1

INTRODUCING ITERATED FUNCTIONS

A *dynamical system* is any mathematical system that changes in time according to a well specified rule. We will look at two different types of dynamical systems in this book: iterated functions and differential equations. We will use these two types of dynamical systems to address a central question: what sorts of behaviors are possible for different types of dynamical systems?

In this chapter I'll introduce iterated functions. I'll say what iterated functions are, present several ways of visualizing their behavior, and introduce some key terminology. This chapter may be a bit abstract. We'll approach iterated functions as simple mathematical systems, without attention to their roles as models of the physical or biological world. In Chapter 2, where I introduce differential equations, we will begin to see how dynamical systems are used in science. This chapter may also be a bit dry; there's nothing too deep in the next few pages. However, it's essential background for the more interesting and surprising results that will come later.

1.1 Iterated Functions

Consider a function f of one variable such as $f(x) = x^2$. A function establishes a relationship between a set of numbers, the inputs x, and another set of numbers, the outputs $f(x)$. We can think of the function f as an action. We start with a number x, apply the function f to it, and get a new number. This new number is called $f(x)$: it is x after it has had f applied to it.

Usually we think of the application of a function as a one-shot deal: Do f to x, get $f(x)$, end of story. For example, if $f(x)=x^2$, then $f(3)=3^2=9$, and $f(-0.5)=(-0.5)^2=0.25$. But if we apply the function repeatedly, using the output at one step as the input to the next, then we have a dynamical system: a mathematical entity that changes in time according to a well-defined rule. For example, we could start with 3, apply f, and obtain 9. Apply f again, and we get $9^2=81$. Apply f yet again, and we get $81^2=6561$. The result is a sequence of numbers:

$$3, 9, 81, 6561, 43046721, \ldots . \tag{1.1}$$

We see that the numbers quite quickly become very large.

This process is known as *iteration*. The application of f is repeated, or iterated, and the output of one step is used as the input for the next step. Our starting number—in this case 3—is known as the *seed* or the *initial condition*. The sequence of numbers in Eq. (1.1) is known as the *orbit* or *itinerary* of 3. The initial condition is usually denoted x_0. The first value in the itinerary is denoted x_1 and is called the first *iterate*. This value is obtained by applying f to x_0. That is, $x_1=f(x_0)$. The second iterate is denoted x_2 and is obtained by applying f to x_1: $x_2=f(x_1)$. Equivalently, we may think of x_2 as resulting from twice applying f to x_0: $x_2=f(f(x_0))$. Subsequent iterates are denoted similarly.

Let's consider another example: $g(x)=\frac{1}{2}x+4$. I'll choose an initial condition of $x_0=1$. The first iterate x_1 results from g acting on x_0:

$$x_1 = g(x_0) = g(1) = \frac{1}{2}(1)+4 = 4.5\,. \tag{1.2}$$

Subsequent iterates are found in a similar manner; $x_2=g(x_1)$, and so on. The first several iterates of this initial condition are shown in

t	x_t
0	1
1	4.5
2	6.25
3	7.125
4	7.5625

Table 1.1 The first several iterates of the initial condition $x_0 = 1$ for the function $g(x) = \frac{1}{2}x + 4$.

Table 1.1. (You might want to grab a calculator and take a moment to verify these numbers.)

It is often useful to display an orbit graphically rather than in a table or list. The orbit in Table 1.1 is plotted in Fig. 1.1. This type of graph is known as a *time series plot*. Such a graph gives a clear view of the orbit's behavior. In Fig. 1.1 we can see that the orbit is approaching 8. Note that a time series plot is not a graph of the function that is being iterated, $g(x) = \frac{1}{2}x + 4$. Instead, it is a plot of the orbit or itinerary.

The value of 8 is a *fixed point* of the function $g(x)$. This means that 8 does not change if g operates on it: $g(8) = 8$, as we can verify:

$$g(8) = \frac{1}{2}8 + 4 = 4 + 4 = 8\,. \tag{1.3}$$

In general, a number x is a fixed point of $f(x)$ if it is a solution to the equation

$$f(x) = x\,. \tag{1.4}$$

Such an equation is called a *fixed-point equation*. In words, Eq. (1.4) says that x, when acted on by f, yields x. A function can have any number of fixed points, including none at all.

