## CHAPTER 1

# Introduction

This chapter is a prelude to this book. It first describes in general terms what the discipline of dynamical systems is about. The following sections contain a large number of examples. Some of the problems treated later in the book appear here for the first time.

## 1.1 DYNAMICS

What is a dynamical system? It is dynamical, something happens, something changes over time. How do things change in nature? Galileo Galilei and Isaac Newton were key players in a revolution whose central tenet is *Nature obeys unchanging laws that mathematics can describe.* Things behave and evolve in a way determined by fixed rules. The prehistory of dynamics as we know it is the development of the laws of mechanics, the pursuit of exact science, and the full development of classical and celestial mechanics. The Newtonian revolution lies in the fact that the principles of nature can be expressed in terms of mathematics, and physical events can be predicted and designed with mathematical certainty. After mechanics, electricity, magnetism, and thermodynamics, other natural sciences followed suit, and in the social sciences quantitative deterministic descriptions also have taken a hold.

### 1.1.1 Determinism Versus Predictability

The key word is determinism: Nature obeys unchanging laws. The regularity of celestial motions has been the primary example of order in nature forever:

> God said, let there be lights in the firmament of the heavens to divide the day from the night and let them be for signs and for seasons and for days and years.

The successes of classical and especially celestial mechanics in the eighteenth and nineteenth centuries were seemingly unlimited, and Pierre Simon de Laplace felt justified in saying (in the opening passage he added to his 1812 *Philosophical Essay* *on Probabilities*):

> We ought then to consider the present state of the universe as the effects of its previous state and as the cause of that which is to follow. An intelligence that, at a given instant, could comprehend all the forces by which nature is animated and the respective situation of the beings that make it up, if moreover it were vast enough to submit these data to analysis, would encompass in the same formula the movements of the greatest bodies of the universe and those of the lightest atoms. For such an intelligence nothing would be uncertain, and the future, like the past, would be open to its eyes.[1]

The enthusiasm in this 1812 overture is understandable, and this forceful description of determinism is a good anchor for an understanding of one of the basic aspects of dynamical systems. Moreover, the titanic life's work of Laplace in celestial mechanics earned him the right to make such bold pronouncements. There are some problems with this statement, however, and a central mission of dynamical systems and of this book is to explore the relation between determinism and predictability, which Laplace's statement misses. The history of the modern theory of dynamical systems begins with Henri Jules Poincaré in the late nineteenth century. Almost 100 years after Laplace he wrote a summary rejoinder:

> If we could know exactly the laws of nature and the situation of the universe at the initial instant, we should be able to predict exactly the situation of this same universe at a subsequent instant. But even when the natural laws should have no further secret for us, we could know the initial situation only *approximately*. If that permits us to foresee the subsequent situation *with the same degree of approximation*, this is all we require, we say the phenomenon has been predicted, that it is ruled by laws. But this is not always the case; it may happen that slight differences in the initial conditions produce very great differences in the final phenomena; a slight error in the former would make an enormous error in the latter. Prediction becomes impossible and we have the fortuitous phenomenon.[2]

His insights led to the point of view that underlies the study of dynamics as it is practiced now and as we present it in this book: The study of long-term asymptotic behavior, and especially that of its qualitative aspects, requires direct methods that do not rely on prior explicit calculation of solutions. And in addition to the qualitative (geometric) study of a dynamical system, probabilistic phenomena play a role.

A major motivation for the study of dynamical systems is their pervasive importance in dealing with the world around us. Many systems evolve continuously in time, such as those in mechanics, but there are also systems that naturally evolve in discrete steps. We presently describe models of, for example, butterfly populations, that are clocked by natural cycles. Butterflies live in the summer, and

1 Pierre Simon marquis de Laplace, *Philosophical Essay on Probabilities*, translated from the fifth French edition of 1925 by Andrew I. Dale, Springer-Verlag, New York, 1995, p. 2.

2 Henri Jules Poincaré, Science et méthode, Section IV.II., Flammarion 1908; see *The Foundations of Science; Science and Hypothesis, The Value of science, Science and Method*, translated by George Bruce Halsted, The Science Press, Lancaster, PA, 1946, pp. 397f; *The Value of Science: Essential Writings of Henri Poincaré*, edited by Stephen Jay Gould, Modern Library, 2001.

we discuss laws describing how next summer's population size is determined by that of this summer. There are also ways of studying a continuous-time system by making it look like a discrete-time system. For example, one might check on the moon's position precisely every 24 hours. Or one could keep track of where it rises any given day. Therefore we allow dynamical systems to evolve in discrete steps, where the same rule is applied repeatedly to the result of the previous step.

This is important for another reason. Such stepwise processes do not only occur in the world around us, but also in our minds. This happens whenever we go through repeated steps on our way to the elusive perfect solution. Applied to such procedures, dynamics provides insights and methods that are useful in analysis. We show in this book that important facts in analysis are consequences of dynamical facts, even of some rather simple ones: The Contraction Principle (Proposition 2.2.8, Proposition 2.2.10, Proposition 2.6.10) gives the Inverse-Function Theorem 9.2.2 and the Implicit-Function Theorem 9.2.3. The power of dynamics in situations of this kind has to do with the fact that various problems can be approached with an iterative procedure of successive approximation by improved guesses at an answer. Dynamics naturally provides the means to understand where such a procedure leads.

### 1.1.2 Dynamics in Analysis

Whenever you use a systematic procedure to improve a guess at a solution you are likely to have found a way of using dynamics to solve your problem exactly. To begin to appreciate the power of this approach it is important to understand that the iterative processes dynamics can handle are not at all required to operate on numbers only. They may manipulate quite complex classes of objects: numbers, points in Euclidean space, curves, functions, sequences, mappings, and so on. The possibilities are endless, and dynamics can handle them all. We use iteration schemes on functions in Section 9.4, mappings in Section 9.2.1 and sequences in Section 9.5. The beauty of these applications lies in the elegance, power, and simplicity of the solutions and insights they provide.

### 1.1.3 Dynamics in Mathematics

The preceding list touches only on a portion of the utility of dynamical systems in understanding mathematical structures. There are others, where insights into certain patterns in some branches of mathematics are most easily obtained by perceiving that underlying the structure in question is something of a dynamical nature that can readily be analyzed or, sometimes, has been analyzed already. This is a range of applications of dynamical ideas that is exciting because it often involves phenomena of a rich subtlety and variety. Here the beauty of applying dynamical systems lies in the variety of behaviors, the surprising discovery of order in bewildering complexity, and in the coherence between different areas of mathematics that one may discover. A little later in this introductory chapter we give some simple examples of such situations.

## EXERCISES

In these exercises you are asked to use a calculator to play with some simple iterative procedures. These are not random samples, and we return to several of these in due course. In each exercise you are given a function $f$ as well as a number $x_0$. The assignment is to consider the sequence defined recursively by the given initial value and the rule $x_{n+1} = f(x_n)$. Compute enough terms to describe what happens in the long run. If the sequence converges, note the limit and endeavor to determine a closed expression for it. Note the number of steps you needed to compute to see the pattern or to get a good approximation of the limit.

**Exercise 1.1.1** $f(x) = \sqrt{2+x}$, $x_0 = 1$.

**Exercise 1.1.2** $f(x) = \sin x$, $x_0 = 1$. Use the *degree* setting on your calculator – this means that (in radians) we actually compute $f(x) = \sin(\pi x/180)$.

**Exercise 1.1.3** $f(x) = \sin x$, $x_0 = 1$. Use the *radian* setting here and forever after.

**Exercise 1.1.4** $f(x) = \cos x$, $x_0 = 1$.

**Exercise 1.1.5**

$$f(x) = \frac{x \sin x + \cos x}{1 + \sin x}, \quad x_0 = 3/4.$$

**Exercise 1.1.6** $f(x) = \{10x\} = 10x - \lfloor 10x \rfloor$ (fractional part), $x_0 = \sqrt{1/2}$.

**Exercise 1.1.7** $f(x) = \{2x\}$, $x_0 = \sqrt{1/2}$.

**Exercise 1.1.8**

$$f(x) = \frac{5 + x^2}{2x}, \quad x_0 = 2.$$

**Exercise 1.1.9** $f(x) = x - \tan x$, $x_0 = 1$.

**Exercise 1.1.10** $f(x) = kx(1-x)$, $x_0 = 1/2$, $k = 1/2, 1, 2, 3.1, 3.5, 3.83, 3.99, 4$.

**Exercise 1.1.11** $f(x) = x + e^{-x}$, $x_0 = 1$.

## 1.2 DYNAMICS IN NATURE

### 1.2.1 Antipodal Rabbits

Rabbits are not indigenous to Australia, but 24 wild European rabbits were introduced by one Thomas Austin near Geelong in Southern Victoria around 1860, with unfortunate consequences. Within a decade they were rampant across Victoria, and within 20 years millions had devastated the land, and a prize of £25,000 was advertized for a solution. By 1910 their descendants had spread across most of the continent. The ecological impact is deep and widespread and has been called a national tragedy. The annual cost to agriculture is estimated at AU$600 million. The unchecked growth of their population makes an interesting example of a dynamical system.

