1

Sequences and Limits

1.1 Calculus: Areas and tangents

The study of calculus begins with questions about change. What happens to the velocity of a swinging pendulum as its position changes? What happens to the position of a planet as time changes? What happens to a population of owls as its rate of reproduction changes? Mathematically, one is interested in learning to what extent changes in one quantity affect the value of another related quantity. Through the study of the way in which quantities change we are able to understand more deeply the relationships between the quantities themselves. For example, changing the angle of elevation of a projectile affects the distance it will travel; by considering the effect of a change in angle on distance, we are able to determine, for example, the angle which will maximize the distance.

Related to questions of change are problems of approximation. If we desire to approximate a quantity which cannot be computed directly (for example, the area of some planar region), we may develop a technique for approximating its value. The accuracy of our technique will depend on how many computations we are willing to make; calculus may then be used to answer questions about the relationship between the accuracy of the approximation and the number of calculations used. If we double the number of computations, how much do we gain in accuracy? As we increase the number of computations, do the approximations approach some limiting value? And if so, can we use our approximating method to arrive at an exact answer? Note that once again we are asking questions about the effects of change.

Two fundamental concepts for studying change are sequences and limits of sequences. For our purposes, a *sequence* is nothing more than a list of numbers. For example,

$$1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$$

might represent the beginning of a sequence, where the ellipsis indicates that the list is to continue on indefinitely in some pattern. For example, the 5th term in this sequence might be

$$\frac{1}{16} = \frac{1}{2^4},$$

the 8th term

$$\frac{1}{128} = \frac{1}{2^7},$$

and, in general, the nth term

$$\frac{1}{2^{n-1}},$$

where $n = 1, 2, 3, \ldots$. Notice that the sequence is completely specified only when we have given the general form of a term in the sequence. Also note that this list of numbers is approaching 0, which we would call the *limit* of the sequence. In the next section of this chapter we will consider in some detail the basic question of determining the limit of a sequence.

The following two examples consider these ideas in the context of the two fundamental problems of calculus. The first of these is to determine the area of a region in the plane; the other is to find the line tangent to a curve at a given point on the curve. As the course progresses, we will find that general methods for solving these two problems are at the heart of the techniques used in calculus. Moreover, we will see that these two problems are, surprisingly, closely related, with the area problem actually being the inverse of the tangent problem. This intimate connection was one of the great discoveries of Isaac Newton (1642–1727) and Gottfried Leibniz (1646–1716), although anticipated by Newton's teacher Isaac Barrow (1630–1677).

Example 1.1.1. Suppose we wish to find the area inside a circle of radius one centered at the origin. Of course, we have all learned that the answer is π. But why? Indeed, what does it mean to find the area of a disk?

Area is best defined for polygons, regions in the plane with line segments for sides. One can start by defining the area of a 1×1 square to be one unit. The area of any other polygonal figure is then determined by how many squares may be fit into it, with suitable cutting as necessary. For example, it is seen that the area of a rectangle with base of length b and height a should be ab. Since a parallelogram with base of length b and height a may be cut and pasted onto a rectangle of length b and height a (see Problem 1), it follows that the area of such a parallelogram is also ab. As a triangle with height a and a base of length b is one-half of a parallelogram of height a and base length b (see Problem 2), it easily follows that the area of such a triangle is $\frac{1}{2}ab$. The area of any other polygon can be calculated, at least in theory, by decomposing it into a suitable number of triangles. However, a circle does not have straight sides and so may not be handled so easily. Hence we resort to approximations.

