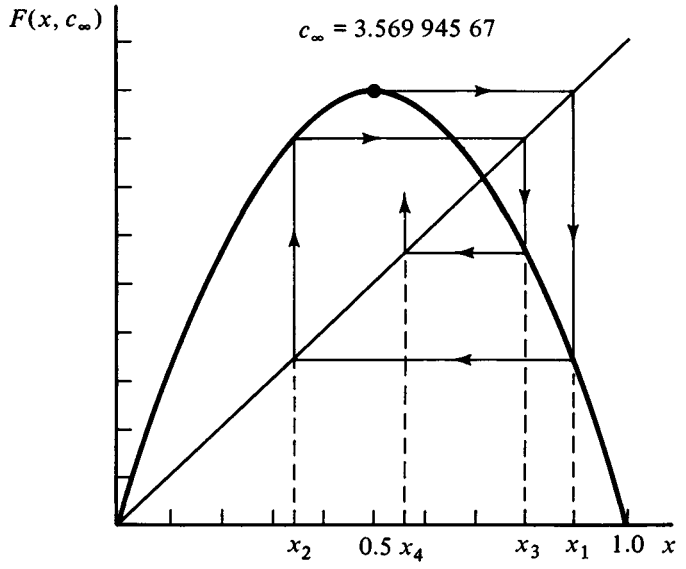
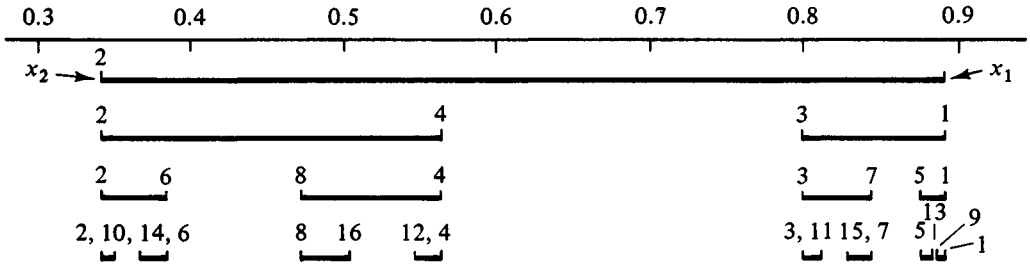


Fig. 4.47



Rather than proceeding in this fashion, we note that the intervals in the above figure have nearly a self-similar structure (allowing for a trivial interchange in the ordering of the long and short intervals). If $L(i, j)$ is the length of the interval (i, j), then we find for the logistic map, with $c = c_\infty = 3.569\,945\,67\ldots\ldots$, the following values: $L(1, 2) = 0.5499$; $L(1, 3) = 0.0885$, $L(2, 4) = 0.2200$; $L(1, 5) = 0.0140$, $L(3, 7) = 0.0380$, $L(4, 8) = 0.0876$, and $L(2, 6) = 0.0385$. If these were exactly similar then, for example, we would find the following equalities on the left, compared with the actual values on the right:

$$L(4,8)/L(2,4) = L(2,4)/L(1,2), \quad \text{whereas } 0.3980 \simeq 0.4001$$
$$L(2,6)/L(2,4) = L(1,3)/L(1,2), \quad \text{whereas } 0.1749 \simeq 0.1609$$
$$L(1,5)/L(1,3) = L(1,3)/L(1,2), \quad \text{whereas } 0.1580 \simeq 0.1609$$
$$L(7,3)/L(1,3) = L(2,4)/L(1,2), \quad \text{whereas } 0.4292 \simeq 0.4001.$$

We see that these intervals are only approximately similar, so the Cantor set being produced is not exactly self-similar on this scale. Moreover, as n is increased, the

successive groups of 2^{2n} intervals do not become increasingly similar to the group of 2^n intervals, relative to the original interval $(1, 2)$. Nevertheless this approximate self-similarity can be used to obtain increasing accurate estimates of d_c, as n is increased.

Exercise 4.13. If you would like to try some computing, set up the ratios of lengths which would have to hold if the group of sixteen intervals, obtained in the next step beyond the above figure, is similar to the four intervals in the figure. Then compute these ratios, and see how similar are the groups. Note that we are now examining whether the Cantor set is self-similar to these four intervals. Note also that adjacent points, producing the intervals (i, j), satisfy $|i - j| = 16$ in the sixteen intervals.

If the Cantor set were self-similar in the two intervals $(1, 3)$ and $(4, 2)$, then the dimension of this set is, according to the discussion given in Exercise 2.29, given by

$$L(4, 2)^{d_c} + L(3, 1)^{d_c} = L(1, 2)^{d_c}. \tag{4.7.2}$$

This gives the lowest approximation of d_c. In the next approximation, we can use the iterated intervals $(2, 6)$, $(8, 4)$, $(3, 7)$, and $(5, 1)$, and assume that the Cantor set is self-similar on this scale. This yields for d_c

$$L(2, 6)^{d_c} + L(8, 4)^{d_c} + L(3, 7)^{d_c} + L(5, 1)^{d_c} = L(1, 2)^{d_c}. \tag{4.7.3}$$

Proceeding in this fashion we can obtain higher approximations for d_c, based on Cantor sets which are self-similar in the $2^n \equiv N$ intervals generated by the 2^{n+1} iterates of $x = 1/2$. The result of this, based on the equations (4.7.2), (4.7.3), and obvious generalizations, are tabulated:

Self-similar Cantor set, based on N intervals	d_c
2	0.5261
4	0.5329
8	0.5346
16	0.5354
32	0.5359

Chang and McCowan (1984) obtained a more rapidly convergent result, based on the use of the universal function (satisfying $\alpha f^{*2}(x/\alpha) = f^*(\alpha), \alpha = -2.5029$) to generate the intervals, and a suitable extrapolation scheme to obtain d_c. They obtained $d_c = 0.5379$, so the above result is accurate to 0.4%. It is clearly much more accurate than the result in Exercise 4.12. They also applied this method to other near self-similar

Cantor sets, generated by bifurcation sequences which are quite different from the common $k2^N$ periodic sequences. They considered the sequence of periodic solutions consisting of periods $N, N^2, N^3, \ldots$, and obtained the dimension of the limiting Cantor sets. For extensive numerical data on both d_c, d_l, and the associated Feigenbaum exponents, see their paper.

4.8 Invariant measures, mixing and ergodicity

From an experimental viewpoint it is important to known something about the likelihood of observing a dynamical feature or characteristic of a system. This likelihood, or probability, is presumably related to some 'measure' of those states in the phase space which have the dynamical feature of interest.

By a *measure*, $\mu(A)$, of a set A in R^n we will mean a set function (i.e., a function which, for any set, gives a real number, $\mu: A \to R$) with the properties: (a) $0 \leqslant \mu(A) \leqslant \infty$, (b) $\mu(A \cup B) = \mu(A) + \mu(B)$ for any disjoint sets $A \cap B = \varnothing$ (the empty set), and (c) $\mu(\varnothing) = 0$. In particular, the measure of all set (e.g., all phase space) does not need to be bounded. If it is bounded, then μ can be normalized to unity.

A classic example where a measure is introduced is the Liouville theorem, where the interest is in the behavior of 'most' solutions rather than the possibly exceptional behavior of a set of states 'of measure zero'. Such a set makes no contribution to the integral invariant. In the general Liouville theorem the local measure, $\mathrm{d}\mu = \rho(x)\,\mathrm{d}x (x \in R^n)$, weights the region $\mathrm{d}x$ with a density, $\rho(x) > 0$, so as to make the total measure of any propagating region, $\Omega(t)$, a constant in time. More succinctly, this is simply called an *invariant measure*. If the dynamics is viewed as a diffeomorphism f^t, discussed in Chapter 2, and A is some set of points in the phase space ($A = \Omega(t)$ in the Liouville theorem), then this invariance of the measure of A can be written

$$\mu(A) = \mu(f^t A) \quad (\text{all } t).$$

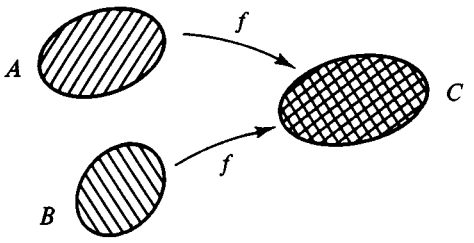
In other words the measure of the set A (which equals $\int_\Omega \rho(x)\,\mathrm{d}x$ in the Liouville theorem) is equal to the measure of the set of points which are obtained from the solutions of $\dot{x} = F(x)$, with the states in A as initial conditions.

However, in the case of maps which are not one-to-one, it is necessary to generalize this concept of an invariant measure. Assume that f is a two-to-one map in some regions of the phase space. For example, assume that the sets A and B both map onto the same region, C, so $f(A) = C$, $f(B) = C$ (Fig. 4.48). In this case we say that the inverse map of C, denoted by $f^{-1}(C)$, is the union of A and B.

$$f^{-1}(C) \equiv A \cup B. \tag{4.7.1}$$

This definition holds whether or not A and B are disjoint. In other words the inverse map of a set of points, C, is simply the set of all points which map into C. Using this

Fig. 4.48



definition we can now generalize the definition of an invariant measure to include the case of maps which are not one-to-one (not homeomorphisms). The measure is said to be an *invariant* of the map $f: R^n \to R^n$ if, for any set C

$$\mu(C) = \mu(f^{-1}C) \tag{4.8.2}$$

where $f^{-1}C$ are all the points that map onto C.

It is useful to consider some explicit examples of such *invariant measures*. One such example has already been obtained for the dynamics on an integral manifold, namely

$d\mu = \Delta(x_1, \ldots, x_{n-1})\, dx_1 \cdots dx_{n-1}$ in (2.4.5), where Δ is given by (2.4.9):

$$\Delta(K_0, x_1, \ldots, x_{n-1}) = \rho(x_1, \ldots, x_n^0(K_0, x_1, \ldots, x_{n-1}))/(\partial K/\partial x_n)_{x_n = x_n^0}.$$

Note, again, that this measure is defined on a manifold of dimension $(n-1)$ in R^n.

We now consider a second example that involves a two-to-one map, namely

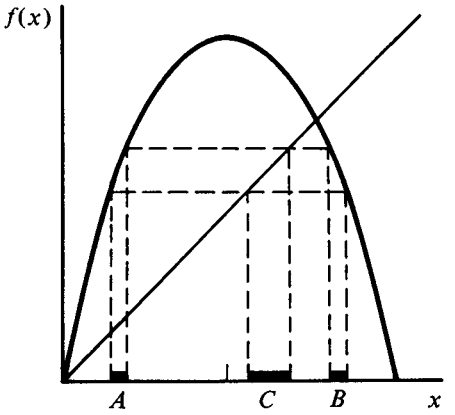
$$x_{n+1} = 4x_n(1 - x_n) \equiv f(x_n)$$

This is the logistic map, with $c = 4$, which is examined also in above exercises. Now consider the set

$$C = \{x : a \leqslant x \leqslant b\},$$

illustrated in Fig. 4.49. Since $x_n = \frac{1}{2} \pm \frac{1}{2}(1 - x_{n+1})^{1/2}$, the two regions which map into C

Fig. 4.49



are

$$A = \{x: \tfrac{1}{2} - \tfrac{1}{2}(1-a)^{1/2} \leqslant x \leqslant \tfrac{1}{2} - \tfrac{1}{2}(1-b)^{1/2}\}$$
$$B = \{x: \tfrac{1}{2} + \tfrac{1}{2}(1-b)^{1/2} \leqslant x \leqslant \tfrac{1}{2} + \tfrac{1}{2}(1-a)^{1/2}\}.$$

For $d\mu = \rho dx$ to be an invariant measure requires

$$\int_a^b \rho(x)\,dx = \int_{\frac{1}{2}-\frac{1}{2}(1-a)^{1/2}}^{\frac{1}{2}-\frac{1}{2}(1-b)^{1/2}} \rho(x)\,dx + \int_{\frac{1}{2}+\frac{1}{2}(1-b)^{1/2}}^{\frac{1}{2}+\frac{1}{2}(1-a)^{1/2}} \rho(x)\,dx.$$

Since this must hold for any region, we can differentiate with respect to b, obtaining the condition

$$\rho(b) = \tfrac{1}{4}(1-b)^{-1/2}\rho(\tfrac{1}{2} - \tfrac{1}{2}(1-b)^{1/2}) + \tfrac{1}{4}(1-b)^{-1/2}\rho(\tfrac{1}{2} + \tfrac{1}{2}(1-b)^{1/2})$$

which must hold for all $0 < b < 1$. From the symmetry of the map about $x = \frac{1}{2}$ it is clear that the two factors on the right are equal, so that what is required is a solution of

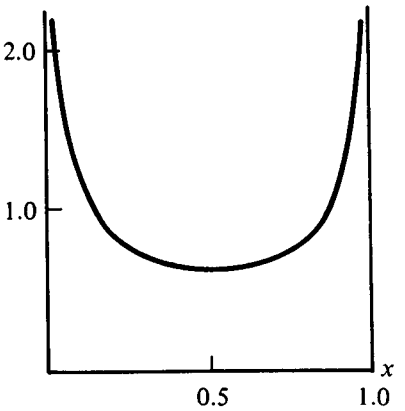
$$\rho(x) = \tfrac{1}{2}(1-x)^{-1/2}\rho(\tfrac{1}{2} - \tfrac{1}{2}(1-x)^{1/2}).$$

The solution of this is (obviously?)

$$\rho(x) = \alpha[x(1-x)]^{-1/2} \tag{4.8.3}$$

where α is an arbitrary constant, which may be selected to normalize the measure ($\int_0^1 \rho(x)\,dx = 1$, if $\alpha = \pi^{-1}$). The invariant density, $\rho(x)$, is illustrated in Fig. 4.50.