In modeling the development of this population we make a few choices. Its large size suggests to count it in millions, and when the number of rabbits is expressed as $x$ million then $x$ is not necessarily an integer. After all, the initial value is 0.000024 million rabbits. Therefore we measure the population by a real number $x$. As for time, in a mild climate rabbits – famously – reproduce continuously. (This is different for butterflies, say, whose existence and reproduction are strictly seasonal; see Section 1.2.9.) Therefore we are best served by taking the time variable to be a real number as well, $t$, say. Thus we are looking for ways of describing the number of rabbits as a function $x(t)$ of time.

To understand the dependence on time we look at what rabbits do: They eat and reproduce. Australia is large, so they can eat all they want, and during any given time period $\Delta t$ a fixed percentage of the (female) population will give birth and a (smaller) percentage will die of old age (there are no natural enemies). Therefore the increment $x(t+\Delta t) - x(t)$ is proportional to $x(t)\Delta t$ (via the difference of birth and death rates). Taking a limit as $\Delta t \to 0$ we find that

$$\frac{dx}{dt} = kx, \tag{1.2.1}$$

where $k$ represents the (fixed) relative growth rate of the population. Alternatively, we sometimes write $\dot{x} = kx$, where the dot denotes differentiation with respect to $t$. By now you should recognize this model from your calculus class.

It is the unchanging environment (and biology) that gives rise to this unchanging evolution law and makes this a dynamical system of the kind we study. The differential equation (1.2.1), which relates $x$ and its rate of change, is easy to solve: Separate variables (all $x$ on the left, all $t$ on the right) to get $(1/x)dx = k\,dt$ and integrate this with respect to $t$ using substitution:

$$\log|x| = \int \frac{1}{x}\,dx = \int k\,dt = kt + C,$$

where log is the *natural* logarithm. Therefore, $|x(t)| = e^C e^{kt}$ with $e^C = |x(0)|$ and we find that

$$x(t) = x(0)e^{kt}. \tag{1.2.2}$$

**Exercise 1.2.1** Justify the disappearance of the absolute value signs above.

**Exercise 1.2.2** If $x(0) = 3$ and $x(4) = 6$, find $x(2)$, $x(6)$, and $x(8)$.

### 1.2.2 The Leaning Rabbits of Pisa

In the year 1202, Leonardo of Pisa considered a more moderate question regarding rabbits, which we explore in Example 2.2.9 and Section 3.1.9. The main differences to the large-scale Australian model above are that the size of his urban yard limited him to small numbers of rabbits and that with such a small number the population growth does not happen continuously, but in relatively substantial discrete steps. Here is the problem as he posed it:[3]

How many pairs of rabbits can be bred from one pair in one year?

[3] Leonardo of Pisa: Liber abaci (1202), published in *Scritti di Leonardo Pisano*, Rome, B. Boncompagni, 1857; see p. 3 of Dirk J. Struik, *A Source Book in Mathematics 1200–1800*, Princeton, NJ, Princeton University Press, 1986.

> A man has one pair of rabbits at a certain place entirely surrounded by a wall. We wish to know how many pairs can be bred from it in one year, if the nature of these rabbits is such that they breed every month one other pair and begin to breed in the second month after their birth. Let the first pair breed a pair in the first month, then duplicate it and there will be 2 pairs in a month. From these pairs one, namely the first, breeds a pair in the second month, and thus there are 3 pairs in the second month. From these in one month two will become pregnant, so that in the third month 2 pairs of rabbits will be born. Thus there are 5 pairs in this month. From these in the same month 3 will be pregnant, so that in the fourth month there will be 8 pairs ... [We have done this] by combining the first number with the second, hence 1 and 2, and the second with the third, and the third with the fourth ...

In other words, he came up with a sequence of numbers (of pairs of rabbits) governed by the recursion $b_{n+1} = b_n + b_{n-1}$ and chose starting values $b_0 = b_1 = 1$, so the sequence goes 1, 1, 2, 3, 5, 8, 13, .... Does this look familiar? (Hint: As the son of Bonaccio, Leonardo of Pisa was known as filius Bonacci or "son of good nature"; Fibonacci for short.) Here is a question that can be answered easily with a little bit of dynamics: How does his model compare with the continuous exponential-growth model above?

According to exponential growth one should expect that once the terms get large we always have $b_{n+1} \approx ab_n$ for some constant $a$ independent of $n$. If we pretend that we have actual equality, then the recursion formula gives

$$a^2 b_n = ab_{n+1} = b_{n+2} = b_{n+1} + b_n = (a+1)b_n,$$

so we must have $a^2 = a + 1$. The quadratic formula then gives us the value of the growth constant $a$.

■ **Exercise 1.2.3** Calculate $a$.

Note, however, that we have only shown that *if* the growth is eventually exponential, then the growth constant is this $a$, not that the growth is eventually exponential. (If we *assume* the recursion $b_{n+1} = 1$ leads to exponential growth, we could come up with a growth parameter if we are quick enough to do it before getting a contradiction.) Dynamics provides us with tools that enable us to verify this property easily in various different ways (Example 2.2.9 and Section 3.1.9). In Proposition 3.1.11 we even convert this recursively defined sequence into closed form.

The value of this asymptotic ratio was known to Johannes Kepler. It is the golden mean or the divine proportion. In his 1619 book *Harmonices Mundi* he wrote (on page 273):

> there is the ratio which is never fully expressed in numbers and cannot be demonstrated by numbers in any other way, except by a long series of numbers gradually approaching it: this ratio is called *divine*, when it is perfect, and it rules in various ways throughout the dodecahedral wedding. Accordingly, the following consonances begin to shadow forth that ratio: 1:2 and 2:3 and 3:5 and 5:8. For it exists most imperfectly in 1:2, more perfectly in 5:8, and still more perfectly if we add 5 and 8 to make 13 and take 8 as the numerator ....[4]

[4] Johannes Kepler, *Epitome of Copernican Astronomy & Harmonies of the World*, Amherst, NY, Prometheus Books, 1995.

We note in Example 15.2.5 that these Fibonacci ratios are the optimal rational approximations of the golden mean.

■ **Exercise 1.2.4** Express $1 + 1 + 2 + 3 + \cdots + b_n$ in terms of $b_{n+2}$.

### 1.2.3 Fine Dining

Once upon a time lobsters were so abundant in New England waters that they were poor man's food. It even happened that prisoners in Maine rioted to demand to be fed something other than lobsters for a change. Nowadays the haul is less abundant and lobsters have become associated with fine dining. One (optimistic?) model for the declining yields stipulates that the catch in any given year should turn out to be the average of the previous two years' catches.

Using again $a_n$ for the number of lobsters caught in the year $n$, we can express this model by a simple recursion relation:

$$a_{n+1} = a_{n-1}/2 + a_n/2. \tag{1.2.3}$$

As initial values one can take the Maine harvests of 1996 and 1997, which were 16,435 and 20,871 (metric) tons, respectively. This recursion is similar to the one for the Fibonacci numbers, but in this case no exponential growth is to be expected. One can see from the recursion that all future yields should be between the two initial data. Indeed, 1997 was a record year. In Proposition 3.1.13 we find a way of giving explicit formulas for future yields, that is, we give the yield in an arbitrary year $n$ in a closed form as a function of $n$.

This situation as well as the Fibonacci rabbit problem are examples where time is measured in discrete steps. There are many other examples where this is natural. Such a scenario from population biology is discussed in Section 1.2.9. Other biological examples arise in genetics (gene frequency) or epidemiology. Social scientists use discrete-time models as well (commodity prices, rate of spread of a rumor, theories of learning that model the amount of information retained for a given time).

### 1.2.4 Turning Over a New Leaf

The word phyllotaxis comes from the words phyllo=leaf and taxis=order or arrangement. It refers to the way leaves are arranged on twigs, or other plant components on the next larger one. The seeds of a sunflower and of a pine cone are further examples. A beautiful description is given by Harold Scott Macdonald Coxeter in his *Introduction to Geometry*. That regular patterns often occur is familiar from sunflowers and pineapples.

In some species of trees the leaves on twigs are also arranged in regular patterns. The pattern varies by species. The simplest pattern is that of leaves alternating on opposite sides of the twig. It is called (1, 2)-phyllotaxis: Successive leaves are separated by a half-turn around the twig. The leaves of elms exhibit this pattern, as do hazel leaves.[5] Adjacent leaves may also have a (2/3) turn between them, which would be referred to as (2, 3)-phyllotaxis. Such is the case with beeches. Oak trees

[5] On which the first author of this book should be an expert!

show a (3, 5)-pattern, poplars a (5, 8), and willows, (8, 13)-phyllotaxis. Of course, the pattern may not always be attained to full precision, and in some plants there are transitions between different patterns as they grow.