Let P_n be a regular n-sided polygon inscribed in the unit circle centered at the origin and let A_n be the area of P_n. For example, Figure 1.1.1 shows P_8 inscribed in the unit circle. We may decompose P_n into n congruent isosceles triangles by drawing line segments from the center of the circle to the vertices of the polygon, as shown in Figure 1.1.2 for P_8. For each of these triangles, the angle with vertex at the center of the circle has measure $\frac{360}{n}$ degrees, or $\frac{2\pi}{n}$ radians, where π represents the ratio of the circumference of a circle to its diameter. Hence, since the equal sides of each of the triangles are of length one, each triangle has a height of

$$h_n = \cos\left(\frac{\pi}{n}\right)$$

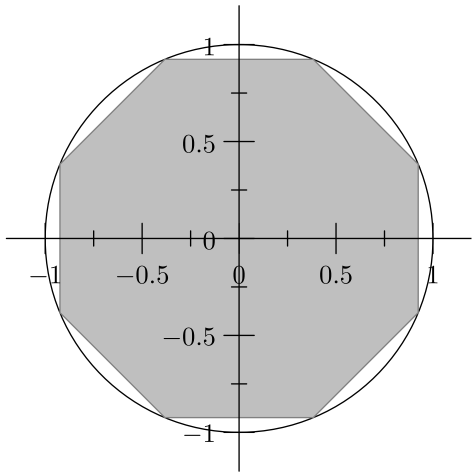

Figure 1.1.1. Regular octagon P_8 inscribed in a circle

and a base of length

$$b_n = 2\sin\left(\frac{\pi}{n}\right)$$

(see Problem 3). Thus the area of a single triangle is given by

$$\frac{1}{2}b_n h_n = \cos\left(\frac{\pi}{n}\right)\sin\left(\frac{\pi}{n}\right) = \frac{1}{2}\sin\left(\frac{2\pi}{n}\right),$$

where we have used the fact that

$$\sin(2\alpha) = 2\sin(\alpha)\cos(\alpha)$$

for any angle α. Multiplying by n, we see that the area of P_n is

$$A_n = \frac{n}{2}\sin\left(\frac{2\pi}{n}\right).$$

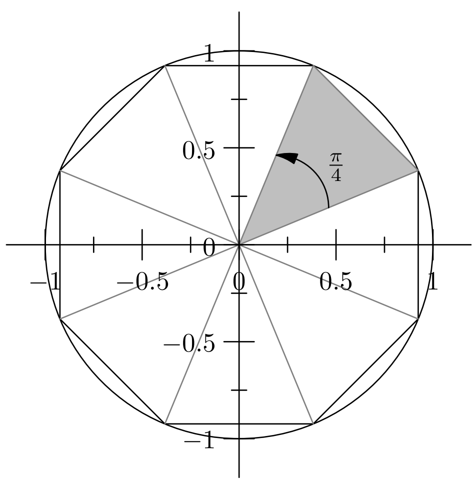

Figure 1.1.2. Decomposition of P_8 into triangles

We now have a sequence of numbers, A_1, A_2, A_3, …, each number in the sequence being an approximation to the area of the circle. Moreover, although not entirely obvious, each term in the sequence is a better approximation than its predecessor since the corresponding regular polygon more closely approximates the circle. For example, to five decimal places we have

$$\begin{aligned} A_3 &= 1.29904, \\ A_4 &= 2.00000, \\ A_5 &= 2.37764, \\ A_6 &= 2.59808, \\ A_7 &= 2.73641, \\ A_8 &= 2.82843, \\ A_9 &= 2.89254, \\ A_{10} &= 2.93893, \\ A_{11} &= 2.97352, \end{aligned}$$

and

$$A_{12} = 3.00000.$$

Continuing in this manner, we find $A_{20} = 3.09017$, $A_{50} = 3.13333$, and $A_{100} = 3.13953$. As we would expect, the sequence is increasing and appears to be approaching π. Indeed, if we take a polygon with 1644 sides, we have $A_{1644} = 3.14159$, which is π to five decimal places.