Fig. 4.50

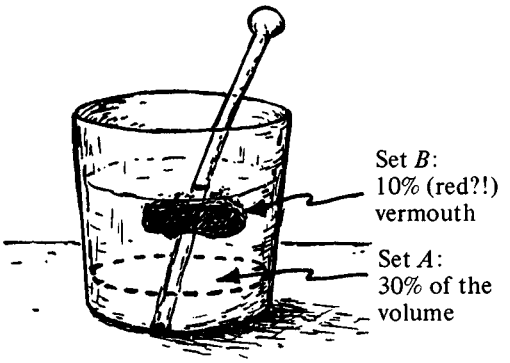


With the use of such invariant measures it is possible to give other descriptions of 'chaotic motion' which take into account the probability that this motion can actually be observed. One such indication of 'scrambled' motion is called *mixing*. A frequent example of mixing dynamics involves stirring a mixed drink, consisting of 20% rum and 80% Coca Cola (for Arnold and Avez, p. 19), or 90% gin and 10% vermouth (for P.R. Halmos, 'Lectures on ergodic theory', *Math. Soc. Japan*, 1956 pp. 36, 37). The drink is

really mixed if, in any fixed spatial region, we find that the fraction of rum in this region is 20%, or 10% vermouth in the stronger drink. This example is a dynamical case involving a one-to-one map (e.g., stroboscopic map of this flow), whereas we need to also include the case of maps which are not one-to-one.

Let μ be an invariant measure of a many-to-one map $f: R^n \to R^n$, and let A be an *arbitrary* fixed region (see Fig. 4.51). Let B be another arbitrary, but dynamic set, where

Fig. 4.51



both A and B have nonzero measure. The then dynamics of B that is generated by this map, $f^k(B)$ (all k), is said to be *mixing* if

$$\lim_{k\to\infty} \mu[f^{-k}(B) \cap A] = \mu(A)\,\mu(B). \tag{4.8.4}$$

Here the notation $A \cap B$ represents those points which are common to the sets A and B (their intersection). A dynamic system is said to be a *mixing system* provided that (4.8.4) holds for all sets A and B. If the map is one-to-one this can be 'run forwards' to produce the condition

$$\lim_{k\to\infty} \mu[f^k(B) \cap A] = \mu(A)\,\mu(B) \tag{4.8.5}$$

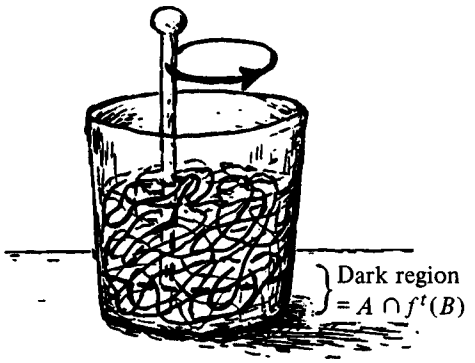
which is somewhat more transparent in its meaning. Thus in the case of Halmos' drink (Figs. 4.51 and 4.52), and for the region $\mu(A) = 0.3$, the mixing dynamics yields

$$\lim_{t\to\infty} \mu(A \cap f^t B) = 0.1 \times 0.3 = .03.$$

Note that these are normalized measures ($\mu(A) \leqslant 1$).

The property of mixing is a strong characterization of chaos. In particular, if a system is mixing then it is chaotic in the sense of Li and Yorke (Section 4.5), but the converse is not necessarily true (e.g., see Oono and Osikawa, 1980). Moreover, if a system is mixing it is necessarily ergodic. A system is *ergodic* under a map f if the measure of every *invariant set*, $A = f(A)$, is either $\mu(A) = 1$ or $\mu(A) = 0$. In other words the only invariant sets either have measure zero, or else represent all states except for a set of measure zero. The importance of ergodicity is presently unclear, but it is at least doubtful whether this

Fig. 4.52



property is either necessary or sufficient for many properties of statistical equilibrium or irreversible properties of systems. It is also unclear whether the strong property of mixing (if, in fact, it exists for real systems) is necessary for the existence of these statistical properties, but it appears to be more relevant. In any case it is a widely investigated property of systems.

To illustrate some of these ideas with a simple dynamic system, we consider a flow (rather than a map) in phase space, given by

$$x = a \sin\theta, \quad \dot{x} = a^2 \cos\theta \quad (\text{all } a \geqslant 0), \tag{4.8.6}$$

where $\dot{a} = 0$, and $\dot{\theta} = a$ is an arbitrary constant satisfying $a \geqslant 0$. This system is a nonlinear oscillator, with a frequency dependent on the amplitude.

Exercise 4.14. Show that a time independent constant of the motion for this oscillator is

$$K(x, \dot{x}) = x^2 + [x^4 + 4\dot{x}^2]^{1/2}.$$

On the manifold $K(x, \dot{x}) = K_0(a)$, what is the function $K_0(a)$? The manifolds ($a = 1.5$) are illustrated in Fig. 4.53. The flow (4.8.6) satisfies $\ddot{x} = -a^2 x$. Why isn't the system a harmonic system? What is the proper second order equation of motion for this system?

On an integral manifold, $K(x, \dot{x}) = K_0$, the arclength$_{\theta_1(t)}^{\theta_2(t)} \mathrm{d}\theta$ is invariant in time, so $\mathrm{d}\mu = \mathrm{d}\theta/2\pi$ is an invariant (normalized) measure on this manifold. Let A be a fixed segment ($\theta_1 \leqslant \theta \leqslant \theta_2$) on $K(x, \dot{x}) = K_0$, and B be a dynamic segment ($\theta_3(t) \leqslant \theta \leqslant \theta_4(t)$). How does $\mu[A \cap f^t B]$ behave as a function of time in this case? It is illustrated in Fig. 4.54, which you should justify for yourself (note that the arclengths are large in this case). Clearly, in this case, this function has no limiting value as t increases, since it is periodic in time, so (4.8.5) is not satisfied. Hence the system is not mixing on integral manifolds (not surprising, since this is R^1). Note, however, that this system is (trivially) *ergodic* on its integral manifolds. On the other hand, the measures of these segments in

Fig. 4.53

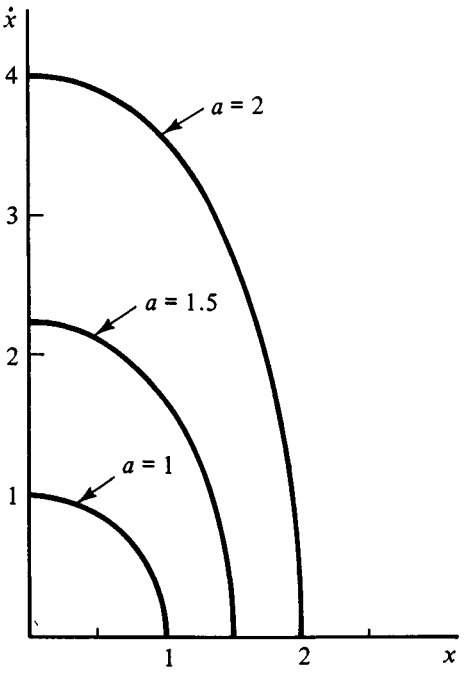
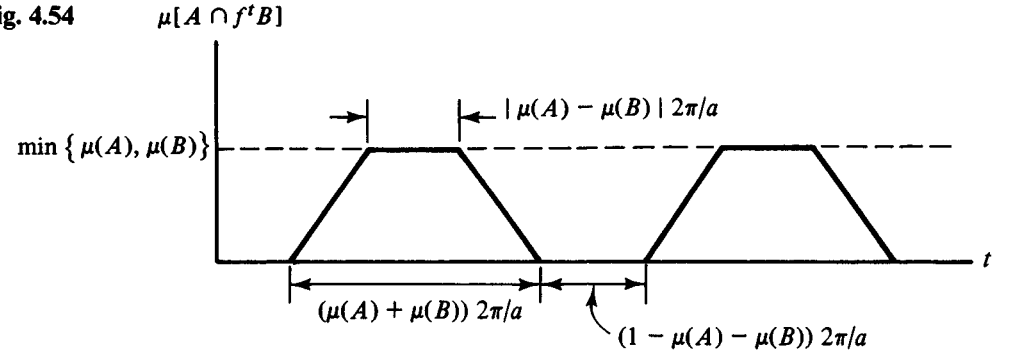


Fig. 4.54



the phase space (R^2) are both zero, so the above result is not relevant to the question of whether (4.8.6) is a mixing system in R^2. These results are, however, useful in what follows.

Next, let us consider some two dimensional regions in the phase space. We note that the measure $d\mu = d\theta da/2\pi(a_2 - a_1)$ is invariant (here $a_1 \leqslant a \leqslant a_2$ is the normalizing range).

Exercise 4.15. Obtain this invariant phase plane measure of (4.8.6) in the form $d\mu = \rho(x, \dot{x})\, dx d\dot{x}$ (i.e., involving only, $x, \dot{x}$, so that $\rho(x, \dot{x}$ is a phase function).

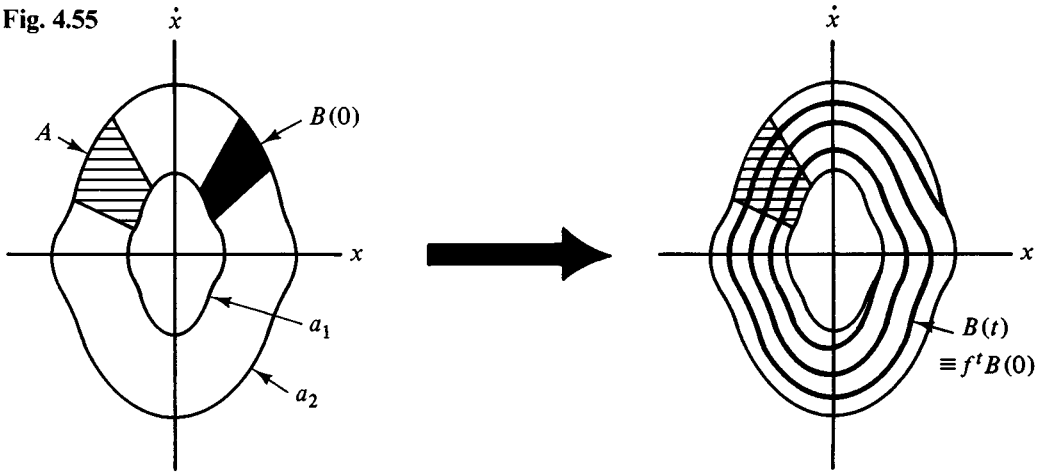
Let A and B be the sets

$$A = \{x, \dot{x} | a_1 \leqslant a \leqslant a_2, \theta_1 \leqslant \theta \leqslant \theta_2\}$$

$$f^t B(0) \equiv \{x, \dot{x} | a_1 \leqslant a \leqslant a_2, \theta_3(t, a) \leqslant \theta \leqslant \theta_4(y, a)\}.$$

Now what happens is much more interesting. Because the arclengths at different values of a rotate at different rates, an initially wedge-shaped region, $B(t = 0)$, will become transformed into a spiral region between the manifolds $K(x, \dot{x}) = K_0(a_1)$ and $K(x, \dot{x}) = K_0(a_2)$. This is illustrated schematically in Fig. 4.55. A little thought will hopefully

Fig. 4.55



convince you that now the function $\mu[A \cap f^t B]$ will settle down to a limiting value as $t \to \infty$. Moreover, it should be clear that this limit is $\mu(A)\mu(B)$, so that (4.8.5) is satisfied.

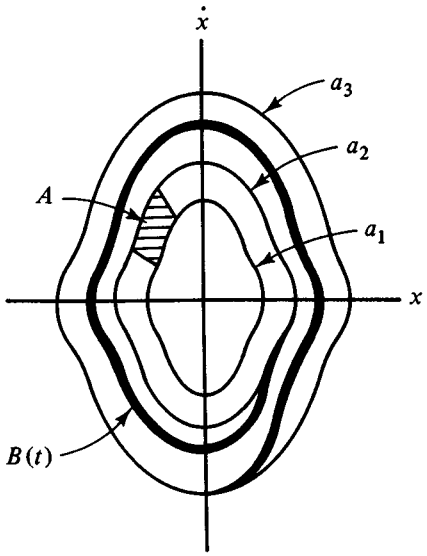
The rate at which the mixing occurs (that is, the time scale required for (4.8.5) to become nearly satisfied) depends on the difference in the two frequencies, a_1, a_2. Typically one expects this relaxation time to be of the order of $(a_2 - a_1)^{-1}$. In the limit $a_2 \to a_1$, this relaxation time goes to infinity, so there is no relaxation, which was our first example above.

Now, although the last two regions A and $B(t)$ do satisfy (4.8.5), this does not mean that the system (4.8.6) is a mixing system in R^2. Indeed the system is not a mixing system, because it does not satisfy (4.8.5) for all A and $B(t)$ in the phase space. This is so because the integral of the motion, $K(x, \dot{x})$, is an isolating integral in R^2, for it is a smooth function of the phase variables. For example, if $B(t)$ is the region

$$B(t) = \{x, \dot{x} | a_2 < a < a_3, \theta_3(t, a) \leqslant \theta \leqslant \theta_4(t, a)\}$$

then $A \cap B(t) = 0$ for all t (see figure), and obviously (4.8.5) is not satisfied. For the system to be mixing in R^2, the region $B(t)$ must become distributed over all of the

Fig. 4.56



phase space (Fig. 4.56). Therefore the system (4.8.6) is neither mixing on a manifold $K(x, \dot{x}) = K_0$ (i.e., in R^1), nor in the full space R^2.