The diamond-shaped seeds of a sunflower are packed densely and regularly. One may perceive a spiral pattern in their arrangement, and, in fact, there are always two such patterns in opposite directions. The numbers of spirals in the two patterns are successive Fibonacci numbers. The seeds of a fir cone exhibit spirals as well, but on a cone rather than flat ones. These come in two families, whose numbers are again successive Fibonacci numbers.

Pineapples, too, exhibit spiral patterns, and, because their surface is composed of approximately hexagonal pieces, there are three possible directions in which one can perceive spirals. Accordingly, one may find 5, 8, and 13 spirals: 5 sloping up gently to the right, say, 8 sloping up to the left, and 13 sloping quite steeply right.

The observation and enjoyment of these beautiful patterns is not new. They were noticed systematically in the nineteenth century. But an explanation for why there are such patterns did not emerge particularly soon. In fact, the case is not entirely closed yet.

Here is a model that leads to an explanation of how phyllotaxis occurs. The basic growth process of this type consists of buds (primordia) of leaves or seeds growing out of a center and then moving away from it according to three rules proposed in 1868 by the self-taught botanist Wilhelm Friedrich Benedikt Hofmeister, while he was professor and director of the botanical garden in Heidelberg:

(1) New buds form at regular intervals, far from the old ones.
(2) Buds move radially from the center.
(3) The growth rate decreases as one moves outward.

A physical experiment designed to mimic these three *Hofmeister rules* produces spiral patterns of this Fibonacci type, so from these rules one should be able to infer that spiral patterns must occur. This has been done recently with methods of the kind that this book describes.[6]

Here is a description of how dynamics may help. To implement the Hofmeister rules we model the situation by a family of $N+1$ concentric circles of radius $r^k$ ($k = 0, \dots, N$), where $r$ stands for growth rate, and we put a bud on each circle. The angle (with respect to the origin) between one bud and the next is $\theta_k$. Possible patterns are now parametrized by angles $(\theta_0, \dots, \theta_N)$. This means that the "space of plants" is a *torus*; see Section 2.6.4. When a new bud appears on the unit circle, all other buds move outward one circle. The angle of the new bud depends on all previous angles, so we get a map sending old angles $\theta_k$ to new angles $\Theta_k$ by

$$\Theta_0 = f(\theta_0, \dots, \theta_N), \quad \Theta_1 = \theta_0, \dots, \Theta_N = \theta_{N-1}.$$

Now $f$ has to be designed to reflect the first Hofmeister rule. One way to do this is to define a natural potential energy to reflect "repulsion" between buds and choosing

[6] Pau Atela, Christophe Golé, and Scott Hotton: A dynamical system for plant pattern formation: A rigorous analysis, *Journal of Nonlinear Science* **12** (2002), no. 6, pp. 641–676.

$f(\theta_0, \ldots, \theta_N)$ to be the minimum. A natural potential is

$$W(\Theta) = \sum_{k=0}^{N} U(\|r^k e^{i\theta_k} - e^{i\Theta}\|),$$

where $U(x) = 1/x^s$ for some $s > 0$. A simpler potential that gives the same qualitative behavior is $W(\Theta) = \max_{0 \le k \le N} U(\|r^k e^{i\theta_k} - e^{i\Theta}\|)$. With either choice one can show that regular spirals (that is, $\theta_0 = \cdots = \theta_N$) are attracting fixed points (Section 2.2.7) of this map. This means that spirals will appear naturally. A result of the analysis is furthermore that the Fibonacci numbers also must appear.

### 1.2.5 Variations on Exponential Growth

In the example of a rabbit population of Section 1.2.1 it is natural to expect a positive growth parameter $k$ in the equation $\dot{x} = kx$. This coefficient, however, is the difference between rates of reproduction and death. For the people of some western societies, the reproduction rate has declined so much as to be lower than the death rate. The same model still applies, but with $k < 0$ the solution $x(t) = x(0)e^{kt}$ describes an exponentially shrinking population.

The same differential equation $\dot{x} = kx$ comes up in numerous simple models because it is the simplest differential equation in one variable.

Radioactive decay is a popular example: It is an experimental fact that of a particular radioactive substance a specific percentage will decay in a fixed time period. As before, this gives $\dot{x} = kx$ with $k < 0$. In this setting the constant $k$ is often specified by the *half-life*, which is the time $T$ such that $x(t + T) = x(t)/2$. Depending on the substance, this time period may be minute fractions of a second to thousands of years. This is important in regard to the disposal of radioactive waste, which often has a long half-life, or radioactive contamination. Biology laboratories use radioactive phosphorus as a marker, which has a half-life of a moderate number of days. A spill on the laboratory bench is usually covered with plexiglas for some two weeks, after which the radiation has sufficiently diminished. On the other hand, a positive effect of radioactive decay is the possibility of radioisotope dating, which can be used to assess the age of organic or geologic samples. Unlike in population biology, the exponential decay model of radioactivity needs no refinements to account for real data. It is an exact law of nature.

■ **Exercise 1.2.5** Express the half-life in terms of $k$, and vice versa.

The importance of the simple differential equation $\dot{x} = kx$ goes far beyond the collection of models in which it appears, however many of these there may be. It also comes up in the study of more complicated differential equations as an approximation that can illuminate some of the behavior in the more complicated setting. This approach of *linearization* is of great importance in dynamical systems.

### 1.2.6 The Doomsday Model

We now return to the problem of population growth. Actual population data show that the world population has grown with increasing rapidity. Therefore we should consider a modification of the basic model that takes into account the progress of civilization. Suppose that with the growth of the population the growing number of researchers manages to progressively decrease the death rate and increase fertility as well. Assuming, boldly, that these improvements make the relative rate of increase in population a small positive power $x^\epsilon$ of the present size $x$ (rather than being constant $k$), we find that

$$\frac{dx}{dt} = x^{1+\epsilon}.$$

As before, this is easy to solve by separating variables:

$$t + C = \int x^{-1-\epsilon}\,dx = -x^{-\epsilon}/\epsilon$$

with $C = -x(0)^{-\epsilon}/\epsilon$, so $x(t) = (x(0)^{-\epsilon} - \epsilon t)^{-1/\epsilon}$, which becomes infinite for $t = 1/(\epsilon x(0)^\epsilon)$. Population explosion indeed!

As far as biology is concerned, this suggests refining our model. Clearly, our assumptions on the increasing growth rate were too generous (ultimately, resources are limited). As an example in differential equations this is instructive, however: There are reasonable-looking differential equations that have divergent solutions.

### 1.2.7 Predators

The reason rabbits have not over taken over the European continent is that there have always been predators around to kill rabbits. This has interesting effects on the population dynamics, because the populations of predators and their prey interact: A small number of rabbits decreases the predator population by starvation, which tends to increase the rabbit population. Thus one expects a stable equilibrium – or possibly oscillations.

Many models of interacting populations of predator and prey were proposed independently by Alfred Lotka and Vito Volterra. A simple one is the *Lotka–Volterra equation*:

$$\frac{dx}{dt} = a_1 x + c_1 xy$$
$$\frac{dy}{dt} = a_2 x + c_2 xy,$$

where $a_1, c_2 > 0$ and $a_2, c_1 < 0$, that is, $x$ is the prey population, which would grow on its own ($a_1 > 0$) but is diminished by the predator ($c_1 < 0$), while $y$ is the predator, which would starve if alone ($a_2 < 0$) and grows by feeding on its prey ($c_2 > 0$). Naturally, we take $x$ and $y$ positive. This model assumes that there is no delay between causes and effects due to the time of gestation or egg incubation. This is reasonable when the time scale of interest is not too short. Furthermore, choosing time continuously is most appropriate when generations overlap substantially. Populations with nonoverlapping generations will be treated shortly.

There is an equilibrium of species at $(a_2/c_2, a_1/c_1)$. Any other initial set of populations turns out to result in oscillations of the numbers of predator and prey. To see this, use the chain rule to verify that

$$E(x, y) := x^{-a_2} e^{-c_2 x} y^{a_1} e^{c_1 y}$$

is constant along orbits, that is, $(d/dt)E(x(t), y(t)) = 0$. This means that the solutions of the Lotka–Volterra equation must lie on the curves $E(x, y) = \text{const}$. These curves are closed.

### 1.2.8 Horror Vacui

The Lotka–Volterra equation invites a brief digression to a physical system that shows a different kind of oscillatatory behavior. Its nonlinear oscillations have generated much interest, and the system has been important for some developments in dynamics.