Alternatively, instead of defining π to be the ratio of the circumference of a circle to its diameter, we could define it to be the area of a circle of radius one. That is, we could define π to be the limiting value of the sequence A_n. Symbolically, we express this by writing

$$\pi = \lim_{n\to\infty} A_n.$$

In that case, let B be the area of a circle of radius r and let B_n be the area of a regular n-sided polygon Q_n inscribed in the circle. If we decompose Q_n into n isosceles triangles in the same manner as P_n above, then each triangle in this decomposition is similar to any one of the triangles in the decomposition of P_n. Since the ratios of the lengths of corresponding sides of similar triangles must all be the same, the sides of a triangle in the decomposition of Q_n must be r times the length of the corresponding sides of any triangle in the decomposition of P_n. Hence each of the triangles in the decomposition of Q_n must have a base of length rb_n and a height of rh_n, where h_n is the height and b_n is the length of the base of one of the isosceles triangles in the decomposition of P_n. Thus the area of one of the triangles in the decomposition of Q_n into isosceles triangles will be

$$\frac{1}{2}(rb_n)(rh_n) = \frac{1}{2}r^2 b_n h_n,$$

from which it follows that

$$B_n = \frac{n}{2}r^2 b_n h_n = r^2\left(\frac{n}{2}b_n h_n\right) = r^2 A_n.$$

Since r is a fixed constant, we would then expect that, in the limit as the number of sides grows toward infinity,

$$B = \lim_{n\to\infty} B_n = \lim_{n\to\infty} r^2 A_n = r^2 \lim_{n\to\infty} A_n = \pi r^2.$$

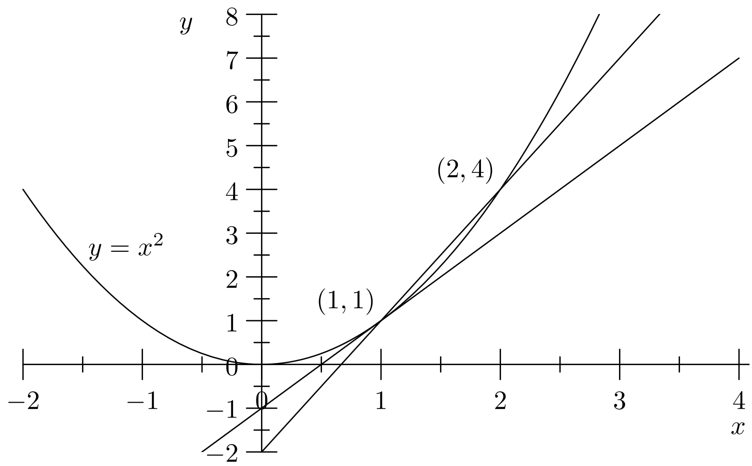

Figure 1.1.3. Parabola with tangent and secant lines

Hence we arrive at the famous formula for the area of a circle of radius r, in which the constant π has been defined to be the area of a circle of radius one.

Example 1.1.2. In this example we wish to find the line tangent to the curve $y = x^2$, a parabola, at the point $(1, 1)$. This problem may not at first seem as useful as that of finding the area of a planar region, but we shall find that the ideas behind the solution have many applications, and are, ultimately, important in the solution of the area problem as well.

First there is the question of exactly what is a tangent line. At the present it will be sufficient to leave the notion at an intuitive level: a tangent line is a line which just touches a given curve at a point, giving a close approximation between curve and line. In Chapter 3, we will see that a line ℓ is tangent to a curve C at a point P on C if ℓ passes through P and, in a sense that we will make precise at that time, gives a better approximation to C for points close to P than any other line.

Now let C be the curve with equation $y = x^2$, let $P = (1, 1)$, and let ℓ be the line tangent to C at P. Since ℓ passes through P, in order to find the equation of ℓ we need only find its slope m. Unfortunately, to find m in the standard way we need to know two points on ℓ, and we know only one, namely P. Hence we will again have to resort to approximations. For example, the line through the points $(1, 1)$ and $(2, 4)$ is not ℓ (it is a secant line, rather than a tangent line), but since it intersects C at P and at another point which is close to P, its slope should approximate m (see Figure 1.1.3). Namely, using "≈" to denote an approximation, we have

$$m \approx \frac{4-1}{2-1} = 3.$$

Since $\left(\frac{3}{2}, \frac{9}{4}\right)$ is on C and is closer to P than $(2, 4)$, a better approximation is given by the slope of the line passing through $(1, 1)$ and $\left(\frac{3}{2}, \frac{9}{4}\right)$, that is,

$$m \approx \frac{\frac{9}{4}-1}{\frac{3}{2}-1} = \frac{\frac{5}{4}}{\frac{1}{2}} = \frac{5}{2}.$$