In higher dimensional phase spaces, it is possible to have isolating integrals, and still have mixing on their integral manifolds. The above system has a phase space with too low a dimension for this mixing to be possible on the integral manifolds.

Classic examples of mixing maps in R^2 are given in Chapter 6.

4.9 The circle map

In this section we will consider a map which has an entirely different physical origin from the logistic map, and which exhibits correspondingly different effects. We are interested here in the influence of one ('driver') oscillator when coupled to a second oscillator, and will attempt to model this effect by a simple one-dimensional map. The dynamics will be governed by two control parameters, the frequency ratio of the (uncoupled) oscillators, Ω, and their coupling strength, K. Moreover the map will be one-to-one for $K < 1$, but is two-to-one if $K > 1$, in which case it has many of the chaotic properties of the logistic map. Since we have already discussed these chaotic properties in some detail, we will largely concentrate here on the one-to-one mapping region, $K < 1$. For more details, see Lanford (1987).

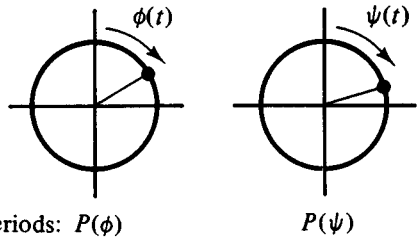
Let ϕ and ψ be the phase angles of two oscillators. If they are decoupled, then their motion can be represented on a torus, T^2, in 'our R^3', as discussed in Chapter 2. Taking a cross section of this torus, we can obtain a Poincaré map for the phase angle of one of these oscillators. This map occurs on a (topological) circle, or one-torus T^1.

We now distinguish between the two oscillators, treating the ψ-oscillator as a

prescribed 'driver' of the other, and proceed to study the Poincaré map of the 'influenced' ϕ-oscillator. The physical nature of the ϕ-oscillator is not specified in the present study, but is implicit in the following models. This will be discussed briefly at the end of this section.

Let $(P(\phi), P(\psi))$ be the periods of the uncoupled (ϕ, ψ) oscillators (Fig. 4.57). The

Fig. 4.57



successive points of the Poincaré map are $\phi_n = \phi(t = nP(\psi))$, and when the oscillators are not coupled, $\phi_{n+1} - \phi_n = 2\pi P(\psi)/P(\phi)$ (Fig. 4.58), and the Poincaré map is simply

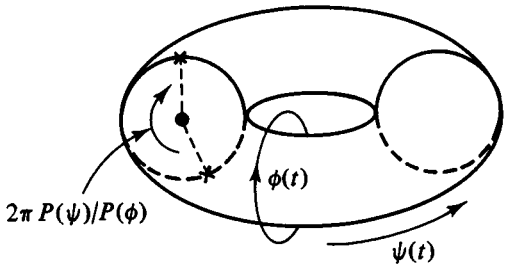
$$T\phi \to \phi + \Phi$$

or

$$\phi_{n+1} = \phi_n + \Phi, \quad \Phi = 2\pi P(\psi)/P(\phi). \tag{4.9.1}$$

This is illustrated in Fig. 4.58. When there is coupling between the oscillators, this map

Fig. 4.58



is modified to

$$\phi_{n+1} = \phi_n + \Phi + F(\phi_n) \equiv \phi_n + t(\phi_n), \tag{4.9.2}$$

where the function $F(\phi)$ must satisfy $F(\phi + 2\pi) = F(\phi)$ for all ϕ. This simply expresses the fact that the effect of the ψ-oscillator on the same phase point of the ϕ-oscillator must be the same (the notion is on T^2). Note that, while ϕ and $\phi + 2\pi$ are the same phase points, the *cumulative* values of the ϕ_n $(n = 1, 2, \ldots)$ are generated by

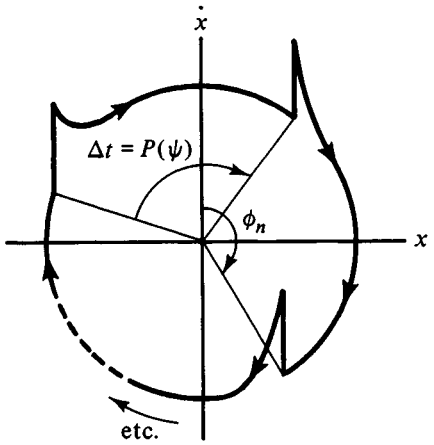
$$T\phi \to \phi + \Phi + F(\phi) \equiv \phi + t(\phi). \tag{4.9.3}$$

In other words $\phi \in R^1$, and therefore can tend to infinity.

Note also that the shift $t(\phi)$ in (4.9.1) is generally only defined up to a multiple of 2π. However, once it is fixed for some ϕ, it remains unique by continuity. The origin of this ambiguity in $t(\phi)$ reflects a physical ambiguity inherent in this Poincaré map; namely, this map does not indicate how many full periods of the ϕ-oscillator occur in the time $P(\psi)$. This physical information must be implicitly contained in the function $t(\phi)$ – because this function reflects the influence of the ψ-oscillator on the ϕ-oscillator for some specific rotational motion of the latter oscillator. This physical fact is not explicit in the present models of coupled oscillators.

To illustrate a simple example of the coupled map, (4.9.2), consider the case of a ϕ-oscillator which tends to rapidly (relative to $P(\psi)$) return to its unperturbed orbit after it is perturbed by an outside source. Such an oscillator is called a relaxation oscillator, and will be described in more detail in the next chapter (we are ahead of the differential equations story!). Next, assume that the ψ-oscillator acts on the ϕ-oscillator impulsively (instantaneously) every $P(\psi)$ seconds, changing its velocity. Moreover, assume that $P(\phi) > P(\psi)$. The resulting motion of the ϕ-oscillator then looks something like shown in Fig. 4.59. If the unperturbed motion is circular in the phase space, $(x, \dot{x})$,

Fig. 4.59



then 'spikes' are added, because of the impulsive interaction (changing $\dot{x}$, but not x). Only three impulses are shown in the figure, for simplicity. The phase shifts caused by the impulsive at ϕ_n is $F(\phi_n)$ in (4.9.2), which can clearly be positive or negative, depending on whether it advances or retards ϕ relative to its unperturbed motion (in the figure, two impulses might produce $F < 0$, and one might yield $F > 0$).

A frequently used form of the map (4.9.2), which may bear some approximate relationship to the above coupled oscillators, is

$$\phi_{n+1} = \phi_n + \Phi - K \sin(\phi_n). \tag{4.9.4}$$

Sometimes the substitution

$$\phi_n = 2\pi\theta_n, \quad \Phi = 2\pi\Omega \tag{4.9.5}$$

is used, so that

$$\theta_{n+1} = \theta_n + \Omega - \left(\frac{K}{2\pi}\right)\sin(2\pi\theta_n). \tag{4.9.6}$$

Ω is not, of course, an angular frequency (despite its notation), but a pure number

$$\Omega \equiv P(\psi)/P(\phi), \tag{4.9.7}$$

the ratio of the periods of the two uncoupled oscillators. The maps (4.9.4) and (4.9.6) are known as *circle maps*. These maps are one-to-one provided that $\mathrm{d}(T\phi)/\mathrm{d}\phi > 0$ for all ϕ, and since $T\phi = \phi + \Phi - K\sin\phi$, this requires

$$1 > K \quad \text{(one-to-one)}. \tag{4.9.8}$$

Examples of oscillators to which this model has been applied are the heartbeat (Glass and Perez, 1982) and the Rayleigh–Bénard vortical motion in fluids (Fein, Heutmaker, and Gollub 1985), and optical devices. Applications of the circle map to other systems (Josephson junctions, charge density waves) have also been proposed (Bohr, Bak and Jensen, 1984).

Exercise 4.16. Refer to the last figure and the circle map (4.9.4). How is K related to the figure? What physical conditions are required for the relaxation properties of the ϕ-oscillator, and for the value of Ω, in order for (4.8.4) to be a (possibly) reasonable approximation?

We use, as usual, the notation

$$T^{n+1}\phi = T(T^n\phi) = \phi_{n+1} \tag{4.9.9}$$

for the iterates of the map. Poincaré introduced the concept of a *rotation number*, defined by

$$2\pi\rho \equiv \lim_{n\to\infty}\frac{T^n\phi - \phi}{n} \tag{4.9.10}$$

which is independent of ϕ. In terms of the variable θ, (4.8.5), the related concept is called the *winding number*

$$W(K,\Omega) = \lim_{n\to\infty}\frac{\theta_n - \theta_0}{n}, \tag{4.9.11}$$

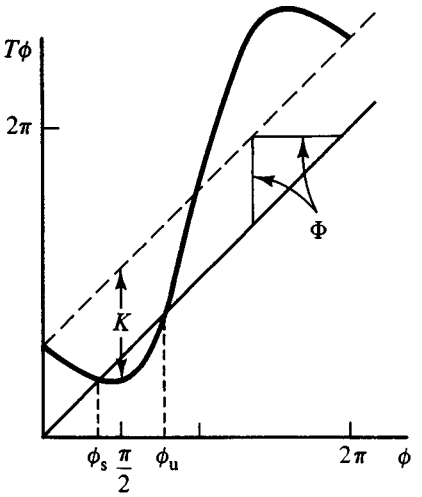
which is simply related to ρ, (4.9.10) by

$$W(K,\Omega) \equiv \rho(K, \Phi = 2\pi\Omega). \tag{4.9.12}$$

The rotation number, $2\pi\rho$, is the time-averaged angular rate of rotation of the ϕ-oscillator. If $K = 0$, then $\phi_n = \phi_0 + n\Phi$, and $2\pi\rho = \Phi$, by (4.9.10). Therefore, in this case

the rotation number is simply the angular step at each iteration, and $\rho = \Omega = W$. When $K \neq 0$ the rotation rate differs from Φ. This is most readily illustrated at the other extreme, when $K > \Phi$ (Fig. 4.60). From (4.9.4) it can be seen that, for some values of ϕ_n, the

Fig. 4.60



map yields $\phi_{n+1} < \phi_n$, so that the rotation is reversed in this region of ϕ. Also, if we plot $T\phi$ vs. ϕ, it is clear that $T\phi$ has two fixed points, ϕ_s (stable) and ϕ_u (unstable). The stability properties follow from

$$\frac{\mathrm{d}}{\mathrm{d}\phi}(T\phi) = 1 - K\cos\phi$$

which is < 1 (> 1) if $\phi < \pi/2$ (respectively, $\phi \gtrsim \pi/2$). Since

$$\lim_{n\to\infty} \phi_n = \phi_s \quad (\text{if } \phi_0 \neq \phi_u),$$

it follows that

$$2\pi\rho = \lim_{n\to\infty} \frac{\phi_n - \phi_0}{n} = 0$$

for all ϕ_0. In other words, asymptotically in time, the points of the Poincaré map do not rotate about the center of the circle (they go to fixed points). Thus, if

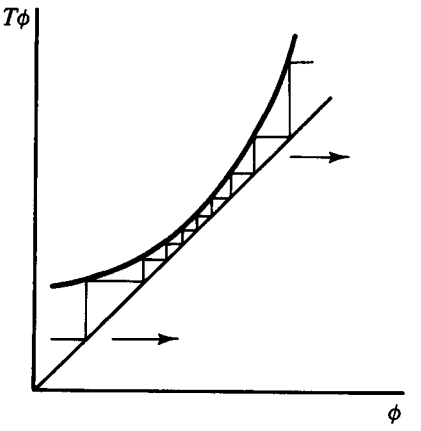
$$1 > K > \Phi, \qquad \text{then } \rho = 0$$

or

$$1 > K/2\pi > \Omega, \quad \text{then } W = 0 \tag{4.9.13}$$

When $K = \Phi$, it can be seen from Fig. 4.60 that $\phi_s = \phi_u = \pi/2$. The mappings in the neighborhood of this fixed point all move 'forward' ($\phi_{n+1} > \phi_n$), and Arnold referred to such points in any cyclic orbit as (forward) *semistable points*. They clearly occur at bifurcation points (the catastrophes set of K, Φ), Fig. 4.61.

Fig. 4.61

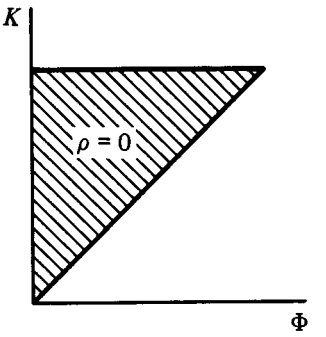


In general, Poincaré proved that if $\rho = m/n$, where m and n are relatively primed integers (ρ is a rational number), then there exists solutions of

$$T^n\phi = \phi + 2\pi m. \tag{4.9.14}$$

Indeed there are usually $2n$ such solutions, which alternate between the n stable values $(\phi_1^s, \ldots, \phi_n^s)$ and the unstable set $(\phi_1^u, \ldots, \phi_n^u)$, where $T\phi_j = \phi_{j+1}$. These sets are called cycles of T. A rotation number $\rho = m/n$ will only occur for a particular set of maps $T(K, \Phi)$, which Arnold referred to as the *level set* of $\rho(T)$. The level set of $\rho = 0$, which we just determined, (4.9.13), is illustrated in Fig. 4.62. Poincaré stated [for a proof, see

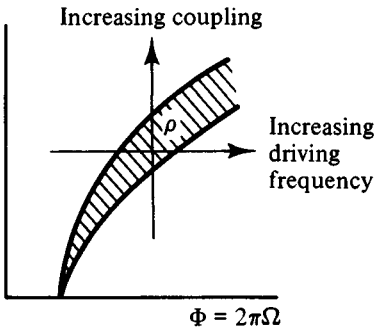
Fig. 4.62



Arnold (1965)] that $\rho(T)$ is a continuous function of T, so $\rho(K, \Phi)$ is a continuous function in the special case of the circle map. Thus the level sets are all continuous regions in the (K, Φ) plane, and have simple boundaries.