The Dutch engineer Balthasar van der Pol at the Science Laboratory of the Philips Light Bulb Factory in Eindhoven modeled a vacuum tube circuit by the differential equation

$$\frac{d^2x}{dt^2} + \epsilon(x^2 - 1)\frac{dx}{dt} + x = 0,$$

which can be rewritten using $y = dx/dt$ as

$$\frac{dx}{dt} = y$$
$$\frac{dy}{dt} = \epsilon(1 - x^2)y - x.$$

If $\epsilon = 1$, the origin is a repeller (Definition 2.3.6). However, solutions do not grow indefinitely, because there is a periodic solution that circles around the origin. Indeed, for $\epsilon = 0$ there are only such solutions, and for $\epsilon = 1$ one of these circles persists in deformed shape, and all other solutions approach it ever more closely as $t \to +\infty$. The numerically computed picture in Figure 1.2.1 shows this clearly. The curve is called a *limit cycle.*

As an aside we mention that there is also the potential for horrifying complexity in a vacuum tube circuit. In 1927, van der Pol and J. van der Mark reported on experiments with a "relaxation oscillator" circuit built from a capacitor and a neon lamp (this is the nonlinear element) and a periodic driving voltage. (A driving voltage corresponds to putting a periodic term on the right-hand side of the van der Pol equation above.) They were interested in the fact that, in contrast to a linear oscillator (such as a violin string), which exhibits multiples of a base frequency, these oscillations were at "submultiples" of the basic frequency, that is, half that frequency, a third, and so on down to 1/40th, as the driving voltage increased. They

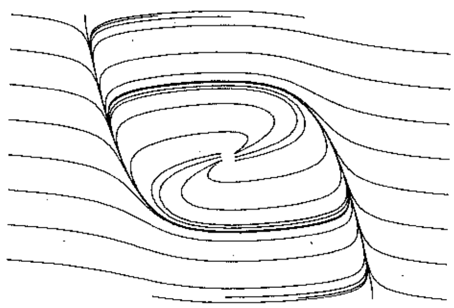

Figure 1.2.1. The van der Pol equation.

obtained these frequencies by listening "with a telephone coupled loosely in some way to the system" and reported that

> Often an irregular noise is heard in the telephone receivers before the frequency jumps to the next lower value. However, this is a subsidiary phenomenon, the main effect being the regular frequency demultiplication.

This irregular noise was one of the first experimental encounters with what was to become known as chaos, but the time was not ripe yet.[7]

### 1.2.9 The Other Butterfly Effect[8]

Population dynamics is naturally done in discrete-time steps when generations do not overlap. This was imposed somewhat artificially in the problem posed by Leonardo of Pisa (Section 1.2.2). For many populations this happens naturally, especially insects in temperate zones, including many crop and orchard pests. A pleasant example is a butterfly colony in an isolated location with a fairly constant seasonal cycle (unchanging rules and no external influence). There is no overlap at all between the current generation (this summer) and the next (next summer). We would like to know how the size of the population varies from summer to summer. There may be plenty of environmental factors that affect the population, but by assuming unchanging rules we ensure that next summer's population depends only on this summer's population, and this dependence is the same every year. That means that the only parameter in this model that varies at all is the population itself. Therefore, up to choosing some fixed constants, the evolution law will specify the population size next summer as a function of this summer's population only. The specific evolution law will result from modeling this situation according to our understanding of the biological processes involved.

**1. Exponential growth.** For instance, it is plausible that a larger population is likely to lay more eggs and produce a yet larger population next year, proportional, in fact, to the present population. Denoting the present population by $x$, we then find that next year's population is $f(x) = kx$ for some positive constant $k$, which is the average number of offspring per butterfly. If we denote the population in year $i$ by $x_i$, we therefore find that $x_{i+1} = f(x_i) = kx_i$ and in particular that $x_1 = kx_0$, $x_2 = kx_1 = k^2x_0$, and so on, that is, $x_i = k^i x_0$; the population grows exponentially. This looks much like the exponential–growth problem as we analyzed it in continuous time.

**2. Competition.** A problem familiar from public debate is sustainability, and the exponential growth model leads to large populations relatively rapidly. It is more realistic to take into account that a large population will run into problems with limited food supplies. This will, by way of malnutrition or starvation, reduce the

[7] B. van der Pol, J. van der Mark, Frequency demultiplication, *Nature* **120** (1927), 363–364.

[8] This is a reference to the statement of Edward Lorenz (see Section 13.3) that a butterfly may flutter by in Rio and thereby cause a typhoon in Tokyo a week later. Or maybe to butterfly ballots in the 2000 Florida election?

number of butterflies available for egg-laying when the time comes. A relatively small number of butterflies next year is the result.

The simplest rule that incorporates such more sensible qualitative properties is given by the formula $f(x) = k(1 - \alpha x)x$, where $x$ is the present number of butterflies. This rule is the simplest because we have only adduced a linear correction to the growth rate $k$. In this correction $\alpha$ represents the rate at which fertility is reduced through competition. Alternatively, one can say that $1/\alpha$ is the maximal possible number of butterflies; that is, if there are $1/\alpha$ butterflies this year, then they will eat up all available food before getting a chance to lay their eggs; hence they will starve and there will be no butterflies next year. Thus, if again $x_i$ denotes the butterfly population in the year $i$, starting with $i = 0$, then the evolution is given by $x_{i+1} = kx_i(1 - \alpha x_i) =: f(x_i)$. This is a deterministic mathematical model in which every future state (size of the butterfly colony) can be computed from this year's state. One drawback is that populations larger than $1/\alpha$ appear to give negative populations the next year, which could be avoided with a model such as $x_{i+1} = x_i e^{k(1-x_i)}$. But tractability makes the simpler model more popular, and it played a significant role in disseminating to scientists the important insight that simple models can have complicated long-term behaviors.[9]

One feature reminiscent of the exponential-growth model is that, for populations much smaller than the limit population, growth is indeed essentially exponential: If $\alpha x \ll 1$, then $1 - \alpha x \approx 1$ and thus $x_{i+1} \approx kx_i$; hence $x_n \approx k^n x_0$ – but only so long as the population stays small. This makes intuitive sense: The population is too small to suffer from competition for food, as a large population would.

Note that we made a slip in the previous paragraph: The sequence $x_n \approx k^n x_0$ grows exponentially *if* $k > 1$. If this is not the case, then the butterfly colony becomes extinct. An interesting interplay between reproduction rates and the carrying capacity influences the possibilities here.

**3. Change of variable.** To simplify the analysis of this system it is convenient to make a simple change of variable that eliminates the parameter $\alpha$. We describe it with some care here, because changing variables is an important tool in dynamics.

Write the evolution law as $x' = kx(1 - \alpha x)$, where $x$ is the population in one year and $x'$ the population in the next year. If we rescale our units by writing $y = \alpha x$, then we must set

$$y' = \alpha x' = \alpha kx(1 - \alpha x) = ky(1 - y).$$

In other words, we now iterate the map $g(y) = ky(1 - y)$. The relationship between the maps $f$ and $g$ is given by $g(y) = h^{-1}(f(h(y)))$, where $h(y) = y/\alpha = x$. This can be read as "go from new variable to old, apply the old map, and then go to the new variable again."

[9] As its title shows, getting this message across was the aim of an influential article by Robert M. May, Simple Mathematical Models with Very Complicated Dynamics, *Nature* **261** (1976), 459–467. This article also established the quadratic model as the one to be studied. A good impression of the effects on various branches of biology is given by James Gleick, *Chaos, Making a New Science*, Viking Press, New York, 1987, pp. 78ff.

The effect of this change of variable is to normalize the competition factor $\alpha$ to 1. Since we never chose specific units to begin with, let's rename the variables and maps back to $x$ and $f$.

**4. The logistic equation.** We have arrived at a model of this system that is represented by iterations of

$$f(x) = kx(1-x).$$

This map $f$ is called the *logistic map* (or logistic family, because there is a parameter), and the equation $x' = kx(1-x)$ is called the logistic equation. The term logistic comes from the French *logistique*, which in turn derived from *logement*, the lodgment of soldiers. We also refer to this family of maps as the *quadratic family*. It was introduced in 1845 by the Belgian sociologist and mathematician Verhulst.[10]

From the brief discussion before the preceding subsection it appears that the case $k \le 1$ results in inevitable extinction. This is indeed the case. For $k < 1$, this is clear because $kx(1-x) < kx$, and for $k = 1$ it is not hard to verify either, although the population decay is not exponential in this case. By contrast, large values of $k$ should be good for achieving a large population. Or maybe not. The problem is that too large a population will be succeeded by a less numerous generation. One would hope that the population settles to an agreeable size in due time, at which there is a balance between fertility and competition.

**Exercise 1.2.6** Prove that the case $k = 1$ results in extinction.

Note that, unlike in the simpler exponential growth model, we now refrained from writing down an explicit formula for $x_n$ in terms of $x_0$. This formula is given by polynomials of order $2^n$. Even if one were to manage to write them down for a reasonable $n$, the formulas would not be informative. We will, in due course, be able to say quite a bit about the behavior of this model. At the moment it makes sense to explore it a little to see what kind of behavior occurs. Whether the initial size of the population matters, we have not seen yet. But changing the parameter $k$ certainly is likely to make a difference, or so one would hope, because it would be a sad model indeed that predicts certain extinction all the time. The reasonable range for $k$ is from 0 to 4. [For $k > 4$, it predicts that a population size of 1/2 is followed two years later by a negative population, which makes little biological sense. This suggests that a slightly more sophisticated (nonlinear) correction rule would be a good idea.]