More generally, let n be a positive integer and let m_n be the slope of the line through the points

$$\left(1+\frac{1}{n},\left(1+\frac{1}{n}\right)^2\right)$$

and P. For example, we have just seen that $m_1 = 3$ and $m_2 = \frac{5}{2}$. Now, in general,

$$\begin{aligned} m_n &= \frac{\left(1+\frac{1}{n}\right)^2-1}{\left(1+\frac{1}{n}\right)-1} \\ &= \frac{1+\frac{2}{n}+\frac{1}{n^2}-1}{\frac{1}{n}} \\ &= n\left(\frac{2}{n}+\frac{1}{n^2}\right) \\ &= 2+\frac{1}{n} \end{aligned}$$

for $n = 1, 2, 3, \ldots$. Hence

$$\begin{aligned} m_3 &= 2+\frac{1}{3} = \frac{7}{3}, \\ m_4 &= 2+\frac{1}{4} = \frac{9}{4}, \\ m_5 &= 2+\frac{1}{5} = \frac{11}{5}, \end{aligned}$$

and so on. Moreover, as n increases, $\frac{1}{n}$ decreases toward 0, and so we would expect that, as n increases, m_n decreases toward 2. At the same time, as n increases m_n more closely approximates m. Thus we should have

$$m = \lim_{n\to\infty} m_n = \lim_{n\to\infty}\left(2+\frac{1}{n}\right) = 2.$$

That is, the slope of the line tangent to C at P is 2. Then the tangent line ℓ has equation

$$y-1 = 2(x-1),$$

or

$$y = 2x-1.$$

Here we have used the fact that the equation of a line with slope m and passing through the point (a, b) is given by

$$y-b = m(x-a).$$

The rest of this chapter will be concerned with the study of sequences and their limits. The next section will consider the basic definitions and computational techniques, while the remaining sections will discuss some applications. We will return to the problem of finding tangent lines in Chapter 3 and the problem of computing areas in Chapter 4.

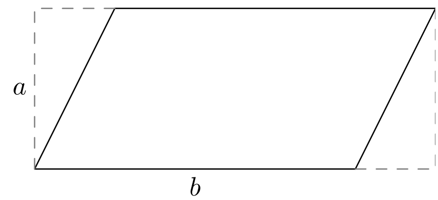

Figure 1.1.4. Parallelogram

Problems 1.1

(1) Use Figure 1.1.4 to verify that a parallelogram with height a and base of length b has area ab.

(2) Explain how any triangle is one-half of a parallelogram, and use this to verify the formula for the area of a triangle.

(3) Use Figure 1.1.5 to verify the formulas given for the height and base of one of the isosceles triangles in the decomposition of P_n.

(4) Try the procedure of the tangent-line example to find the equation of the line tangent to the following curves at the indicated point.

(a) $y = 2x^2$ at $(1, 2)$ (b) $y = x^2 + 1$ at $(1, 2)$

(c) $y = x^3$ at $(1, 1)$ (d) $y = x^2$ at $(2, 4)$

(5) For the area example, find the number of sides necessary for the area of the inscribed polygon to approximate π to 6, 7, 8, 9, and 10 digits after the decimal point.

(6) For the tangent-line example, how large would n have to be in order for $|m_n - 2|$ to be less than 0.005?

(7) For the tangent-line example, let p be the smallest positive integer such that $|m_p - 2| < 0.01$.

(a) What is p?