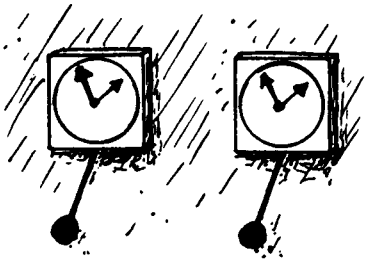
From a physical point of view, a level set may be considered in several different ways. Consider, for example, increasing the frequency of the ψ-oscillator, while maintaining a constant coupling, K, to the ϕ-oscillator (Fig. 4.63). If $\rho = m/n$ over some range of $P(\psi)$, this means that the ϕ-oscillator tends to a cyclic mode which is '*locked in*' to the driving oscillator's period (ρ remains constant as $P(\psi)$ is varied). The phenomena is called the

Fig. 4.63



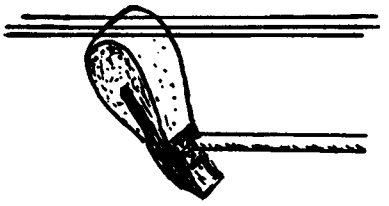
entrainment of the ϕ-oscillator by the ψ-oscillator, and is discussed further in Chapter 5. A classic (17th century) example of this entrainment was Christian Huygens' observation that when two clocks are hung back to back on a wall (Fig. 4.64), they tend

Fig. 4.64



to synchronize their motion (noted by van der Pol, *Phil. Mag.* **3**, 65–80 (1927)). A more recently discovered application of entrainment is in coupled piano strings (Fig. 4.65),

Fig. 4.65



which are intentionally mistuned, to produce a sustained aftersound – but we will save this for Chapter 5, where other physiological examples will also be noted.

While it is less common in practice, we might also take two oscillators, with fixed physical properties (specified by Φ), and increase the coupling constant between them. Once again, we may obtain entrainment at a given value of ρ only for a limited range of K, and another entrainment value may occur for larger K.

Exercise 4.17. It is useful at this point to give a physical interpretation of the level set of $\rho = 0$. Specifically, consider the $P(\phi) \gg P(\psi)$, with $K > \Phi$, and interpret the meaning of

$\rho = 0$. Make the physical interpretation again, if $P(\psi) = P(\phi)$ (see the discussion following (4.8.32)). Prove that generally,

$$\rho(\Phi, K) + l = \rho(\Phi + 2\pi l, K) \quad (l\text{:any integer}).$$

The boundaries of the level sets occur where the stable and unstable solutions become semistable points, ϕ^{ss}. This occurs when the slope of $T^n\phi$ equals one at ϕ^{ss} (see our previous tangency condition). Hence the boundary of the level set $\rho = m/n$ is given by

$$T^n\phi = \phi + 2\pi m, \tag{4.9.14}$$

together with the tangency requirement

$$[\mathrm{d}(T^n\phi)/\mathrm{d}\phi]_\phi = 1,$$

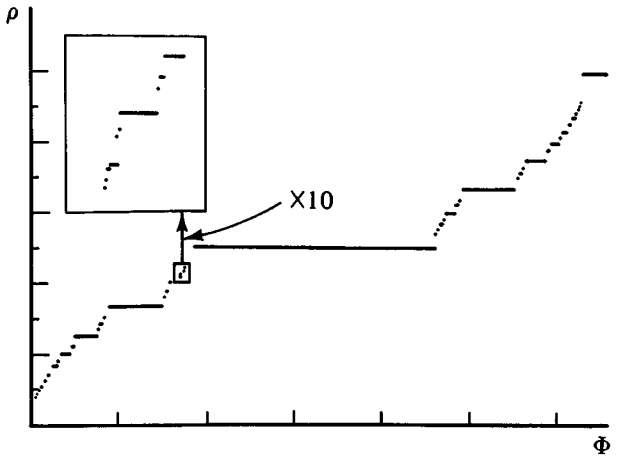
which is the same as (recall (4.2.11)),

$$\prod_{j=0}^{n-1} (\mathrm{d}T/\mathrm{d}\phi)_\phi = 1, \tag{4.9.15}$$

where $\phi_j = T^j\phi$ are the semistable points.

These level sets are separated by irrational ρ values. In any neighborhood of an irrational value of ρ is a level set, because the rational values of ρ are dense. Thus the function $\rho(K, \Omega)$ has a very strange property, as a function of Φ, which is schematically illustrated in Fig. 4.66. ρ is a continuous function, which is constant in some finite

Fig. 4.66



interval of Φ, whenever ρ takes on a rational value, $\rho = m/n$ (read the sentence again!). That is indeed a strange function, which is an example of functions known as *Cantor–Lebesgue functions*. The schematic figure of ρ (it can't really be drawn!) is known, quite appropriately, as the *'devil's staircase'*.

To illustrate these level sets, for $\rho = \frac{1}{2}$ the equations (4.9.14) and (4.9.15) yield

$$2\pi = 2\Phi - K\sin\phi - K\sin(\phi + \Phi - K\sin\phi)$$

$$1 = (1 - K\cos\phi)(1 - K\cos(\phi + \Phi - K\sin\phi).$$

Expanding Φ and $(\sin\phi, \cos\phi)$ in powers of K, one finds, after some labor, the relationship

$$\Phi = \pi \pm (K^2/4) + O(K^4). \tag{4.9.16}$$

For $K = 1$, the actual boundary value is $\Phi = \pi \pm 0.232\,37\ldots$ (Arnold), compared with this result of $\pi \pm 0.25$. Arnold reported that a similar calculation for the boundaries of the level set $\rho = \frac{1}{3}$ yields

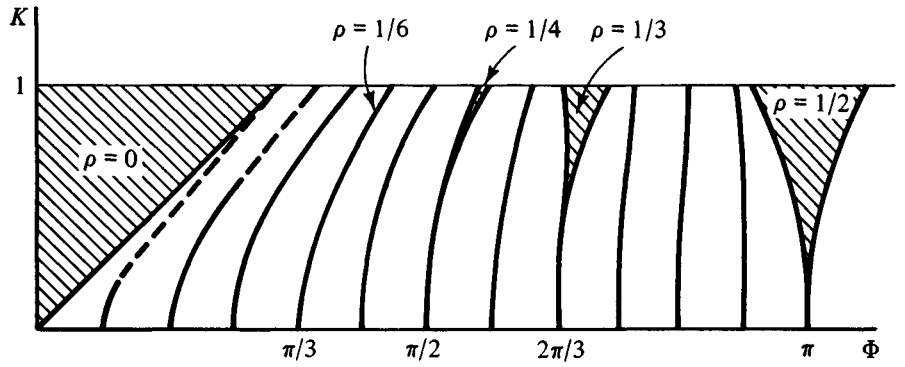
$$\Phi = \frac{2\pi}{3} + \frac{3^{1/2}}{12}K^2 \pm \frac{7^{1/2}}{24}K^3 + O(K^4). \tag{4.9.17}$$

Generally, for $\rho = m/n$

$$\Phi = 2\pi m/n + O(K^{n-1})$$

but the leading term may vanish, as in (4.9.16). In any case the level sets are very narrow for larger values of n. It is because of the rapid decrease in the range of the level sets with increasing n that there remains a finite measure for the irrational values of ρ. Their structure is illustrated in Fig. 4.67, which is adopted from Arnold. These structures

Fig. 4.67



have since become known as 'Arnold tongues'. Note that Arnold's figure may be extended using (see Exercise 4.16)

$$\rho(\Phi, K) + l = \rho(\Phi + 2\pi l, K).$$

When $K > 1$, the circle map becomes two-to-one in regions of ϕ, and period doubling bifurcation series, together with many other bifurcations can occur. These

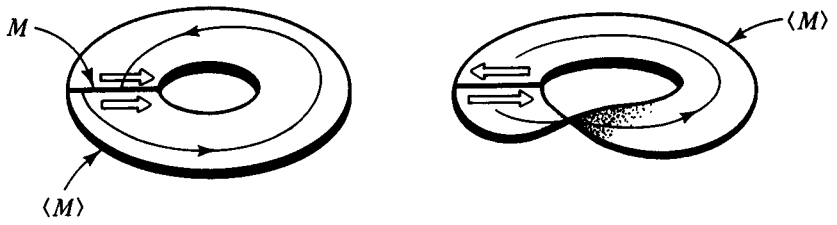
studies go beyond our present concerns, but can be found discussed in the literature (e.g., Ostlund *et al.*, 1983, Aronson *et al.*, 1982, Bélair and Glass, 1985).

4.10 The 'suspension' of a tent map

In the introduction of this chapter, we considered briefly the possible relationship between maps and the continuous dynamics (flows) in the phase space associated with a system of ODE. The reason that this may be of interest is that physical systems are frequently modeled by ODE. It is nice, therefore, to have some insight into the possible relationshp (map → flow) in order to better understand the possible relationship between a map and physical systems. In particular we can determine how complicated the system must be (e.g., the dimensions of the phase space R^n).

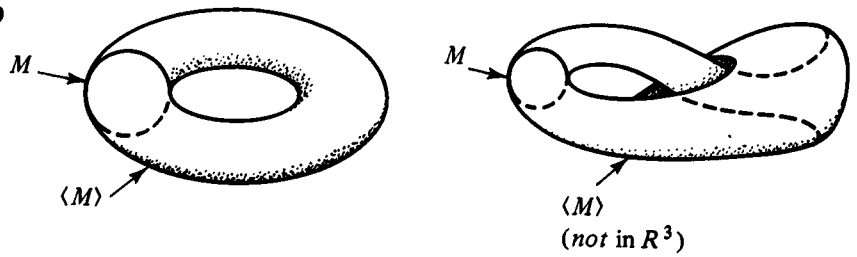
When the map is a diffeomorphism (one-to-one, etc.) then it can always be regarded as the first-return map of a flow in some manifold, by a very simple operation, called a *suspension* (Nitecki, 1971). The idea is that, if the diffeomorphism, f, maps the manifold M onto itself, $f: M \to M$ then we construct a new manifold, $\langle M \rangle$, which is the product space of M and the unit interval $I = (0, 1)$, $\langle M \rangle = M \times I$, and satisfying the relationship $(x, 0) = (f(x), 1)$. Thus, if $M = I$ (the unit interval), then $\langle M \rangle$ is either an annulus or a Mobius strip (Fig. 4.68), depending on whether the map $f(x)$ is orientation preserving or

Fig. 4.68



not. Similarly, if $M = S^1$ (a circle), the new manifold, $\langle M \rangle$, is either a two-torus, T^2, or else a Klein bottle (Fig. 4.69), again depending on whether or not $f(x)$ is orientation

Fig. 4.69



preserving. While this construction of a flow is nice from a mathematical point of view, the resulting flow on the manifold $\langle M \rangle$ may not be clearly related to any physical system of interest. That, of course, requires a separate investigation.

When a map $f(x)$ is continuous, but two-to-one, it is much more difficult to find a flow in some manifold M^* for which $f(x)$ can be related to (but is not) a Poincaré map. To illustrate the problems which are involved, we follow Haken's study (1985) of this question for a tent map. Following the reasons discussed in the introduction of this chapter, it is clear that a two-to-one map cannot be a first-return map of a flow in some other manifold. This is simply because flows have unique orbits through each point. Therefore a two-to-one map can only be related to one coordinate of a two or more dimensional surface of section, rather than the entire surface of section. Therefore chaotic motion can only result from a flow in three or more dimensions.

Let us construct a rectangular region, with one side parallel to the x-coordinate, which is mapped according to some tent map (4.2.2). The other coordinate we will denote as z. The question is, can this rectangular region have a continuous one-to-one first-return map

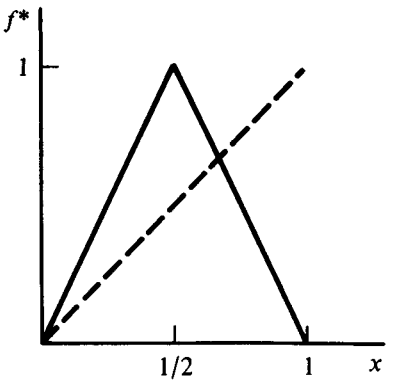
$$x' = f^*(x, z), \quad z' = g^*(x, z) \quad (0 \leqslant x \leqslant 1, 0 \leqslant z \leqslant 1) \tag{4.10.1}$$

which is produced by some flow in the R^3 phase space (x, y, z), and is such that x maps according to the tent map,

$$f(x) = f^*(x, z) = \begin{matrix} 2x (\text{if } 0 \leqslant x \leqslant \frac{1}{2}) \\ 2(1-x) (\text{if } \frac{1}{2} \leqslant x \leqslant 1) \end{matrix} \tag{4.10.2}$$

for all $0 \leqslant z \leqslant 1$? Here we take the tent map which maps $[0, 1]$ onto itself (as in the case of the logistic map with $c = 4$), Fig. 4.70.