**5. Experiments.** Increasing $k$ should produce the possibility of a stable population, that is, to allow the species to avoid extinction. So let's start working out the model for some $k > 1$. A simpleminded choice would be $k = 2$, halfway between 0 and 4.

**Exercise 1.2.7** Starting with $x = 0.01$, iterate $2x(1-x)$ until you discern a clear pattern.

[10] Pierre-François Verhulst, Récherches mathématiques sur la loi d'accroissement de la population, *Nouvelles Mémoires de l'Academie Royale des Sciences et Belles-Lettres de Bruxelles* **18** (1845), 1–38.

Starting from a small population, one obtains steady growth and eventually the population levels off at 1/2. This is precisely the behavior one should expect from a decent model. Note that steady states satisfy $x = 2x(1-x)$, of which 0 and 1/2 are the only solutions.

**Exercise 1.2.8** Starting with $x = 0.01$ iterate $1.9x(1-x)$ and $2.1x(1-x)$ until you discern a clear pattern.

If $k$ is a little less than 2, the phenomenon is rather the same, for $k$ a little bigger it also goes that way, except for slightly overshooting the steady-state population.

**Exercise 1.2.9** Starting with $x = 0.01$, iterate $3x(1-x)$ and $2.9x(1-x)$ until you discern a clear pattern.

For $k = 3$, the ultimate behavior is about the same, but the way the population settles down is a little different. There are fairly substantial oscillations of too large and too small population that die out slowly, whereas for $k$ near 2 there was only a hint of this behavior, and it died down fast. Nevertheless, an ultimate steady state still prevails.

**Exercise 1.2.10** Starting with $x = 0.01$, iterate $3.1x(1-x)$ until you discern a clear pattern.

For $k = 3.1$, there are oscillations of too large and too small as before. They do get a little smaller, but this time they do not die down all the way. With a simple program one can iterate this for quite a while and see that no steady state is attained.

**Exercise 1.2.11** Starting with $x = 0.66$, iterate $3.1x(1-x)$ until you discern a clear pattern.

In the previous experiment, there is the possibility that the oscillations die down so slowly that the numerics fail to notice. Therefore, as a control, we start the same iteration at the average of the two values. This should settle down if our diagnosis is correct. But it does not. We see oscillations that grow until their size is as it was before.

These oscillations are stable! This is our first population model that displays persistent behavior that is not monotonic. No matter at which size you start, the species with fertility 3.1 is just a little too fertile for its own good and keeps running into overpopulation every other year. Not by much, but forever.

Judging from the previous increments of $k$ there seems only about $k = 4$ left, but to be safe let's first try something closer to 3 first. At least it is interesting to see whether these oscillations get bigger with increasing $k$. They should. And how big?

**Exercise 1.2.12** Starting with $x = 0.66$, iterate $3.45x(1-x)$ and $3.5x(1-x)$ until you discern a clear pattern.

The behavior is becoming more complicated around $k = 3.45$. Instead of the simple oscillation between two values, there is now a secondary dance around each of these values. The oscillations now involve four population sizes: "Big, small, big, Small" repeated in a 4-cycle. The period of oscillation has doubled.

**Exercise 1.2.13** Experiment in a manner as before with parameters slightly larger than 3.5.

A good numerical experimenter will see some pattern here for a while: After a rather slight parameter increase the period doubles again; there are now eight population sizes through which the model cycles relentlessly. A much more minute increment brings us to period 16, and it keeps getting more complicated by powers of two. This cascade of period doublings is complementary to what one sees in a linear oscillator such as a violin string or the column of air in wind instruments or organ pipes: There it is the frequency that has higher harmonics of double, triple, and quadruple the base frequency. Here the frequency is halved successively to give *subharmonics*, an inherently nonlinear phenomenon.

Does this period doubling continue until $k = 4$?

**Exercise 1.2.14** Starting with $x = .5$, iterate $3.83x(1-x)$ until you discern a clear pattern.

When we look into $k = 3.83$ we find something rather different: There is a periodic pattern again, which we seem to have gotten used to. But the period is 3, not a power of 2. So this pattern appeared in an entirely different way. And we don't see the powers of 2, so these must have run their course somewhat earlier.

**Exercise 1.2.15** Try $k = 3.828$.

No obvious pattern here.

**Exercise 1.2.16** Try $k = 4$.

There is not much tranquility here either.

**6. Outlook.** In trying out a few parameter values in the simplest possible nonlinear population model we have encountered behavior that differs widely for different parameter values. Where the behavior is somewhat straightforward we do not have the means to explain how it evolves to such patterns: Why do periods double for a while? Where did the period-3 oscillation come from? And at the end, and in experiments with countless other values of the parameter you may choose to try, we see behavior we cannot even describe effectively for lack of words. At this stage there is little more we can say than that in those cases the numbers are all over the place.

We return to this model later (Section 2.5, Section 7.1.2, Section 7.4.3 and Chapter 11) to explain some of the basic mechanisms that cause these diverse behaviors in the quadratic family $f_k(x) = kx(1 - x)$. We do not provide an exhaustive analysis that covers all parameter values, but the dynamics of these maps is quite well understood. In this book we develop important concepts that are needed to describe the complex types of behavior one can see in this situation, and in many other important ones.

Already this purely numerical exploration carries several lessons. The first one is that *simple systems can exhibit complex long-term behavior*. Again, we arrived at this example from the linear one by making the most benign change possible. And immediately we ran into behavior so complex as to defy description. Therefore such complex behavior is likely to be rather more common than one would have thought.

The other lesson is that it is worth learning about ways of understanding, describing, and explaining such rich and complicated behavior. Indeed, the important insights we introduce in this book are centered on the study of systems where explicit computation is not feasible or useful. We see that even in the absence of perfectly calculated results for all time one can make precise and useful qualitative and quantitative statements about such dynamical systems. Part of the work is to develop concepts adequate for a description of phenomena of such complexity as we have begun to glimpse in this example. Our study of this particular example begins in Section 2.5, where we study the simple behaviors that occur for small parameter values. In Section 7.1.2 and Section 7.4.3 we look at large parameter values. For these the asymptotic behavior is most chaotic. In Chapter 11 we present some of the ideas used in understanding the intermediate parameter regime, where the transitions to maximal complexity occur.

As an interesting footnote we mention that the analogous population with continuous time (which is quite reasonable for other species) has none of this complexity (see Section 2.4.2).

### 1.2.10 A Flash of Inspiration

As another example of dynamics in nature we can take the flashing of fireflies. Possibly the earliest report of a remarkable phenomenon is from Sir Francis Drake's 1577 expedition:

> [o]ur general... sailed to a certaine little island to the southwards of Celebes,... throughly growen with wood of a large and high growth.... Among these trees night by night, through the whole land, did shew themselves an infinite swarme of fiery wormes flying in the ayre, whose bodies beeing no bigger than our common English flies, make such a shew of light, as if every twigge or tree had been a burning candle.[11]

A clearer description of what is so remarkable about these fireflies was given by Engelbert Kämpfer, a doctor from eastern Westphalia who made a 10-year voyage through Russia, Persia, southeast Asia, and Japan. On July 6, 1690, he traveled down the Chao Phraya (Meinam) River from Bangkok and observed:

> The glowworms (Cicindelae) represent another shew, which settle on some trees, like a fiery cloud, with this surprising circumstance, that a whole swarm of these

[11] Richard Hakluyt (pronounced Hack-loot), *A Selection of the Principal Voyages, Traffiques and Discoveries of the English Nation*, edited by Laurence Irving, Knopf, New York, 1926.

> insects, having taken possession of one tree, and spread themselves over its branches, sometimes hide their light all at once, and a moment after make it appear again with the utmost regularity and exactness, as if they were in perpetual systole and diastole.[12]

So, in some locations large numbers of the right species of flashing fireflies in a bush or a collection of bushes synchronize, turning their arboreal home into a remarkable christmas-tree-like display. Or do they? This is such a striking phenomenon that for a long time reports of it had a status not entirely unlike that of tales of dragons and sea monsters. As late as 1938 they were not universally accepted among biologists. Only with increased affordability and speed of travel could doubters see it for themselves.[13] Once there was some belief that this really happens, it took many decades to develop an understanding of how this is possible. Early on it was supposed that some subtle and undetected external periodic influence caused this uniform behavior, but it is the fact that these fireflies naturally flash at close to the same rate combined with a tendency to harmonize with the neighbors that causes an entire colony to wind up in perfect synchrony.