Iterated functions are our first example of a dynamical system, a mathematical system that changes in time according to a

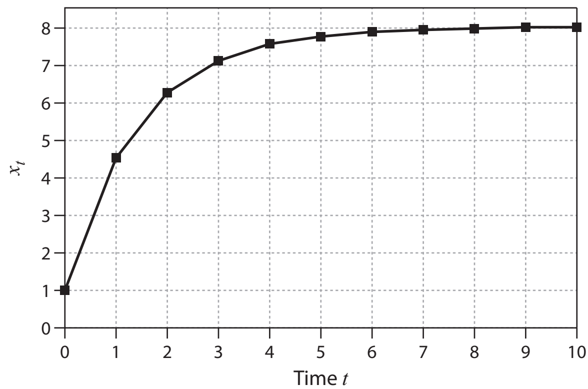

Figure 1.1. The time series plot for the initial condition $x_0 = 1$ iterated with $g(x) = \frac{1}{2}x + 4$.

well-defined rule. In the example we just considered, the rule is given by the function $g(x) = \frac{1}{2}x + 4$. The dynamical system that results from iterating this equation is sometimes written as:

$$x_{t+1} = \frac{1}{2}x_t + 4\,. \tag{1.5}$$

This notation makes the dynamical nature of the equation clearer. We can see that the next value of x is determined by the current value of x. That is, x_{t+1} is a function of x_t. Thus, as long as we know the initial value x_0, we can figure out all subsequent values of x by repeated application of Eq. (1.5).

Note that Eq. (1.5) does not directly tell us x_t as a function of t. If, say, we want to know x_{13}, we can't just plug in $t = 13$ somewhere. Rather, we need to start at some known value of x, usually x_0, and iterate forward, one step at a time, using Eq. (1.5). Doing so might be a bit time consuming, but it is at base a very simple procedure. There is a rule—namely the function $f(x)$—and that rule is applied again and again. For all but the simplest such functions one almost always turns to a computer to carry out the iterations. I used a short python program to iterate the function $g(x) = \frac{1}{2}x + 4$ and make the time series plot shown in Fig. 1.1.

Before going on I should mention some additional terminology. A function f takes input values and returns output values. So one sometimes refers to f as a *mapping* from input to output. Iterated functions are also often called *maps*. In mathematics, a map is synonymous with function. I will usually refer to functions as functions, but the term map is very commonly used in dynamical systems so you will likely see it elsewhere.

1.2 Thinking Globally

In the study of dynamical systems we are often interested not in the numerical values of a particular orbit, but in its long-term behavior. We want to know about the big picture—the global dynamics of the function—not the local details of each and every point in a particular orbit. For example, when describing the itinerary of $x_0 = 1$ when iterated with $g(x) = \frac{1}{2}x + 4$, we simply say that it approaches 8, rather than list all the data in Table 1.1.

Let's consider another example: the square root function $f(x) = \sqrt{x}$. Our goal will be to figure out the long-term behavior of *all* initial conditions. (Since the square root of a negative number results in a complex number—also known as an imaginary number—we will limit our analysis to non-negative numbers.) To get us started, let's choose the seed $x_0 = 4$. Then x_1 is the result of applying the function to x_0. Since $f(4) = \sqrt{4} = 2$, the first iterate is 2. The next iterate is approximately 1.414, since $\sqrt{2} \approx 1.414$. We keep on square rooting and obtain the itinerary shown in Table 1.2.

The orbit for $x_0 = 4$ is shown in Fig. 1.2. Also on this figure are the time series plots for three other initial conditions: 2, 0.5, and 0.25. You can obtain orbits for these initial conditions by entering the seed and then repeatedly hitting the $\sqrt{}$ key on your calculator. However, without using a calculator we can understand the shape of the time series plots qualitatively.

t	x_t
0	4
1	2
2	1.414
3	1.189
4	1.091
5	1.044

Table 1.2 The first several iterates of the initial condition $x_0 = 4$ for the function $f(x) = \sqrt{x}$.

When you take the square root of a number larger than one, the result is a smaller number. For example, $\sqrt{10000} = 100$, $\sqrt{16} = 4$, and $\sqrt{1.5} \approx 1.225$. Numbers larger than 1 get closer and closer to 1 when successively square rooted. We can see this in Fig. 1.2. The seeds 4.0 and 2.0 are both getting smaller and approaching 1.