Fig. 4.70



The flow is required to take all points $(0 \leqslant x \leqslant \frac{1}{2}, z)$ onto the interval $0 \leqslant x \leqslant 1$, suitably changing z so as to leave room for the other half of the rectangular region, $x \geqslant \frac{1}{2}$ (Fig. 4.71). A simple (area preserving) way of accomplishing this is to cut the rectangle along $x = \frac{1}{2}$, and put a 'hinge' at $x = \frac{1}{2}, z = 1$. As the points flow around the z axis the hinge moves to $(x = 1, z = \frac{1}{2})$ and the initial region $(\frac{1}{2} \leqslant x \leqslant 1, 0 \leqslant z \leqslant 1)$, rotates around the hinge, so that it is mapped onto $1 \geqslant z > \frac{1}{2}$. The z distances are

Fig. 4.71

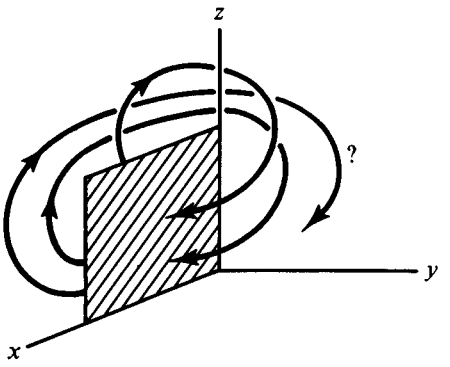
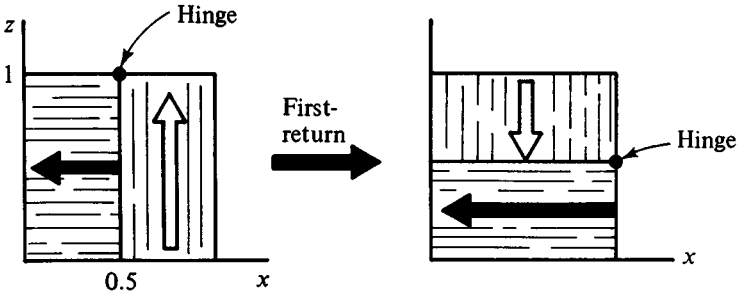


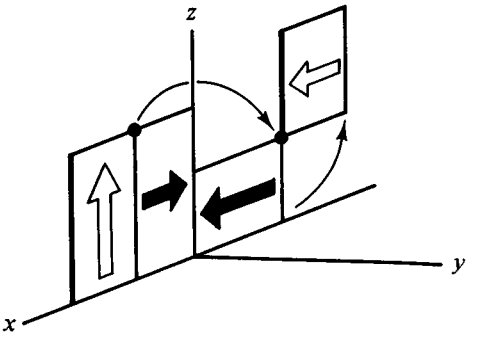
Fig. 4.72



compressed by a factor $\frac{1}{2}$, as the x distances increasing by a factor 2 (Fig. 4.72).

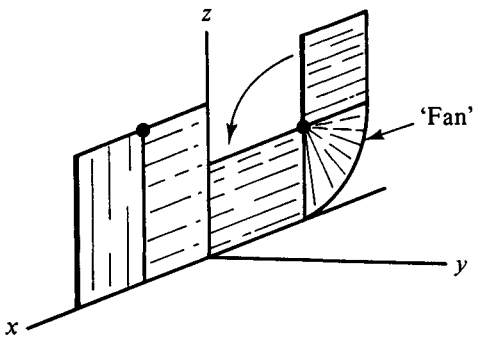
This proposed flow is, of course, not unique. There might be (for example) any number of complete rotations of the present entire flow about any periodic 'hinge' circling the z axis (Fig. 4.73).

Fig. 473



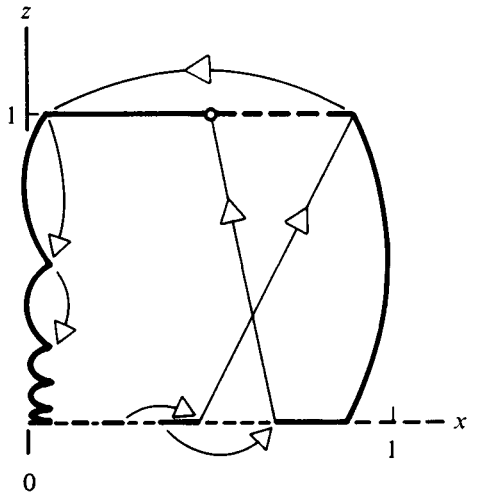
The above proposed flow does not, however, satisfy continuity nor closure in R^3. The first difficulty is there cannot be a cut associated with a flow that depends continuously on the initial conditions. That is, Fig. 4.73 must look more like shown in Fig. 4.74

Fig. 4.74



(containing some 'fan' which preserves continuity). A second problem is that the two boundaries which are coming together (see Fig. 4.74) cannot both be closed, for this would lead to a two-to-one map of the points on these boundaries, and that cannot occur for unique flows. To overcome these difficulties for a flow in R^3, appears to be possible only if the Poincaré surface of section has a fractal boundary structure, such as the one schematically illustrated in Fig. 4.75. Part of the fractal structure arises from the

Fig. 4.75



'continuity fan', which produces a self-similar fractal structure on the left side. The other fractal aspect is the closed (solid line) and open (dashed line) boundaries, which is essentially self-similar on the lower 'edge'. The lower left hand corner of this boundary must apparently be a sophisticated 'intersection' of two types of fractals.

These results indicate, that in order to obtain a physically meaningful system, which has a Poincaré surface of section yielding a tent-like two-to-one map, requires that the flow occur in R^n, with $n \geqslant 4$. In other words, physically realistic systems which have the dynamic complexity of a logistic map, apparently must have a dynamical dimension greater than, or equal to, four. (Also see Sverdlove (1977), Mayer-Kress and Haken (1987)).

4.11 Mathematics, computations, and empirical sciences

Even at this early stage in our exploration of nonlinear dynamics, we encounter an issue which may be one of the most basic lessons, which will come from the study of nonlinear dynamics. The general issue concerns the interface between mathematics, computations, and empirical sciences. The present specific issue is related to the complicated, 'chaotic' dynamics which we have encountered in the simple two-to-one maps of this chapter. Many features of this dynamics were uncovered because we had the help of computers to perform the necessary iterations. In certain select cases, such as the logistic map with $c = 4$, we could mathematically establish that chaotic motion does occur, by determining that the Lyapunov exponent is positive, and actually obtain its (precise) value. Such cases are rare indeed, but that is not the issue we will consider here.

What we are concerned with at present is the fact that both empirical sciences and computations are fundamentally restricted to observations/computations involving a finite amount of information. We can only write down so many numbers in 'our lifetime'! In other words, these occupations deal with THE FINITE – some finite 'precision' (rational numbers), manipulated for some finite amount of time. By way of contrast, as Leibniz expressed it, mathematics is the science of THE INFINITE, which contains the concept of the continuum, and such related concepts as limit point, derivatives, irrational numbers, and Cantor sets. These concepts belong to the realm of THE INFINITE, and of mathematics, not to THE FINITE, which is the realm of computations and empirical sciences.

The idea that computations and empirical sciences should be intimately related, has been concisely expressed by Fredkin (1982) in the form of the postulate: 'There is a one-to-one mapping between what is possble in the real world, and what is theoretically possible in the digital simulation world,' and the corollary 'That which cannot, in principle, be simulated on a computer, cannot be part of physics', where 'physics' presumably means any physical system. Such statements can easily stimulate long discussions, some of which we will return to in Chapter 10.

The point we will discuss here is that the 'chaotic' motion, discovered in this chapter, brings THE INFINITE and THE FINITE face to face (Fig. 4.76)! It does so because

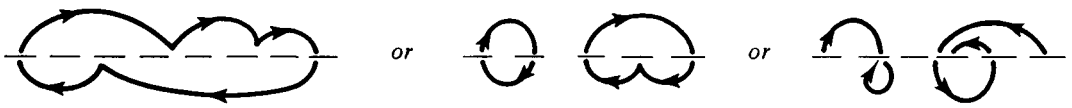
Fig. 4.76

the various forms of 'chaos' which we have defined mathematically were naturally based on concepts from THE INFINITE, whereas our computations and observ-

ations fall within the realm of THE FINITE. In most dynamics, which involves regular forms of motion, this distinction is not so clearly apparent because the dynamics never explores the 'fine structures' contained in THE INFINITE, particularly when only finite times are considered. Thus, even for regular motion, infinitesimal distances can be explored, as in the Poincaré recurrence theorem, but this involves arbitrarily long intervals of time. What the chaotic motion of THE INFINITE does is to force us to consider whether, within a very finite time, our actions in THE FINITE have any relationship to these theoretical ideas from THE INFINITE. It is clearly a very basic and important issue, about which there is presently only a very limited understanding. However, since this issue may ultimatey influence our understanding of nature itself, it is certainly worth a brief discussion, despite the present very incomplete insight.

To be much less philosophical, let us consider some really finite dynamics, say a system with only five possible states. We would probably all agree that the deterministic dynamics of such a system cannot be chaotic, in any 'real sense', because after no more than five 'steps' the motion must repeat. In other words, all motion is ultimately periodic, possibly ergodic (left example), or containing several 'short' periods (middle), or several basins of attraction (right) (Fig. 4.77). But what if the system has

Fig. 4.77



100 states, or 100,000 states, or 10^7 states? Does it make empirical sense to talk about a system with 10^{100} states, even if it is finite? (A.N. Kolmogorov, 1984, indeed suggested that all numbers may be meaningfully, even if not precisely divided into small, medium, large, and extra-large numbers.) Which of these systems can have 'chaotic' or 'complex' forms of dynamics, or can we not make such a distinction in the realm of THE FINITE? The question is rhetorical, of course, since we presently have no generally accepted definitions for 'chaos' or 'complexity' within THE FINITE.

There are a variety of topics that can be raised in this area, such as:

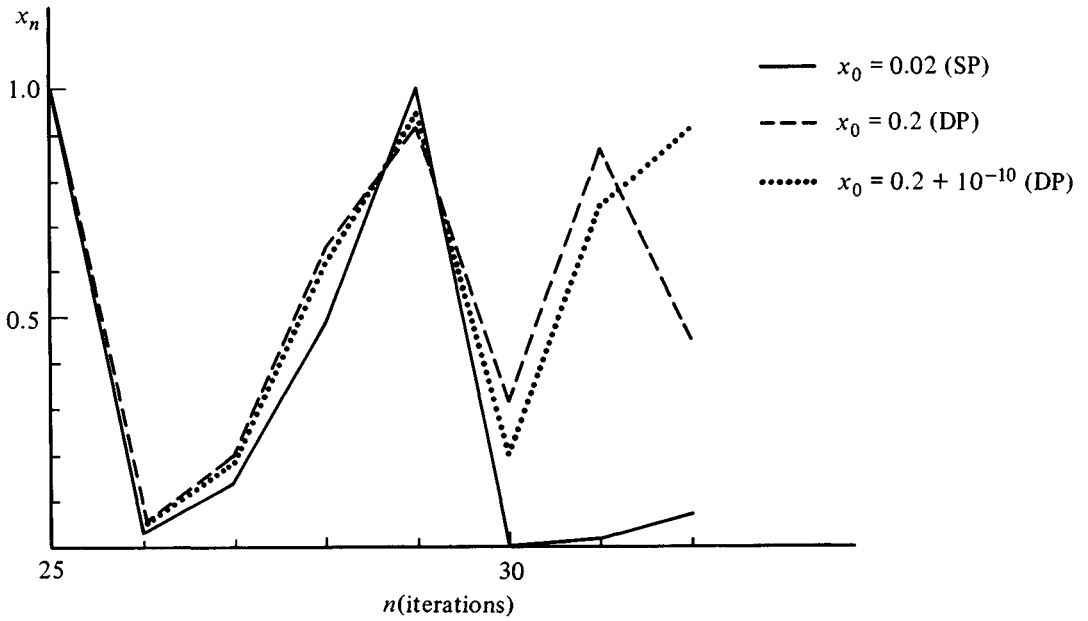
(A) Introduce concepts which deal with questions concerning the possibility of computing numbers (and in particular, dynamical orbits) with finite algorithms (programs). This gets into such topics as Turing's universal computer, computable functions, computable irrational numbers, Kolmogorov's concept of complexity, and questions of polynomial vs. nonpolynomial computing time. These important topics are beyond the scope of our present study, and bear at least questionable relevance to the understanding of the dynamics in THE FINITE (e.g., see Bennett, 1985, 1986).

(B) Try to establish relationships between computed properties and associated

properties of the solutions in THE INFINITE. We will examine briefly several examples of such enquiries.

Consider first the computation of an orbit using a computer with its necessarily limited accuracy. If there is an exponential separation of nearby solutions, small differences in the calculation of an orbit rapidly produces totally different results. This, of course, means that the computed result no longer has anything to do with THE INFINITE, but is a consequence of the possible finite numbers of some particular computer. Fig. 4.78 illustrates this fact for the logistic map, with $c = 4$. One computer

Fig. 4.78



was used, with either single precision (SP: 7 digits) or double precision (DP: 17 digits) accuracy, for the 'same' initial condition $x_0 = 0.2$. In addition a DP iteration was done for an initial condition which differs by 1×10^{-10}.