An analogous situation much closer to home is the study of circadian rhythms, where periodic changes in our body (the sleep cycle) synchronize with the external cues of day and night. In the absence of clocks and other cues to the time of day, the human wake–sleep cycle reverts to its natural period, which is for most people slightly longer than 24 hours. Those external cues affect the system of neurons and hormones that make up our complicated internal oscillator and gently bring it up to speed. In this case, the rate at which the adjustment happens is fairly quick. Even the worst jet lag usually passes within a few days, that is, a few cycles.

These systems are instances of coupled oscillators, which also appear in numerous other guises. The earth–moon system can be viewed as such a system when one looks for an explanation why we always see the same side of the moon, that is, why the moon's rotation and revolution are synchronized. Here simple tidal friction is the coupling that has over eons forced the moon's rotation into lockstep with its revolution and will eventually synchronize the earth's rotation as well, so a day will be a month long – or a month a day long, making the moon a geostationary satellite. It is amusing to think that at some intermediate time the longer days may match up with our internal clocks, as if human evolution is slightly ahead of its time on this count.

[12] Engelbert Kämpfer, *The history of Japan*, edited by J. G. Scheuchzer, Scheuchzer, London, 1727. The translation is not too good. The German original apparently remained unpublished for centuries: "Einen zweiten sehr angenehmen Anblik geben die Lichtmücken (cicindelae), welche einige Bäume am Ufer mit einer Menge, wie eine brennende Wolke, beziehn. Es war mir besonders hiebei merkwürdig, daß die ganze Schaar dieser Vögel, so viel sich ihrer auf einem Baume verbunden, und durch alle Aeste desselben verbreitet haben, alle zugleich und in einem Augenblik ihr Licht verbergen und wieder von sich geben, und dies mit einer solchen Harmonie, als wenn der Baum selbst in einer beständigen Systole und Diastole begriffen wäre." (Geschichte und Beschreibung von Japan (1677–79). Internet Edition by Wolfgang Michel. In: Engelbert-Kaempfer-Forum, Kyushu University, 1999.).

[13] An account of this sea change is given by John Buck, Synchronous rhythmic flashing of fireflies, *Quarterly Review of Biology* **13**, no. 3 (September 1938), 301–314; II, *Quarterly Review of Biology* **63**, no. 3 (September 1988), 265–289. The articles include the quotes given here and many more reports of flashing fireflies from various continents.

We will look at systems made up of two simple oscillators in Section 4.4.5, where relatively simple considerations suggest that this kind of synchronization is somewhat typical.[14]

## EXERCISES

**Exercise 1.2.17** In 1900, the global human population numbered 1.65 million, and in 1950 it was 2.52 billion. Use the exponential growth model (Equation 1.2.2) to predict the population in 1990, and to predict when the population will reach 6 billion. (The actual 1990 population was some 5.3 billion, and around July 1999 it reached 6 billion. Thus the growth of the world population is accelerating.)

**Exercise 1.2.18** Denote by $a_n$ the number of sequences of 0's and 1's of length $n$ that do not have two consecutive 0's. Show that $a_{n+1} = a_n + a_{n-1}$. (Note that this is the same recursion as for the Fibonacci numbers, and that $a_1 = 2$ and $a_2 = 3$.)

**Exercise 1.2.19** Show that any two successive Fibonacci numbers are relatively prime.

**Exercise 1.2.20** Determine $\lim_{n\to\infty} a_n$ in (1.2.3) if $a_0 = 1$ and $a_1 = 0$.

## 1.3 DYNAMICS IN MATHEMATICS

In this section we collect a few examples of a range of mathematical activity where knowledge of dynamical systems provides novel insights.

### 1.3.1 Heroic Efforts with Babylonian Roots

Sometime before 250 A.D., in his textbook *Metrica*, Heron of Alexandria (often latinized to Hero of Alexandria) computed the area of a triangle with sides 7, 8, and 9 by first deriving the formula $\text{area}^2 = s(s-a)(s-b)(s-c)$, where $a, b, c$ are the sides and $2s = a + b + c$. To compute the resulting square root of $12 \cdot 5 \cdot 4 \cdot 3 = 720$ he took the following approach, which may have been known to the Babylonians 2000 years before:

> Since $[z=]720$ has not its side rational [that is, 720 is not a perfect square], we can obtain its side within a very small difference as follows. Since the next succeeding square number is 729, which has $[x=]27$ for its side, divide 720 by 27. This gives $[y=]26\frac{2}{3}$. Add 27 to this, making $53\frac{2}{3}$, and take half of this or $[x' = \frac{1}{2}(x+y) =]26\frac{1}{2}\,\frac{1}{3}$. The side of 720 will therefore be very nearly $26\frac{1}{2}\,\frac{1}{3}$... If we desire to make the difference still smaller... we shall take $[x' = \frac{1}{2}(x+y) = 26\frac{1}{2}\,\frac{1}{3} = 26\frac{5}{6}$ instead of $x = 27]$ and by proceeding in the same way we shall find that the resulting difference is much less...[15]

Heron used that, in order to find the square root of $z$, it suffices to find a square with area $z$; its sides have length $\sqrt{z}$. A geometric description of his procedure is

[14] We omit a full treatment of coupled linear oscillators. The subject of fireflies is treated by Renato Mirollo and Steven Strogatz, Synchronization of Pulse-Coupled Biological Oscillators, *SIAM Journal of Applied Mathematics* **50** no. 6 (1990), 1645–1662.

[15] Thomas L. Heath, *History of Greek Mathematics: From Aristarchus to Diophantus*, Dover, 1981, p. 324. This sequence of approximations also occurs in Babylonian texts; as related by Bartels van der Waerden: *Science awakening*, Oxford University Press, Oxford, 1961, p. 45, who gives a geometric interpretation on pp. 121ff. Some variant was known to Archimedes.

that as a first approximation of the desired square we take a rectangle with sides $x$ and $y$, where $x$ is an educated guess at the desired answer and $xy = z$. (If $z$ is not as large as in Heron's example, one can simply take $x = 1$, $y = z$.) The procedure of producing from a rectangle of correct area another rectangle of the same area whose sides differ by less is to replace the sides $x$ and $y$ by taking one side to have the average length $(x+y)/2$ (arithmetic mean) and the other side to be such as to get the same area as before: $2xy/(x+y)$ (this is called the harmonic mean of $x$ and $y$). The procedure can be written simply as repeated application of the function

$$f(x, y) = \left(\frac{x+y}{2}, \frac{2xy}{x+y}\right) \tag{1.3.1}$$

of two variables starting with $(x_0, y_0) = (z, 1)$ [or $(x_0, y_0) = (27, 26\frac{2}{3})$] in Heron's example). Archimedes appears to have used a variant of this. One nice thing about this procedure is that the pairs of numbers obtained at every step lie on either side of the true answer (because $xy = z$ at every step), so one has explicit control over the accuracy. Even before starting the procedure Heron's initial guess bracketed the answer between $26\frac{2}{3}$ and 27.

**Exercise 1.3.1** To approximate $\sqrt{4}$, calculate the numbers $(x_i, y_i)$ for $0 \le i \le 4$ using this method, starting with $(1, 4)$, and give their distance to 2.

**Exercise 1.3.2** Carry Heron's approximation of $\sqrt{720}$ one step further and use a calculator to determine the accuracy of that approximation.

**Exercise 1.3.3** Starting with initial guess 1, how many steps of this procedure are needed get a better approximation of $\sqrt{720}$ than Heron's initial guess of 27?

What happens after a few steps of this procedure is that the numbers $x_n$ and $y_n$ that one obtains are almost equal and therefore close to $\sqrt{z}$. With Heron's intelligent initial guess his first approximation was good enough ($26\frac{5}{6}$ is within .002% of $\sqrt{720}$), and he never seems to have carried out the repeated approximations he proposed. It is a remarkable method not only because it works, but because it works so quickly. But why does it work? And why does it work so quickly? And exactly how quickly does it work? These are questions we can answer with ease after our start in dynamical systems (Section 2.2.8).

### 1.3.2 The Search for Roots

Many problems asking for a specific numerical solution can be easily and profitably rephrased as looking for a solution of $f(x) = 0$ for some appropriate function $f$. We describe two well-known methods for addressing this question for functions of one variable.

**1. Binary Search.** There is a situation where we can be sure that a solution exists: The Intermediate-Value Theorem from calculus tells us that if $f\colon [a, b] \to \mathbb{R}$ is continuous and $f(a) < 0 < f(b)$ [or $f(b) < 0 < f(a)$, so we could say $f(a)f(b) < 0$], then there is some $c \in (a, b)$ such that $f(c) = 0$.

■ **Exercise 1.3.4** Show that this statement of the Intermediate-Value Theorem is equivalent to the standard formulation.

Knowing that a solution exists is, however, not quite the same as knowing the solution or at least having a fairly good idea where it is. Here is a simple reliable method for getting to a root.

Given that $f(a) < 0 < f(b)$, consider the midpoint $z = (a + b)/2$.

CASE 0: If $f(z) = 0$, we have found the root. Otherwise, there are two cases.