On the other hand, numbers between 0 and 1 get *larger* when square rooted. For example $\sqrt{0.25} = 0.5$. It might be easier to see this using fractions:

$$\sqrt{\frac{1}{4}} = \frac{1}{2}, \text{ because } \left(\frac{1}{2}\right)^2 = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}. \tag{1.6}$$

So for this dynamical system—iterated square rooting—any number between 0 and 1 will increase and approach 1, and any number larger than 1 will decrease and approach 1. The numbers 0 and 1 are fixed points; they are unchanged when square rooted: $\sqrt{0} = 0$, and $\sqrt{1} = 1$.

With these observations, we can describe the global dynamics of the square root function. That is, we can specify the long-term behavior of all non-negative initial conditions. Any initial condition larger than 1 will get smaller and move closer and closer to 1. Any initial condition between 0 and 1 will get larger and get closer

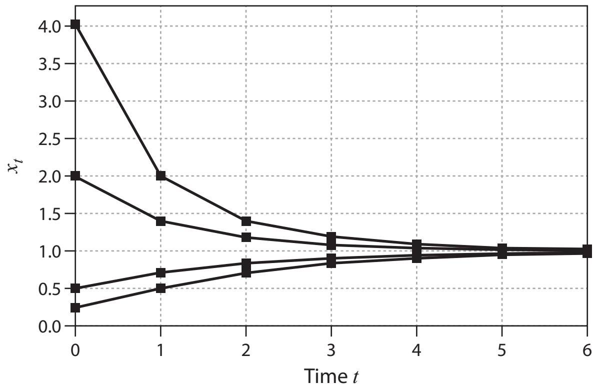

Figure 1.2. The time series plot for four different initial conditions iterated with $f(x) = \sqrt{x}$.

and closer to 1. The initial conditions 0 and 1 are fixed points. They do not change when acted upon by the function: $\sqrt{0} = 0$ and $\sqrt{1} = 1$.

1.3 Stability: Attractors and Repellors

The fixed points of $f(x) = \sqrt{x}$ are 0 and 1, but these two fixed points have a rather different character. The fixed point at 1 is called *stable* or *attracting*. Nearby orbits are pulled toward it; it attracts nearby points. It is called stable because if one is at the fixed point and then a perturbation moves you a little bit away, you will return to the fixed point. That is, if you are at 1 and something happens and you get moved to 1.1, the square-rooting function will move you back, closer and closer to 1. The first several iterates of 1.1 are: 1.1, 1.049, 1.024, 1.012. The orbit is getting closer to 1.

The fixed point at 0 is different. You will not be surprised to learn that this fixed point is *unstable* or *repelling*. If you are at 0 and something happens and you get bumped to 0.05, you will not return to 0. Instead, you will get pushed away from 0, never

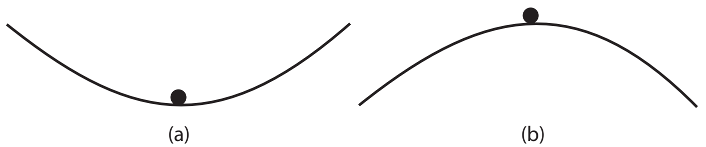

Figure 1.3. A schematic illustration of (a) a stable and (b) an unstable fixed point.

to return. The first several iterates of 0.05 are: 0.05, 0.224, 0.473, 0.688. The orbit is not getting closer to 0.

Stable and unstable fixed points are illustrated schematically in Fig. 1.3. On the left is shown a stable fixed point—a ball at the bottom of a valley. If the ball is moved a small amount it will return to the bottom of the valley. On the right, the fixed point is unstable. If the ball is moved a little bit it will roll down one side of the hill, not to return.

For completeness, I should mention that it is possible for a fixed point to be poised between stability and unstability. In this case, if one moves away from the fixed point one neither returns to the fixed point nor is pushed away. Fixed points with these properties are called *neutral*. In terms of the schematic representation of fixed points shown in Fig. 1.3, neutral fixed points look like a ball resting on a perfectly flat table. If the ball is moved to the left or right, it will stay there; it won't return to its original location, but it also won't roll further away.