The SP value of x_n deviates noticably for $n > 26$, and the different initial states are clear for $n > 28$. When $n \geqslant 30$ there is no correlation between the SP and DP solution with the same initial condition. Also, a difference of 10^{-10} in the initial condition produces 'uncorrelated' solutions for $n \geqslant 32$. Thus, roughly speaking, THE FINITE dominates the computed (hence known) solutions after about thirty iterations for the map

$$x_{n+1} = 4x_n(1 - x_n) \quad (0 \leqslant x_0 \leqslant 1). \tag{4.11.1}$$

The nature of the chaos in the iterates of (4.11.1) can be better understood once we

use the Ulam–von Neumann transformation.

$$x_n = \sin^2(\pi y_n), \tag{4.11.2}$$

where y is now only defined modulo 1 (i.e., y is on a unit circle, S^1). Using $4\sin^2(\pi y_n)\cos^2(\pi y_n) = \sin^2(2\pi y_n)$, it is clear that (4.11.1) transforms to the map

$$y_{n+1} = 2y_n \bmod (1), \tag{4.11.3}$$

and y_n is uniquely defined by requiring $1 > y_0 \geqslant 0$. The mod (1) simply means that we are to discard the integer part of y_{n+1} at each step. Using this mod(1) notation, we can readily write down the formal solution to (4.11.3), namely

$$y_n = 2^n y_0 \bmod (1).$$

However, this does not help to clarify much, since the numbers y_n must still be computed.

A much more useful representation of the dynamics of (4.11.3) is obtained by using *binary notation*. This is, of course, also the natural language of computers, with their 'on' and 'off' switches. Since $0 \leqslant y < 1$, we are interested in the binary fractions $2^{-n}(n = 1, 2, 3, \ldots)$, and the binary representation of a number y is given by first making the identification

$$\sum_{n=1}^{\infty} a_n 2^{-n} \quad (a_n = 0 \text{ or } 1), \tag{4.11.4}$$

and then recording these coefficients, a_n. We do this by writing

$$y = .a_1 a_2 a_3 \ldots, \tag{4.11.5}$$

with the understanding that the value of y is given by (4.11.4). A sequence of such symbols, $(a_1, a_2, a_3 \ldots)$, is sometimes referred to as a 'string'.

The binary representation (4.11.5) has the great advantage that if

$$y = \cdot a_1 a_2 a_3 \ldots$$

and

$$y' = 2y \bmod (1) \equiv \cdot a_1' a_2' a_3' \ldots \tag{4.11.6}$$

then $a_k' = a_{k+1}$. In other words, the operation $2y$ mod (1) simplify shifts the numbers a_k one place to the left, and discards a_1. For example, if $y = 0.100\,101\,0\ldots$, then $2y$ mod $(1) = 0.001\,010\ldots$. Thus the dynamics (4.11.3), in binary notation (4.11.5), simply amounts to shifting the a_k one place to the left, discarding the previous a_1 because the mod (1) operation. [If $y_{n+1} = 2y_n$, without the mod (1) operation, so that y is any real number, then the shift of all a_k in the binary representation, $y = \sum_{n=-\infty}^{\infty} a_n 2^n$, is an example of the so-called *Bernoulli shift*, $a_n' = a_{n+1}$. Also see Smale's horseshoe map, Appendix K.]

Because THE INFINITE dynamics of the particular logistic map (4.11.1) (i.e., $c = 4$) amounts to the shift operation (4.11.6), and this discards the previous a_1 at each

iteration, we clearly lose information about the previous state after each iteration. Thus to get a chaotic solution out of the dynamics (4.11.3) requires an initial y_0 which is an infinite, random, binary digit string, (4.11.5). Most $y_0 \in R$ (in a Lebesgue measure sense) are in fact truely random and cannot be computed by any deterministic finite program. Thus, within the realm of THE INFINITE, we see that the chaos of (4.11.3), and hence (4.11.1), resides in this randomness in the real numbers, which are 'used' (in principle) for the initial conditions. Thus, as Ford (1986) has emphasized, the randomness of the dynamics of (4.11.3) and (4.11.1), can be viewed as a 'transcription' of this randomness which is intrinsic to most numbers. But as he and others (e.g., Prigogine and Stengers, 1984) have also emphasized, we do not deal with the real number in computations on experiments. There we must use finite amounts of information.

What happens in the realm of THE FINITE? Because (4.11.3) discard the previous a_1 information from the binary string (4.11.5) at each interation, it is clear that all information in the initial condition

$$y_0 = \cdot a_1 a_2 \cdots a_N \tag{4.11.7}$$

is discarded after N iterations. Thus, putting (4.11.7) on a computer (assuming it can hold the N bits of information) will simply yield $y_k = 0$ for $k \geqslant N + 1$, which is hardly chaotic. Thus there is not anything that looks even vaguely chaotic for (4.11.3) in THE FINITE.

On the other hand, the calculations at the beginning of this section illustrate that calculations (in THE FINITE) of the dynamics of (4.11.1) continue indefinitely in what may appear to be a reasonably 'chaotic', if not altogether 'accurate' fashion (from the point of view of THE INFINITE). It should be noted that this great differences between the calculated behavior of (4.11.1) and (4.11.3), is due to the fact that THE FINITE is very different in the two systems. This is because they are only related within THE INFINITE, namely through (4.11.2), which has no meaning in THE FINITE. Thus there is no reason for their calculated behavior to be related after N steps, given (4.11.7).

This leaves open the question as to the possible relationship between the computed orbits of (4.11.1) and the mathematical orbits, in THE INFINITE – including, perhaps, the question 'Who Cares?'. After all, the only orbits we know about are the computed orbits. So a scientist might take the point of view that she will let the mathematicians worry about THE INFINITE. However, sometimes certain properties of the dynamics in THE INFINITE can be established even though detailed orbits are unknown. Thus one may be able to determine the Lyanpunov exponent, or the invariant measure, or the dimensions of an attractor, or whether the system is ergodic, and so forth. In that case, it would be nice to know if these properties also appear in THE FINITE – and, more specifically, which FINITE? That is, how much information do we need to manipulate to 'approximate' THE INFINITE properties? This is a very large subject, which presently has very few answers, but let us look at a few ideas.

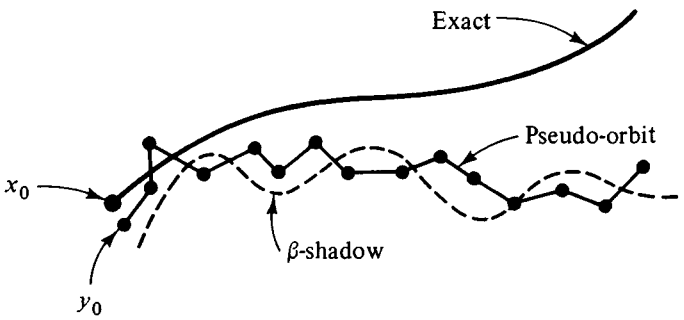
Following Benettin *et al.* (1978) we call a computed orbit (in THE FINITE) a *pseudo-orbit*, and ask 'To what extent does a pseudo-orbit have the unpredictable properties of the exact mathematical orbit (in THE INFINITE), in a deterministically chaotic system?' One response to similar questions has been to introduce the concept of a *shadowing orbit* (The term 'shadowing' was introduced by Bowen, 1978).

To introduce this concept, let

$$x_{n+1} = F(x_n) \quad (0 < x < 1) \tag{4.11.8}$$

yield an exact 'orbit', associated with an initial state x_0. We illustrate this as a continuous curve (for clarity) in Fig. 4.79. A *δ-pseudo-orbit* $(y_0, y_1, y_2, \ldots)$ is one such

Fig. 4.79



that $|F(y_n) - y_{n+1}| \leqslant \delta$ for all n. It allows for the introduction of a numerical error, or noise, at each step of the iteration. A simple example is a roundoff to some accuracy δ (e.g., $\delta = 10^{-6}$)

$$y_{n+1} = \delta \,\mathrm{Int}\,[F(y_n)/\delta] \tag{4.11.9}$$

where $\mathrm{Int}(z)$ equals the integer part of z. In this case $|F(y_n) - y_{n+1}| = |F(y_n) - \delta \,\mathrm{Int}\,[F(y_n/\delta)]| \leqslant \delta$, as can be seen by dividing by δ. Since the dynamics of (4.11.9) occurs on a set of rational numbers (a 'lattice') we illustrate the pseudo-orbit as a series of straight line segments (again for clarity) connecting these lattice points (Fig. 4.79). If the dynamics of (4.11.8) is bounded (e.g., $0 \leqslant x \leqslant 1$), then the number of lattice points associated with the pseudo-orbit is finite. (e.g., δ^{-1} in the case (4.11.9)), in which case all pseudo-orbits are eventually periodic. That is y_0 is either a periodic point or y_n is periodic for some n.

In recent years there have been studies (e.g., Bowen, 1978; Guckenheimer and Holmes, 1985; McCauley and Palmore, 1986; Coven, Kan, and Yorke, 1986) to understand under what conditions there exists exact orbits, $\bar{x}_n = F(\bar{x}_n)$, which can 'shadow' pseudo-orbits. More specifically, $F(x)$ is said to have the *shadowing property* if, for every $\beta > 0$ there is a $\delta(\beta) > 0$ such that every δ-pseudo-orbit, y_n, can be β-shadowed

by some exact solution $\bar{x}_n$, as defined by

$$|\bar{x}_n - y_n| \leqslant \beta \quad (\text{all } n \geqslant 0). \tag{4.11.10}$$

If this holds, then $\bar{x}_n$ is said to β-shadow the pseudo-orbit y_n. Put another way, even though the δ-pseudo-orbit diverges from the orbit (starting at x_0) that it was intended to describe, the pseudo-orbit may be shadowed within a distance β by another exact solution, provided $\delta(\beta)$ is sufficiently small. Notice that this property may not be of much use since, given β, we may not have a computer with the necessary accuracy, $\delta(\beta)$, in which case that computer need not have β-shadowing pseudo-orbits.

If the β-shadowing orbits exist, and are in some sense 'typical' of the entire family of exact orbits (4.11.8), then the computed pseudo-orbits will reflect this same 'typical' behavior, and this feature of the dynamics will not be lost by the finite numerical accuracy. At present it is not known that 'typical properties' of the entire family of exact orbits, if any, are retained by the β-family – (provided they exist!) and hence our original question remains unanswered at present.

Exercise 4.18. (McCauley and Palmore 1986) To explicitly illustrate a β-shadowing orbit, consider the Bernouli shift dynamics $x_{n+1} = 2x_n \bmod (1)$. Assume that some computation method makes a constant error, e (a rational number), in the calculation of this shift, so that the pseudo-orbit obeys $y_{n+1} = 2y_n + e \bmod (1)$. This is a heuristic rather than a realistic example, since no acceptable computation method would make such a systematic error – but let us consider it anyway. Show that, if $x_0 = y_0 + e$, then x_n β-shadows y_n. Determine the minimum β, and also determine a fundamental property of the β-shadowing orbit. Is this property a typical property of the orbits of $x_{n+1} = 2x_n \bmod (1)$?

THE FINITE can also greatly alter the number of basins of attraction of a map, as well as its bifurcation sequences. To be more specific, we consider again the logistic map, but now consider the pseudo-orbits which are in a dynamical space with some finite 'lattice' of values, as in (4.11.9). That is, we divide the interval into a set of rational numbers,

$$x(j) = j/(N+1) \quad (j = 1, 2, \ldots, N), \tag{4.11.11}$$

so that the dynamics takes place on N points (the 'lattice'). It is certainly appropriate at this point to recall the biological origin of the logistic map, in which x represented the (normalized) population of a species in succeeding generations. Clearly this number is rational, even if the maximum population (N in (4.11.11)) is very large. Thus the mathematical logistic map, in THE INFINITE, is quite an extrapolation from its empirical origin in THE FINITE. The dynamics will now be defined so that it is 'similar' to the logistic map, but satisfies the lattice restriction (4.11.11). We will use the

N	$P(N)$
500	2;7;10
1000	10;10
100 000	81
200 000	5;355
400 000	96;508
800 000	159;192
1 600 000	304

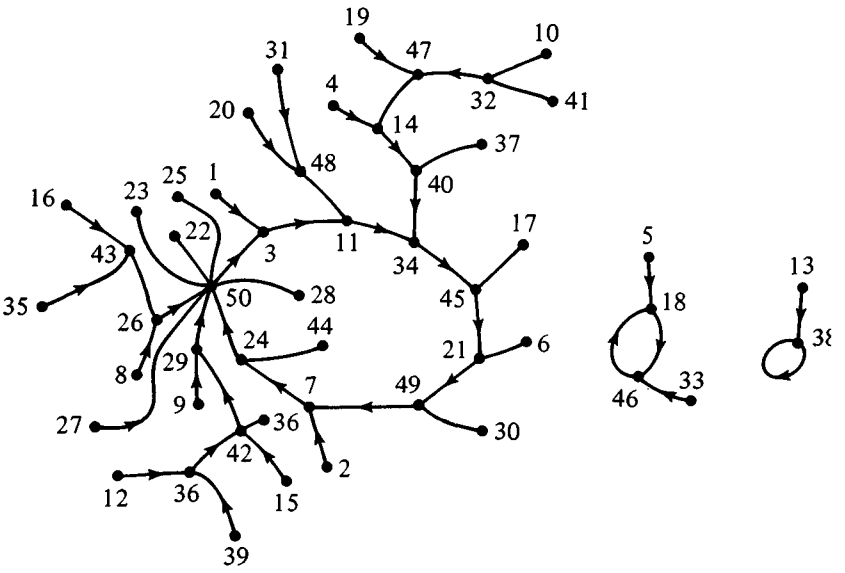
form (4.11.4), now with $\delta = 1/(N+1)$, to define this lattice map $x \to x'$ by

$$x' = \text{Int}\,[(N+1)cx(1-x)]/(N+1) \tag{4.11.12}$$

where $\text{Int}\,(z)$ is the integer part of z.