CASE 1: If $f(z) > 0$, replace the interval $[a, b]$ by the interval $[a, z]$, which is half as long and contains a root by the Intermediate-Value Theorem because $f(a) < 0 < f(z)$. Repeat the procedure on this interval.

CASE 2: If $f(z) < 0$, replace $[a, b]$ by $[z, b]$, which is also half as long, and apply the procedure here.

This binary search produces a sequence of nested intervals, cutting the length in half at every step. Each interval contains a root, so we obtain ever-better approximations and the limit of the right (or left) endpoints is a solution.

Note that this procedure is iterative, but it does not define a dynamical system. Not one that operates on numbers anyway. One could view it as a dynamical system operating on intervals on whose endpoints $f$ does not have the same sign.

■ **Exercise 1.3.5** Carry out three steps of this procedure for $f(x) = x - \cos x$ on $[0, 1]$. Conclude with an approximate solution and its accuracy.

This method is reliable: It gives ever-better approximations to the solution at a guaranteed rate, and this rate is respectable and the error can be calculated. For example, nine steps give an error less than $(b - a)/1000$.

**2. The Newton Method.** The Newton Method (or Newton–Raphson Method) was devised as a solution of the same problem of finding zeros of functions. It is more flamboyant than the binary search: It is ingenious and can work rapidly, but it is not always reliable.

For this method we need to assume that the function $f$, whose zero we are to find, is differentiable, and, of course, that there is a zero someplace. One begins by making an educated guess $x_0$ at the solution. How to make this guess is up to the user and depends on the problem. A reasonable graph might help, or maybe the situation is such that the binary search can be applied. In the latter case a few steps give an excellent initial guess.

Newton's method endeavors to give you a vastly improved guess. If the function is linear, then your initial guess combined with the slope of the graph immediately gives the exact solution. Being differentiable, the function $f$ is well approximated by tangent lines. Therefore the initial guess $x_0$ and the equation of the tangent line to the graph of $f$ tell us the $x$-intercept of the tangent line. This is the improved guess. In terms of formulas the calculation amounts to

$$x_1 = F(x_0) := x_0 - \frac{f(x_0)}{f'(x_0)}.$$

**Exercise 1.3.6** Verify that this formula encodes the geometric description above.

**Exercise 1.3.7** Apply this method four times to $x^2 - 4 = 0$ with initial guess 1. Compare with Exercise 1.3.1. (Look also at Exercise 1.3.18 and Exercise 1.3.19.)

This simple procedure can be applied repeatedly by iterating $F$. It gives a sequence of hopefully ever-better guesses. In Section 2.2.8 we give a simple criterion to ensure that the method will succeed.

**Exercise 1.3.8** Several of the exercises in Section 1.1 are examples of Newton's method at work. Find the ones that are and give the equation whose solution they find.

Since this method defines a dynamical system, it has been studied as such. This is in large part because some initial choices provide situations where the asymptotic behavior is complicated. Especially when this is done with complex numbers, one can produce beautiful pictures by numerical calculations. An important development was an adaptation of this method to work on points in function spaces usually called the Kolmogorov–Arnol'd–Moser or KAM method, which provided a tool for one of the furthest advances in studying whether our solar system is stable. This is an outstanding example where knowledge about simple asymptotics of a dynamical system in an auxiliary space gives insight into another dynamical system.

### 1.3.3 Closed Geodesics

If an airplane pilot were to tie down the wheel[16] and had a lot of fuel, the plane would go around the earth all the way along a great circle, returning precisely to the starting point, and repeat. One could try the same with a vehicle on the surface, but some more attention would be required because of intervening mountains, oceans, rainforests, and such. The idealized model of this kind of activity is that of a particle moving freely on the surface of a sphere. Because there are no external forces (and no friction, we assume), such a particle moves at constant speed with no change of direction. It is quite clear that the particle always returns to the starting point periodically. So there are infinitely many ways of traveling (freely) in a periodic fashion.

What if your sphere is not as round and shiny as the perfect round sphere? It may be slightly dented, or maybe even badly deformed. One could adorn it with a mushroom-like appendage or even make it look like a barbell. Only, no tearing or glueing of the surface is allowed. And no crimping. A smooth but not ball-shaped "sphere." Now a freely moving particle has no obvious reason to automatically return home. Almost any way of deforming the sphere produces many nonperiodic motions. Here is a hard question: Are there still infinitely many ways, on a given deformed sphere, of moving freely and periodically?

[16] This means that the plane flies horizontally and straight, and the proper technical term would be "yoke" instead of "wheel".

One beautiful aspect of free particle motion is that the path of motion is always the shortest connection between any two points on it that are not too far apart. (Obviously, a closed path is not the shortest curve from a point to itself.) This is familiar for the round sphere when these paths are great circles, but it is universally true, and such paths are called *geodesics*. Therefore, the above question can also be asked in terms of geodesics: On any sphere, no matter how deformed, are there always infinitely many closed geodesics?

This is a question from geometry, and it was posed long ago. It was solved (not so long ago) by dynamicists using the theory of dynamical systems. We explain how geodesics are related to dynamics in Section 6.2.8 and outline an approach to this question in Section 14.5.

### 1.3.4 First Digits of the Powers of 2

As an illustration of the power of dynamics to discern patterns even of a subtle and intricate nature, consider the innocuous sequence of powers of 2. Here are the first 50 terms of this sequence:

| | | | | |
|---|---|---|---|---|
| 2 | 2048 | 2097152 | 2147483648 | 2199023255552 |
| 4 | 4096 | 4194304 | 4294967296 | 4398046511104 |
| 8 | 8192 | 8388608 | 8589934592 | 8796093022208 |
| 16 | 16384 | 16777216 | 17179869184 | 17592186044416 |
| 32 | 32768 | 33554432 | 34359738368 | 35184372088832 |
| 64 | 65536 | 67108864 | 68719476736 | 70368744177664 |
| 128 | 131072 | 134217728 | 137438953472 | 140737488355328 |
| 256 | 262144 | 268435456 | 274877906944 | 281474976710656 |
| 512 | 524288 | 536870912 | 549755813888 | 562949953421312 |
| 1024 | 1048576 | 1073741824 | 1099511627776 | 1125899906842624. |

This list looks rather complicated beyond the trivial pattern that these numbers grow. There are some interesting features to be observed, however. For example, the last digits repeat periodically: 2, 4, 8, 6. That this must be so is quite obvious: The last digit of the next power is determined by the last digit of the previous one; so once a single repetition appears, it is bound to reproduce the pattern. (Furthermore, the last digit is always even and never 0.)

A similar argument shows that the last two digits jointly must also eventually start repeating periodically: By the previous observation there are at most 40 possibilities for the last two digits, and since the last two digits of the next power are determined by those of the previous one, it is sufficient to have one repetition to establish a periodic pattern. Looking at our sequence we see that, indeed, the last two digits form the following periodic sequence with period 20 beginning from the second term: 04 08 16 32 64 28 56 12 24 48 96 92 84 68 36 72 44 88 76 52.

Note that this sequence has a few interesting patterns. Adding its first and eleventh terms gives 100, as does adding the second and twelfth, the third and thirteenth, and so on. One way of developing this sequence is to start from 04 and apply the following rule repeatedly: If the current number is under 50, double it; otherwise, double the difference to 100. The simpler 2,4,8,6 above exhibits analogous patterns.

Now look at the sequence of the *first* digits. Reading off the same list:

| | | | | |
|---|---|---|---|---|
| 2 | 2048 | 2097152 | 2147483648 | 2199023255552 |
| 4 | 4096 | 4194304 | 4294967296 | 4398046511104 |
| 8 | 8192 | 8388608 | 8589934592 | 8796093022208 |
| 16 | 16384 | 16777216 | 17179869184 | 17592186044416 |
| 32 | 32768 | 33554432 | 34359738368 | 35184372088832 |
| 64 | 65536 | 67108864 | 68719476736 | 70368744177664 |
| 128 | 131072 | 134217728 | 137438953472 | 140737488355328 |
| 256 | 262144 | 268435456 | 274877906944 | 281474976710656 |
| 512 | 524288 | 536870912 | 549755813888 | 562949953421312 |
| 1024 | 1048576 | 1073741824 | 1099511627776 | 1125899906842624 |

one finds the first digits of the 50 entries to be

$$\begin{array}{c}2481361251\\2481361251\\2481361251\\2481361251\\2481371251.\end{array}$$

This is tantalizingly close to being periodic, but a small change creeps in at the end, so no truly periodic pattern appears – and there is no reason to expect any. (If you calculate further entries in this sequence, this behavior continues; little changes keep appearing here and there.)

Since this sequence is not as regular as the previous one, a statistical approach might be helpful. Look at the *frequency* of each digit – how often does a particular digit figure in this list? We have:

| digit : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| number of times : | 15 | 10 | 5 | 5 | 5 | 4 | 1 | 5 | 0. |

These frequencies look somewhat uneven. In particular, seven and nine seem to be disfavored. Seven appears for the first and only time at the 46th place in our sequence, and nine appears for the first time as the first digit of $2^{53}$. Calculation of the first 100 entries gives slightly less lopsided frequencies, but they seem to be smaller for larger digits.