The stability of fixed points—or of other dynamical behavior that we will encounter later—is an important notion. Typically, in a mathematical model or the real world, one only expects to observe stable fixed points. An unstable fixed point is susceptible to a small perturbation; a tiny external influence will move the system away from the unstable fixed point. For example, it is possible to carefully balance a pencil on its eraser. However, it will not stay this way for long. A small vibration or bit of wind will make the pencil fall over and lie on its side. Or, returning to Fig. 1.3,

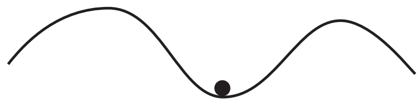

Figure 1.4. A stable fixed point. A large perturbation and the ball will not return to the fixed point, but for small perturbations it will return.

we would not expect to observe the situation depicted in part (b). A rock balanced on the top of a hill will not remain there indefinitely. A small gust of wind or a little push will cause it to roll downhill. The upshot is that in dynamical systems one is particularly focused on stable behavior, as usually it is only stable behavior that is observed.

Before moving on, I should define stable and unstable fixed points just a bit more carefully. A fixed point x is stable if there is an open interval around x such that any initial conditions in this interval get closer and closer to the fixed point x. In terms of the schematic view of Fig. 1.3, this says that a fixed point is a point in the bottom of a valley, regardless of how wide or narrow the valley is. This is illustrated in Fig. 1.4.

1.4 Another Example

Let's consider another example: the cubing function $f(x) = x^3$. What are its dynamics? How many fixed points are there and what are their stability? Let's start by solving for the fixed points. A point x is fixed if it is unchanged by the function. That is, $f(x) = x$. Here, the function is $f(x) = x^3$, so the equation for fixed points is:

$$x^3 = x \,. \tag{1.7}$$

This equation has three solutions: $-1, 0,$ and 1. Each of these numbers, when cubed, does not change. For example, $(-1)^3 = (-1)(-1)(-1) = -1$.

Are these fixed points stable or unstable? Let's think about what happens to different initial conditions. A number larger than 1 will get larger when cubed. For example, the itinerary of the initial condition $x_0 = 2$ is:

$$2, 8, 512, 134217728, \ldots . \tag{1.8}$$

The orbit grows very rapidly and will keep getting larger. One describes this situation by saying that the orbit grows without bound or tends toward infinity. A number less than -1 will get "bigger and more negative." (Strictly speaking this means that the orbit gets smaller; as one moves to the left on a number line the numbers get smaller. Negative three is less than negative two.) So we say that the orbit of -1 decreases without bound or tends toward negative infinity. Lastly, numbers between -1 and 1 will approach zero when cubed. For example, the orbit of $x_0 = -0.9$ is:

$$-0.9, -0.729, -0.38742, -0.058450, \ldots . \tag{1.9}$$

Thus, 0 is a stable, or attracting fixed point; it pulls in all initial conditions between -1 and 1. The fixed points at ± 1 are unstable, or repelling.

1.5 One More Example

I'll end this chapter with one more example. We'll consider the function $f(x) = x^2 - 1$. Does this function have any fixed points? Yes—two of them, in fact. The fixed point equation

$$f(x) = x \tag{1.10}$$

has two solutions:

$$x = \frac{1}{2}(1 + \sqrt{5}) \approx 1.61803\,, \tag{1.11}$$

and

$$x = \frac{1}{2}(1 - \sqrt{5}) \approx -0.61803\,. \tag{1.12}$$

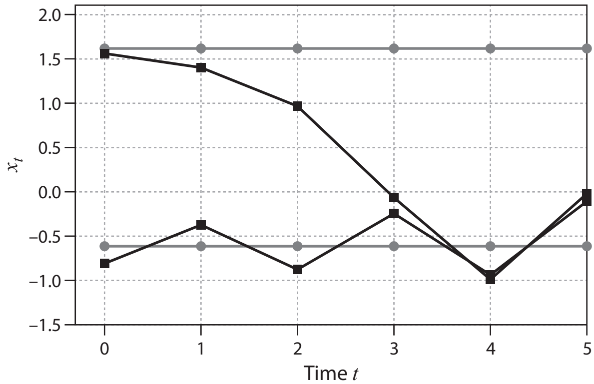

Figure 1.5. The time series plot for four different initial conditions iterated with $f(x) = x^2 - 1$. The two orbits plotted with gray circles are the fixed points, 1.618 and −0.618.