As noted at the beginning of this section, for any finite lattice, all initial states must tend to periodic orbits. That is, all x_0 are eventually periodic. Fig. 4.80 illustrates this

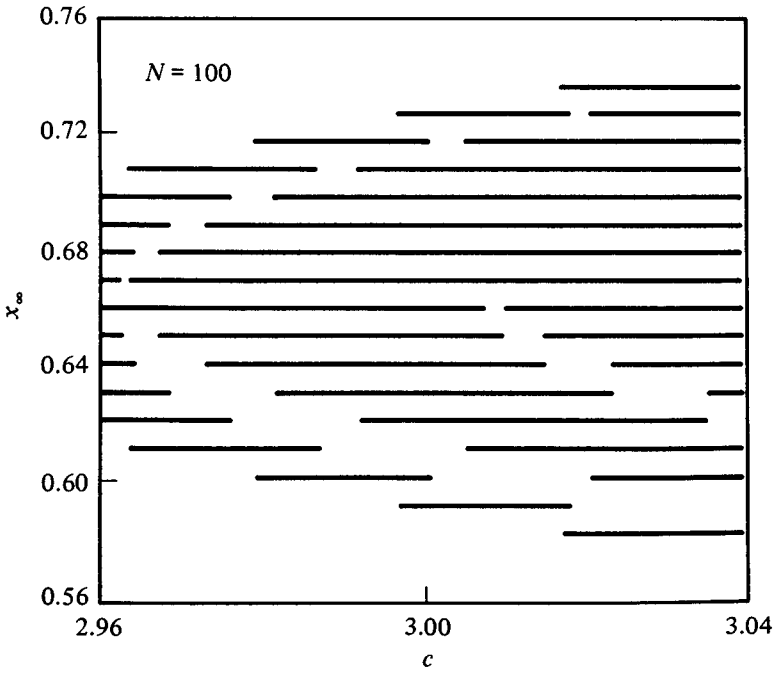
Fig. 4.80



for the case $N = 50, c = 4$. It shows the 'flow' of the iterates of all 50 initial states, and the fact that there are three basins of attraction to periodic orbits with periods $P = 1, 2$, and 9. This is obviously not a very large value of N, and hence not necessarily representative of what happens for larger N.

It is presently unclear in what sense chaotic motion is approximated within THE FINITE, for large values of N. The periods of the asymptotic motion, $P(N)$ apparently do not become proportional to N, nor does the number of such basins increase. This is indicated by the table, obtained from limited samples of initial conditions (for $c = 4$).

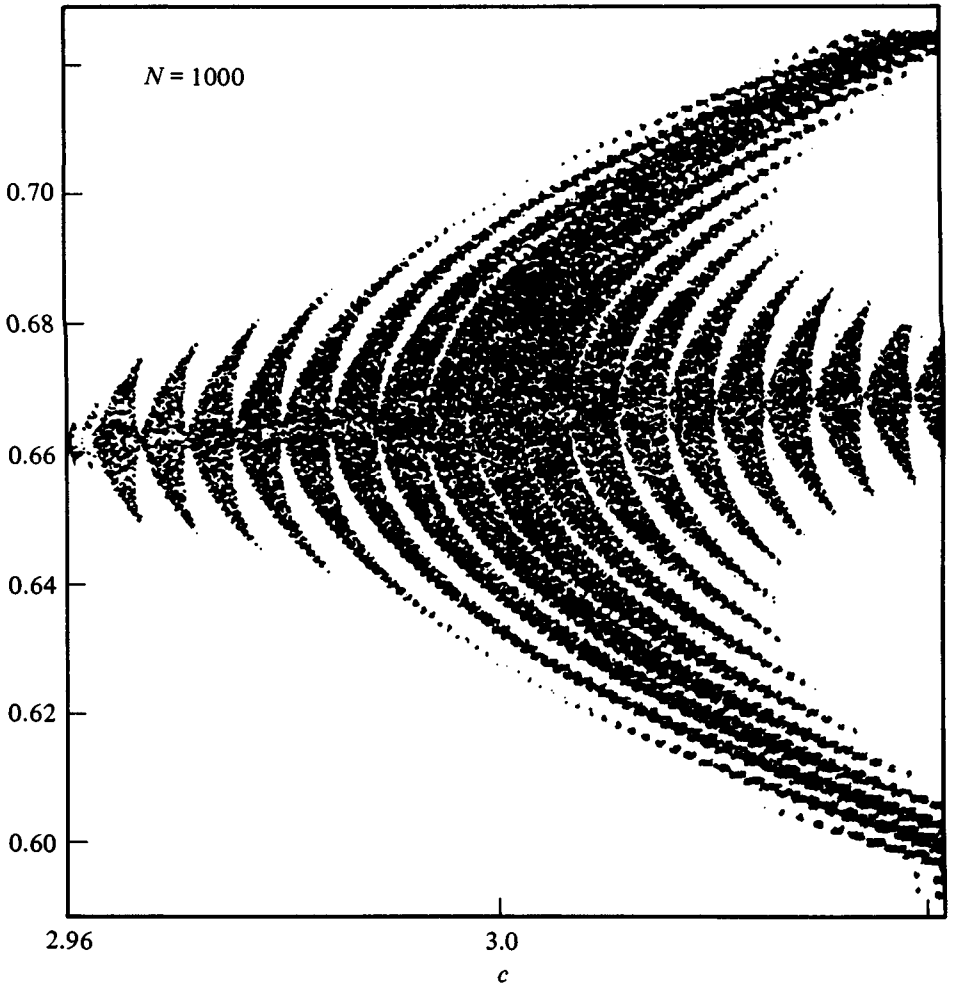
Fig. 4.81



(*a*)

Thus most states are not periodic. However, this alone is not an indication of chaos (e.g., all states could, in one step, go to a fixed point). It is not the asymptotic motion which can characterize 'chaos' but rather some form of 'initial' scrambling of the orbits (e.g., discrete and finite-time forms of Lyapunov exponents). These details are yet to be understood. For two-dimensional examples, see Section 6.14.

In addition to the chaotic dynamics at $c = 4$, it is interesting to see the difference which THE FINITE makes on the dynamics in the region of THE INFINITE bifurcation points. This can be explored at all of THE INFINITE bifurcation points, but we illustrate the modified bifurcation (new definition?), by considering the first period-two bifurcation region, $c \simeq 3$. Fig. 4.81 illustrate all of the periodic points (for all initial conditions) in the attracting periodic cycles, for different numbers of lattice points. Essentially all points come in pairs (period two cycles). Using $x_{n+1} = \text{NINT}(cx_n(1-x_n)N)/N$, with NINT$(z)$ being the nearest integer of z (differing slightly from (4.11.12)), the number of pairs at $c = 3$ is 7 when $N = 100, 30$ when $N = 1000$, 44 when $N = 2000$, and 63 when $N \sim 3{,}300$. Note that the graphic dots have been reduced in size going from the cases $N \leqslant 1000$ to $N \geqslant 2000$. The differences between these structures and the single period-one, ($c \leqslant 3$), or period-two ($c > 3$) attractor of the logistic map in THE INFINITE is very striking. The bifurcation features for large values of c are also dramatically altered, due to the discrete nature of the Int(z) function. It appears to require quite large values of N to recover the



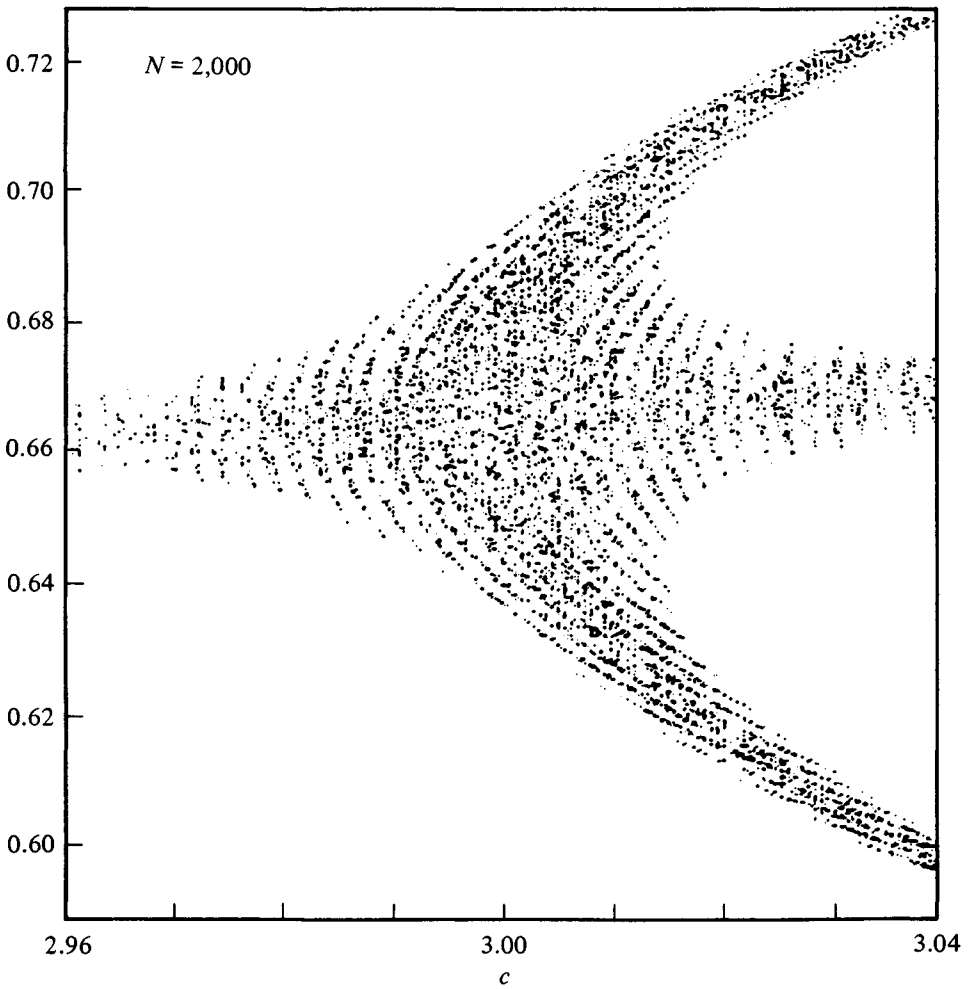
(*b*)

continuum behavior. However, note in the above figures the reduction in both the range and the gaps, as N is increased from $N = 2000$ to $N \simeq 3300$. Of course, the higher order bifurcations, the Cantor set attractor, and periodic windows are largely lost in THE FINITE, unless N is very large.

Comments on exercises

(4.1) Also see Section 10 of this chapter.

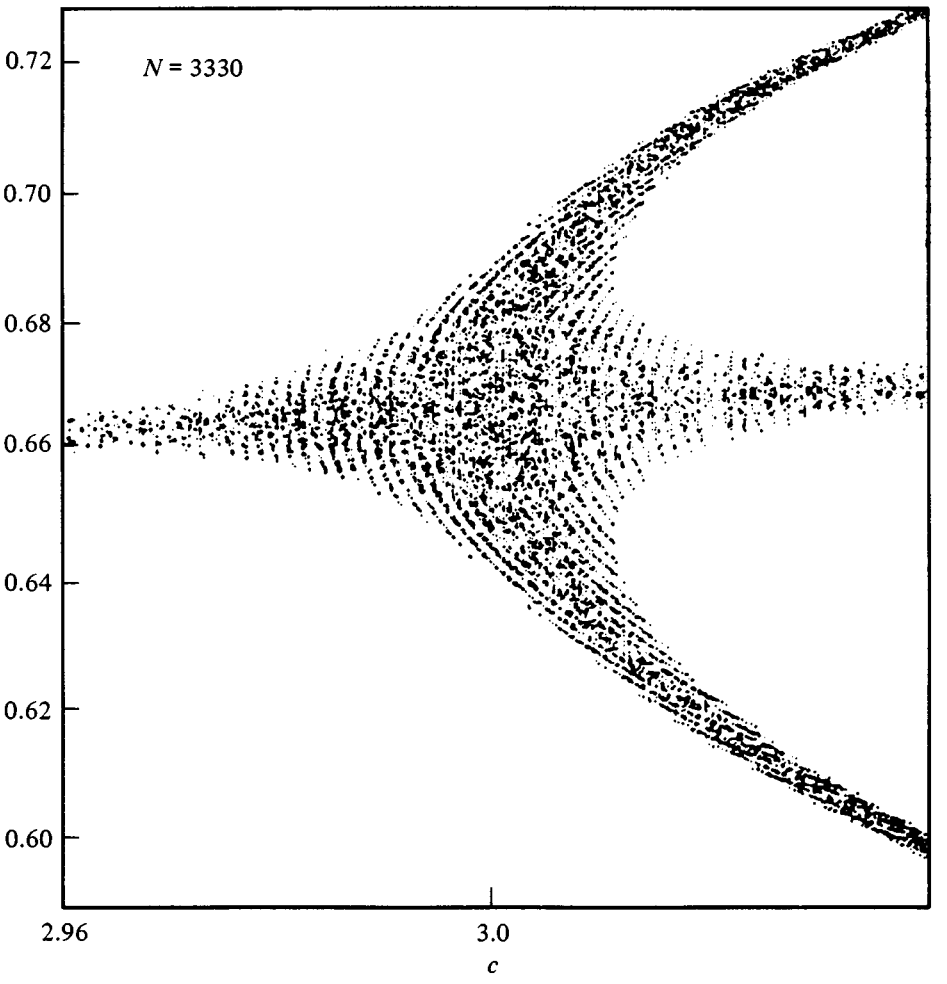
(4.2) If $F^n(x_0) = x_0$, then $F^{m+n}(x_0) = F^m(x_0)$, so $F(x_m) = x_m$ $(m = 0, \ldots, n-1)$. There are n periodic points.