Thus, all nine digits appear as the first digit of some power of 2. We would like to know more, however. Does every digit appear infinitely many times? If yes, do they appear with any regularity? Which of the digits appear most often?

In order to discuss them we need to formulate these questions precisely. To that end we count for each digit $d$ and every natural number $n$ the number $F_d(n)$ of those powers $2^m$, $m = 1, \dots, n$ that begin with $d$. Thus, we just listed the 10 values of $F_d(50)$. The frequency with which $d$ appears as the first digit among the first $n$ powers of 2 is $F_d(n)/n$. Thus, one of our questions is whether each of these quantities has a limit as $n$ goes to infinity and how these limits, if they exist, depend on $d$. Once these questions have been answered, one can also ask them about powers of 3 and compare the limit frequencies.

In Proposition 4.2.7 we obtain existence of these limits and give a formula for them which in particular implies that all limit frequencies are positive and that they decrease as $d$ increases. Thus, contrary to the evidence from the first 50 powers (but in agreement with what one sees among the first 100), seven eventually appears more often than eight. The relationship between these limits for powers of 3 versus powers of 2 is also striking.

### 1.3.5 Last Digits of Polynomials

In the previous example we had immediate success in studying patterns of last digits and noted that some dose of dynamics provides the tools for understanding the behavior of the first digits. Let us look at another problem of integer sequences where similar questions can be asked about last digits.

Instead of an exponential sequence consider the sequence $x_n = n^2$ for $n \in \mathbb{N}_0$. The last digits come out to be 01496569410 and repeat periodically thereafter.

**Exercise 1.3.9** Prove that these digits repeat periodically.

**Exercise 1.3.10** Explain why this sequence is *palindromic*, that is, unchanged when reversed.

This is about as simple as it was earlier, so let's try $x_n = n^2 p/q$ instead, for some $p, q \in \mathbb{N}$. Unless $q = 1$, these won't all be integers, so we make explicit that we are looking at the digit before the decimal point. You may want to experiment a little, but it is easy to see directly that we still get a periodic pattern, with period at most $10q$. The reason is that

$$a_{n+10q} - a_n = (n + 10q)^2\, p/q - n^2\, p/q = 10(2np + 10\,pq)$$

is an integer multiple of 10, so the digit before the decimal point (as well as all the ones after) is the same for $a_{n+10q}$ and $a_n$.

**Exercise 1.3.11** Prove that the initial $10q$ results form a palindromic string.

This was interesting, but not subtle. It is natural to replace $p/q$ by an irrational number, because that should cause an "infinite period," as it were, that is, no periodicity at all.

So, consider $x_n = n^2\sqrt{2}$. The sequence of last digits (before the decimal point) begins with the following 100 terms: 47764935641602207257751690074812184811 07379985035540580084923206134316133205911072577527011950343171.

There are no obvious reasons for periodicity, nor is any such pattern apparent. Certainly all digits make an appearance. However, the questions we asked about first digits of powers of 2 are also appropriate here: Do all digits appear infinitely often? Do they appear with well-defined relative frequencies? Relative frequencies are defined as before: Let $P_n(d)$ be the number of times the last digit is $d$ in the set $\{i^2\sqrt{2}\}_{i=0}^{n-1}$ and consider $P_n(d)/n$ for large $n$. Among the first 100 values we get the frequencies

| $i$: | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| $P_{100}(i)/100$: | 0.14 | 0.15 | 0.09 | 0.10 | 0.09 | 0.11 | 0.06 | 0.13 | 0.06 | 0.07. |

This list does not suggest any answer to this question, and the same list for larger $n$ might not either.

Dynamics is able to address these questions as well as many similar ones completely and rigorously. In this particular example it turns out that all relative frequencies converge to $1/10$. Thus we have an example of *uniform distribution*, which is one of the central paradigms in dynamics as well as in nature. We outline a solution of the problem of distribution of last digits in Section 15.1.

### 1.3.6 Cellular Automata

A game of sorts called the game of life was popular in the 1980s. It is intended to model a simple population of somethings that live in fixed locations. Each of the "organisms" is at a point of a fixed lattice, the points in the plane with integer coordinates, say, and can have several states of health. In the simplest version such organisms might have only the two states "present" and "not there" (or 1 and 0). But one may also take a model with a larger number of possible states, including, for example, "sickly" or "happy." The rule of the game is that the population changes in discrete time steps in a particular way. Each organism checks the states of some of its neighbors (up to some distance) and, depending on all these, changes its own state accordingly. For example, the rule might say that if all immediate neighbors are present, the organism dies (overpopulation). Maybe the same happens if there are no neighbors at all (too lonely or exposed). This game was popular because from relatively simple rules one could find (or design) intriguing patterns, and because computers, even early ones, could easily go through many generations in a short time.

If the number of cells is finite, then from our perspective of asymptotic long-term behavior there is not too much to say about the system. It has only finitely many states, so at some point some state must be attained for a second time. Because the rules are unchanged, the pattern thereafter cycles again through the same sequence of states since the last time, and again and again. No matter how interesting the patterns may be that emerge, or how long the cycle, this is a complete qualitative description of the long-term behavior.

When there are infinitely many cells, however, there is no reason for this kind of cycling through the same patterns, and there may be all kinds of long-term behaviors.

Systems of this kind are called cellular automata. Since the rules are so clearly described, one can easily make mathematics out of them. To keep the notation simple we look not at the integer points in the plane, but only those on the line. Accordingly, a state of the system is a sequence, each entry of which has one of finitely many values (states). If the states are numbered $0, \dots, N-1$, then we can denote the space of these sequences by $\Omega_N$. All organisms have the same rule for their development. It is given by a function $f\colon \{0, \dots, N-1\}^{2n+1} \to \{0, \dots, N-1\}$, that is, a function that maps $2n+1$-characters-long strings of states $(0, \dots, N-1)$ to a state. The input consists of the states of all neighbors up to distance $n$ in either direction, and the output is the future state of the individual. Therefore, each step of the evolution of the whole system is given by a map $\Phi\colon \Omega_N \to \Omega_N$ such that $(\Phi(\omega))_i = f(\omega_{i-n}, \dots, \omega_{i+n})$. By way of example, take $N = n = 1$ and $f(x_{-1}, x_0, x_1) = x_1$. This means that every individual just chooses to follow its right neighbor's lead (today's $x_1$ is tomorrow's $x_0$). You might call this example "the wave," because whatever pattern you begin with, it will relentlessly march leftward.

This is a general description of cellular automata, whose interest goes well beyond the game of life. The same mathematical concept admits a rather different interpretation. If one thinks of each of these sequences as a stream of data, then the map $\Phi$ transforms these data – it is a code. This particular class of codes is known as *sliding block codes,* and this kind is suitable for real-time streaming data encoding or decoding. For us, it is a transformation on a nice space that can be repeated, a dynamical system. The general class of dynamical systems whose states are given by sequences (or arrays) is called *symbolic dynamics,* and some of our most useful models are of this kind. "The wave" is actually our favorite, and we call it the (left) shift. As a class, sliding block codes play an important role, although under a different name (conjugacies).

Symbolic dynamics is introduced in Section 7.3.4 and studied in Section 7.3.7. It provides a rich supply of examples that are simple to describe but produce a variety of complicated dynamical phenomena.

## EXERCISES

**Exercise 1.3.12** Prove that in the binary search for a root the sequences of left and right endpoints both converge and that they have the same limit.

**Exercise 1.3.13** In the binary search for a root assume $a = 0$, $b = 1$ and that the procedure never terminates. Keep track of the choices at each step by noting a 0 whenever Case 1 occurs and noting a 1 whenever Case 2 occurs. Prove that the string of 0's and 1's thus obtained gives the binary representation of the solution found by the algorithm.

**Exercise 1.3.14** In the preceding exercise assume that the search terminates. How does the finite string of 0's and 1's relate to the binary representation of the root?

**Exercise 1.3.15** Solve $\cos x = x$ with the Newton Method and the initial guess $x_0 = 3/4$.

**Exercise 1.3.16** Approximate $\sqrt{5}$ to the best possible accuracy of your calculator by the Newton Method with initial guess 2.

**Exercise 1.3.17** Use the Newton Method to solve $\sin x = 0$ with initial guess 1 and note the pattern in the size of the absolute error.

**Exercise 1.3.18** Try to solve $\sqrt[3]{x} = 0$ with the Newton Method, *not* taking 0 as initial guess.

**Exercise 1.3.19** For the Greek method of arithmetic/harmonic mean, express the successive arithmetic means as the iterates of some function, that is, write down a recursive formula for the first components alone.

**Exercise 1.3.20** Finding the root of a number $z$ can be done in various ways. Compare the Greek method of arithmetic/harmonic mean with the Newton Method, taking 1 as the initial guess.