Are these fixed points stable? Let's iterate and see. In Fig. 1.5 I have plotted the orbits for four different initial conditions. The orbits shown with gray circles are the two fixed points, $x \approx 1.61803$ and $x \approx -0.61803$. The orbits plotted with squares begin close to the fixed points; the top orbit has an initial condition of $x_0 = 1.55$, and the initial condition for the bottom orbit is $x_0 = -0.8$. We see in the figure that the two square orbits are not pulled in toward the fixed points, so the fixed points are not stable.

It looks like the two orbits plotted with squares are getting closer together. By $t = 4$ or 5 they are almost on top of each other. What could be going on? To address this question, in Fig 1.6 I have plotted the two square orbits out to $t = 15$. The orbits of the two fixed points are again shown as gray circles. As in the previous figure, we see that the two square orbits do not get pulled toward the fixed points. Instead, the two orbits both approach periodic behavior; they oscillate between −1 and 0. These two points, −1 and 0, form a cycle of period 2.

To see that the orbit of −1 is periodic, first, we let f act on −1:

$$f(-1) = (-1)^2 - 1 = 1 - 1 = 0\,. \tag{1.13}$$

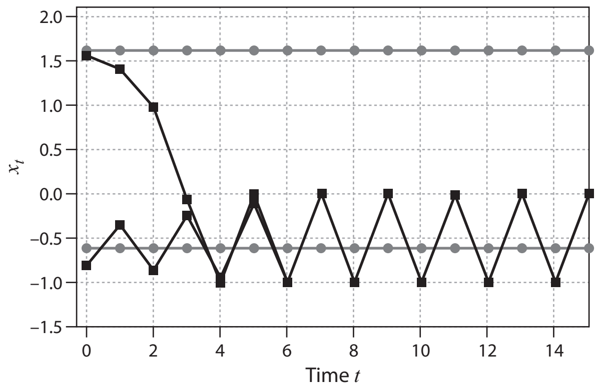

Figure 1.6. The time series plot for the four different initial conditions iterated with $f(x) = x^2 - 1$. The two orbits plotted with grey circles are the fixed points, 1.618 and −0.618.

Then, we let f act on 0:

$$f(0) = 0^2 - 1 = -1. \tag{1.14}$$

Thus, −1 is periodic with period two. The period is two, because it takes two iterations to return to the initial condition.[1] In other words, $f(f(-1)) = -1$. It thus follows, of course, that 0 is also periodic with period two.

The period-two cycle is attracting or stable. Nearby orbits are pulled in to the cycle. Figure 1.7 gives us another way to see this. In this figure I have made time series plots for 200 different initial conditions distributed uniformly between −1.6 and 1.6. One can see in the figure that all initial conditions get pulled quite quickly into the period-two attractor. Not all of the orbits are in phase.

1. The initial condition −1 is also periodic with period *four*, because −1 will return to itself after four iterations. The period of a periodic point is quite sensibly defined to be the smallest number of iterations needed for the point to return to itself. (In more formal mathematical settings, it is common to use the term *prime period* to refer to the smallest number of iterations needed for a point to return to itself. Then one would say that −1 is periodic with period two, four, six, and so on, but that it has a prime period of two.)

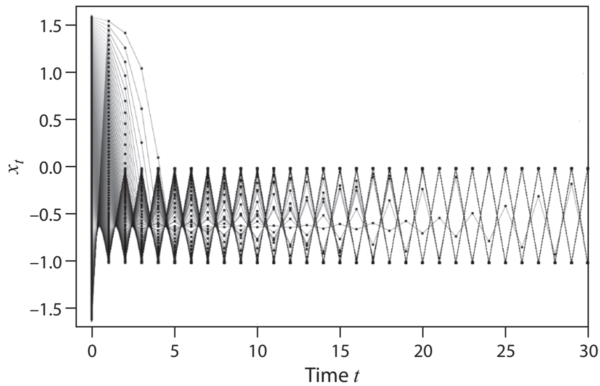

Figure 1.7. The time series plot for 200 initial conditions for the function $f(x) = x^2 - 1$. All initial conditions are pulled toward the period-two cycle at -1 and 0.