(*c*)

(4.3) Continuity is the essential element here. Any curve which joins the vertical at $x_n = 0$ to the one at $x_n = 1$ (defining the map), and such that $0 \leqslant F(x_n) \leqslant 1$, must (because of continuity) intersect the diagonal, $x_{n+1} = x_n$, at least at one point (Fig. 4.83).

(4.4) For $c = 3$, $x_0 = \frac{2}{3}$ is a fixed point (period one) of $F = cx(1-x)$. If $F^2(x_e) = F(F(x_e)) = \frac{2}{3}$, then $F(x_1) = \frac{2}{3}$ where $x_1 = F(x_e)$. We first obtain x_1, $3x_1(1-x_1) = \frac{2}{3}$, so $x_1 = \frac{1}{3}$, and $F(x_e) = \frac{1}{3}$; $x_e^2 - x_e + \frac{1}{9} = 0$, $x_e = \frac{1}{2} \pm 5^{1/2}/6$. Similarly, if $F^3(x_3) = \frac{2}{3}$, then $F^2(F(x_e)) = \frac{2}{3}$, so we need $F(x_e) = \frac{1}{2} \pm \frac{5^{1/2}}{6}$, or $x_e = \frac{1}{2}$



(*d*)

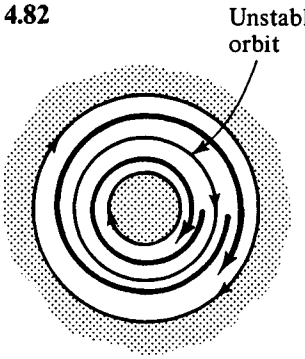
$+\frac{1}{6}(3+2(5)^{1/2})^{1/2}$. Note that $F(x_e)=\frac{1}{2}+\frac{5^{1/2}}{6}$ yields a complex x_e. (Also see Exercise 4.5.)

(4.5)

(4.6) (a) It has a chaotic set, but it cannot be observed because it is unstable. In other words, not all mathematically chaotic sets are of physical (observational) importance.

(b) By continually folding-over the interval [0, 1], it may seem clear that any initial interval, A, will ultimately map to intersect A (i.e., $F^n A \cap A \neq \phi$). This

Fig. 4.82



(*a*)

The flow is on a Möbius strip (see Section 4.9), with a 'central' periodic orbit which is unstable (see (4.2.5) and the following discussion)

(*b*)

Fig. 4.83

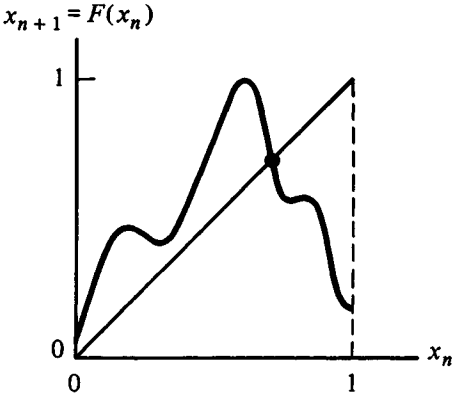
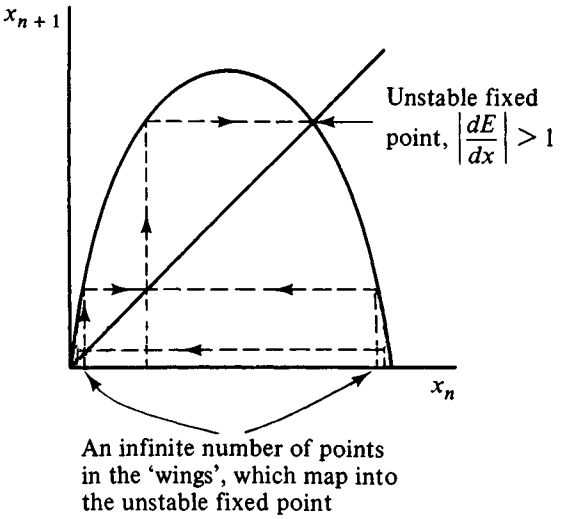


Fig. 4.84



intersect must have a fixed point of F^n, by Brouwer's theorem, and hence contain a period-n point.

(4.7) $x_{n+1} = \sin^2(2^{n+1}\pi\theta) = [2\sin(2^n\pi\theta)\cos(2^n\pi\theta)]^2 = 4x_n[1-x_n]$; if θ is irrational the solution is nonperiodic. Hence the nonperiodic solutions are a set of Lebesgue measure 1 on the interval $0<\theta<1$, but there is also a dense set of periodic solutions (namely all the rationals, $\theta = l2^{-n}/(1+2^k)$, which have period k). Note also that we have here an 'analytic solution (??)' of chaotic motion!

(4.8) We want to show that $G\cdot h = h\cdot F$, or:
For $0<x<1/2$, $4/\pi \arcsin x^{1/2} = 2/\pi \arcsin(4x(1-x))^{1/2}$
For $1>x>\frac{1}{2}$, $2[1-2/\pi \arcsin x^{1/2}] = 2/\pi \arcsin(4x(1-x)^{1/2}$
(a) if $x<\frac{1}{2}$, multiply by $\pi/2$, and take the sine. We need to show $\sin \cdot(2\arcsin x^{1/2}) = (4x(1-x))^{1/2}$, but $\sin(2\arcsin x^{1/2}) = 2\sin(\arcsin x^{1/2}) \cdot[1-\sin^2(\arcsin x^{1/2})]^{1/2}$ (qed)
(b) if $x>\frac{1}{2}$, similarly we need to show $\sin[\pi - 2\arcsin x^{1/2}] = (4x(1-x))^{1/2}$ but $\sin[\pi - 2\arcsin x^{1/2}] = \sin\pi\cos(2\arcsin x^{1/2}) - \cos\pi\sin \cdot(2\arcsin x^{1/4}\sin x^{1/2}) = \sin(2\arcsin x^{1/2}) =$ same as in (a).
All of the fixed points of G^n are unstable ($|\partial G^n/\partial x|>1$), so all fixed points of F^n (period-n points of F) are unstable.

(4.9) $\int_0^1[dx/(\pi(x(1-x))^{1/2})] = \int_0^{\pi/2}[d(\sin^2\theta)/(\pi\sin\theta\cos\theta)] = \int_0^{\pi/2}(2d\theta/\pi) = 1$. $P(y) = \int_0^1 \delta(y-4x(1-x))\,[\pi(x(1-x))^{1/2}]^{-1}\,dx$; the δ function contains an $f(x)$ which gives the same y for two values of x, so split it up, and set $z = 4x(1-x)$; $P(y) = \int_0^{1/2}\delta(y-z)[dx(\pi(z/4)^{1/2})] + \int_{1/2}^1 \delta(y-z)[dx/(\pi(z/4)^{1/2})]$ and $dx = dz/(4-8x)$ and $x = \frac{1}{2} - \frac{1}{2}(1-z)^{1/2}$ in the first integral and $x = \frac{1}{2} + \frac{1}{2}(1-z)^{1/2}$ in the second. So $P(y) = \int_0^1 \delta(y-z)[1/(\pi(z/4)^{1/2})][dz/(4(1-z)^{1/2})] + \int_1^0 \delta(y-z)[1/(\pi(z/4)^{1/2})][dz/(-4(1-z)^{1/2})] = \int_0^1 \delta(y-z)[dz/(\pi(z)^{1/2}(1-z)^{1/2})] = [\pi(y(1-y))^{1/2}]^{-1}$. The Lyapunov exponent can be obtained from (4.6.2) $\lambda = \int_0^1[\ln|4-8x|/(\pi(x(1-x))^{1/2})]\,dx = \int_0^1[\ln 2^2 + \ln|1-2x|/(\pi(x(1-x))^{1/2})]\,dx = 2\ln 2 + 2\int_0^{1/2}[\ln(1-2x)/(\pi(x(1-x))^{1/2})]\,dx$ since $d(-\sin^{-1}(1-2x)) = dx/(x(1-x))^{1/2}$; set $(1-2x) = \sin\theta$, then $\lambda = 2\ln 2 + 2/\pi\int_{\pi/2}^0 \ln(\sin\theta)(-d\theta)$; now $\int_0^{\pi/2}\ln(\sin\theta)\,d\theta = -\pi/2\ln 2$ so $\lambda = 2\ln 2 - \ln 2 = \ln 2$.

(4.10) For a periodic orbit (period m), x only 'exists' on the points x_n, such that $f^m(x_n) = x_n (n = 1,\ldots,m)$, so the probability (note $\int_0^1 P(x)\,dx = 1$) is $P(x) = (1/m)\sum_{n=1}^m \delta(x-x_n)$. Hence $\lambda = \int \ln|df/dx|P(x)\,dx = (1/m)\sum_{n=1}^m \ln|df/dx|_{x_m} = (1/m)\ln|df^m/dx|_{x_j}$ where x_j is any period-m point. The last equality is due to (4.2.11). Thus the sign of λ is directly related to whether $|df^m/dx|_{x_j} > 1$ (unstable) or <1 (stable).

(4.11) $P(x) = 1$ is invariant, according to (4.6.3), if $1 = \int_0^1 \delta(y-f(x))\,dx$. Substituting $f(x) = z$ in each interval yields $x_0\int_0^1\delta(y-z)\,dz + \int_1^0\delta(y-z)(1-x_0)(-dz)$, $x_0 + (1-x_0) = 1$, so it is indeed invariant. (4.6.2) yields $\bar{\lambda} = -x_0\ln(x_0) -$

$(1-x_0)\ln(1-x_0)$. The most unstable map is given by $d\bar{\lambda}/dx_0 = 0$, which yields $x_0 = \frac{1}{2}$, so $\bar{\lambda}_{\max} = \ln 2$.

(4.12) The number of occupied cells is $N(\varepsilon) = 69$, $\varepsilon = 10^{-3}$, so $d_c \simeq \ln(69)/3\ln 10 = .613$.

(4.13) If you work this exercise out in detail, the general answer will become clear. First of all, 2^{N+1} iterations of the point $x_0 = \frac{1}{2}$ produces 2^N intervals. For a Cantor set which is self-similar to 2^N intervals in $(1, 2)$, the $(2^N)^2$ intervals obtained in the next level of similarity construction must satisfy the following equalities

$$L(K + I2^N, 2^{2N} + K + I2^N)/L(K, K + 2^N) = L(1 + I, 1 + I + 2^N)/L(1, 2)$$

where $K = 1, 2, \ldots, 2^N$, and $I = 0, \ldots (2^N - 1)$. Here

$$L(i, j) = |x_i - x_j|, \quad x_n = F^n(1/2, c_\infty).$$

This result is not particularly obvious, but it is correct. The exercise corresponds to the case $N = 2$. From this, one finds that the Cantor set generated by the iterations of $x_0 = \frac{1}{2}$, is not more self-similar to 2^N intervals than it is to 2^{N-1} intervals (i.e., the percentage deviation from the above equalities does not decrease as N increases). Nonetheless, the estimate of d_c is improved, simply because it is based on the more accurate starting point of 2^N intervals.

(4.14) Eliminate the θ dependence between the equations of (4.8.6) to obtain $a^4 - a^2x^2 - \dot{x}^2 = 0$. We then obtain $K(x, \dot{x}) = 2a^2$. The constant a is arbitrary, so $\ddot{x} = -a^2x$ is not a harmonic oscillator. The proper equation is $\ddot{x} = -(x/2)(x^2 + (x^4 + 4\dot{x}^2)^{1/2})$. Is this a Lipschitz system?

(4.15) $\rho(x, \dot{x}) = [2\pi(a_2 - a_1)(x^4 + 4\dot{x}^2)^{1/2}]^{-1}$.

(4.16) K is proportional to the height of the 'spikes' in the figure, provided that the relaxation of the ϕ-oscillator is fast relative to $P(\phi)$. Otherwise the response to the disturbance would not simply depend on the single point ϕ_n (i.e., $K\sin(\phi_n)$). Also the relaxation must be rapid compared to $P(\psi)$, for otherwise there would be an overlap effect, not accounted for in (4.9.4). If the ϕ-oscillator has a relaxation which is some fraction of $P(\phi)$, then the last condition means that Ω cannot be too large. Presumably a 'safe' value would always be $\Omega \lesssim 1$.

(4.17) The set $\rho = 0$ occurs at $\Phi = 0$ when there is no coupling, $K = 0$, $\Phi = 0$ implies that $P(\psi) = 0$ so that the driving oscillator is very fast. If now $P(\psi)$ is very small, the rotation number, (4.9.10) is the average change of ϕ over this short period. In the case (4.9.13) the coupling is strong enough to cause the 'naturally slow' ϕ-oscillator to retain a zero (time-averaged) phase shift relative to the fast ψ-oscillator, i.e., when viewed at each period of the ψ-oscillator.

For the proof, let $\Phi' = \Phi + 2\pi l$. $T_{\Phi'}\phi = T_\Phi\phi + 2\pi l$ so

$$\rho(\Phi') = \lim_{n\to\infty} \frac{T^n_{\Phi'} - \phi}{2\pi n} = l + \rho(\Phi).$$

(4.18) We find that $x_1 = 2y_0 + 2e \bmod (1)$ whereas $y_1 = 2y_0 + e \bmod (1)$. More generally $x_n = y_n + e$, so $|x_n - y_n| = e \leqslant \beta$, if $\beta = e$. All y-solutions are eventually periodic, whereas only a set of measure zero of the x-solutions are periodic. Hence, in this example, the β-shadowing trajectories are not typical of most x-solutions.