At, say, $t = 35$, about half of the orbits will be at (or very near to) -1 and half at 0. But all initial conditions get pulled in to the cycle. Said another way, the long-term behavior of all initial conditions between -1.6 and 1.6 is period two.

Actually, this is not quite right. It is not strictly the case that *all* initial conditions become period-two. The two fixed points, $x \approx 1.61803$ and $x \approx -0.61803$ are not period-two; they are fixed and so remain constant. So I need to amend the statements at the end of the last paragraph. I should have said: *almost* all initial conditions between -1.6 and 1.6 are pulled toward a period-two attractor. The word "almost" in this sentence has a technical meaning: it means that there are infinitely many more points that are pulled toward the period-two attractor than are not. Another way to say this is that if I choose an initial condition at random between -1.6 and 1.6, with probability 1 the orbit will get pulled toward the period-two attractor. This points out again the importance of stability and instability. The two unstable fixed points do not appear at all in Fig. 1.7. In order to observe the unstable behavior I would have to choose an initial condition *exactly* on the fixed point, something that is vanishingly improbable.[2]

To summarize, the function $f(x) = x^2 - 1$ has an attracting cycle of period two: -1 and 0. Equivalently, one says that the cycle is stable. If an orbit is cycling between -1 and 0 and then is perturbed slightly, it will return to the cycle. The function has two fixed points, but they are unstable, and thus do not affect the long-run behavior of almost all initial conditions.

1.6 Determinism

Before concluding, I have a few initial thoughts on an important concept and a recurring theme in the study of dynamical systems: *determinism*. The iterated functions we have studied in this chapter are all examples of *deterministic* dynamical systems. This means that there is no element of chance in the rule. For such dynamical systems the current value of x determines the next value, that value of x then determines the next value, and so on.

Thus, if one knows the function and the initial condition, then the entire future—that is, the itinerary—follows. One might think that deterministic systems are rather dull; once one writes down the rule and specifies the initial condition the story is essentially over. But one of the central lessons of dynamical systems is that deterministic systems still hold plenty of surprises. In Chapter 3 I will make some more extensive remarks on determinism and

2. There is a bit more mathematical fine print. It could also be the case that I chose an initial condition that after a finite number of iterations lands exactly on the unstable fixed point. This is also exceedingly unlikely; it occurs with probability zero. There are a countably infinite number of initial conditions that eventually land exactly on one of the unstable fixed points, but there are an uncountably infinite number of points on the interval between -1.6 and 1.6. Thus, there is zero probability that an orbit lands exactly on one of the fixed points after a finite number of iterations. The bottom line is that we do not expect to observe the unstable fixed points.

related issues. And then in later chapters we will encounter examples of deterministic dynamical systems that behave in ways that are counter-intuitive and produce results that are, in a sense, random.

1.7 Summary

A dynamical system is a mathematical system that changes in time according to a well-specified rule. In this chapter I introduced a simple type of dynamical system: iterated functions. Iterating a function is a repetitious and simple-minded task, requiring only a calculator and a bit of patience. One just applies a rule—in this case a function—to an initial condition over and over and over again.[3] Typically we're interested in a global view of the dynamics. How many fixed points does the dynamical system have and what are their stabilities? What is the long-term behavior of almost all initial conditions? In the examples in this chapter we have seen several types of long-term behavior. An orbit can tend toward positive or negative infinity, or get pulled to an attracting fixed point or an attracting cycle. In Chapter 4 we will see that iterated functions are capable of other, much more complex behavior.

To be honest, I hope this chapter was almost boring. My aim was to introduce a very simple type of dynamical system and to present some key terminology and concepts: initial condition or seed, orbit or itinerary, fixed points, and stable/unstable

3. I have presented the study of iterated functions as an experimental endeavor: choose a seed, grab your calculator, iterate, and see what happens. This is the approach that I'll take in this book. However, there are analytic and less computational approaches to studying the properties of iterated functions. See, e.g., Devaney (1989); Peitgen et al. (1992). These analytic techniques are a lot of fun and are a useful and important complement to the experimental approach I take here.

or attracting/repelling behavior. There is nothing deep or profound in this chapter. We will soon see, however, that simple iterated functions similar to the ones introduced here are capable of surprising—and definitely not boring—behavior. Before doing so, in the next chapter, I will introduce another type of dynamical system: differential equations.