(b) What can you say about $|m_n - 2|$ for values of n greater than p?

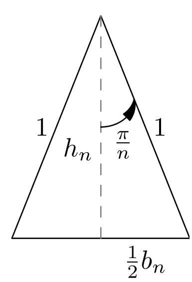

Figure 1.1.5. Isosceles triangle

(8) For each of the following sequences $\{a_n\}$, compute a_{10}, a_{20}, a_{100}, a_{500}, and a_{1000}.

(a) $a_n = n\sin\left(\frac{1}{n}\right)$

(b) $a_n = \left(1+\frac{1}{n}\right)^n$

(c) $a_n = \frac{10^n}{n!}$, where $n! = n(n-1)(n-2)\cdots(2)(1)$

(9) As we saw in the area example, there is more than one way to define the number π. For example, we can define it either as the area of a circle of unit radius or as the ratio of the circumference of a circle to its diameter (of course, if the latter approach is taken, one has to show that this ratio is the same for every circle). Suppose we define π as the area of a circle of unit radius. Consider a circle with radius r, diameter d, circumference C, and area A. Then we have seen that $A = \pi r^2$. The following steps show that we also have $\pi = \frac{C}{d}$.

(a) Let P_n be a regular n-sided polygon inscribed in the circle. Let s be the length of a side of P_n. By dividing P_n into n equal isosceles triangles as we did in the area example, argue that

$$A \approx \frac{nrs}{2}.$$

(b) Can you see why as n goes to infinity, ns approaches C?

(c) Now can you see why

$$A = \lim_{n\to\infty} \frac{nrs}{2} = \frac{rC}{2}?$$

(d) Use the result in part (c) to show that

$$\pi = \frac{C}{d}.$$

(10) You may find an interesting discussion of techniques for computing areas and volumes up to the time of Archimedes (287–212 B.C.) in the first two chapters of *The Historical Development of Calculus* by C. H. Edwards (Springer-Verlag New York Inc., 1979). In particular, there is a discussion on pages 31–35 of Archimedes' proof that the two definitions of π mentioned in the area example yield the same number.

1.2 Sequences

Recall that a sequence is a list of numbers, such as

$$1, 2, 3, 4, \ldots,$$
$$2, 4, 6, 8, \ldots,$$
$$0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots,$$
$$1, -\frac{1}{2}, \frac{1}{4}, -\frac{1}{8}, \ldots,$$

or

$$1, -1, 1, -1, \ldots.$$

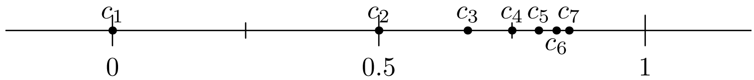

Figure 1.2.1. First seven values of $c_n = 1 - \frac{1}{n}$

As we noted in Section 1.1, listing the first few terms of a sequence does not uniquely specify the remaining terms of the sequence. To fully specify a sequence, we need a formula that describes an arbitrary term in the sequence. For example, the first example above lists the first four terms of the sequence $\{a_n\}$ with

$$a_n = n$$

for $n = 1, 2, 3, \ldots$; the second example lists the first four terms of $\{b_n\}$ with

$$b_n = 2n$$

for $n = 1, 2, 3, \ldots$; the third example lists the first four terms of $\{c_n\}$ with

$$c_n = 1 - \frac{1}{n}$$

for $n = 1, 2, 3, \ldots$; the fourth lists the first four terms of $\{d_n\}$ with

$$d_n = \frac{(-1)^n}{2^n}$$

for $n = 0, 1, 2, 3, \ldots$; and the fifth lists the first four terms of $\{e_n\}$ with

$$e_n = (-1)^n$$

for $n = 0, 1, 2, \ldots$. Note, however, that although these are in some sense the natural formulas for these sequences, they are not the only possibilities.

As indicated in Section 1.1, we are often interested in the value, if one exists, which a sequence approaches. For example, the sequences $\{a_n\}$ and $\{b_n\}$ increase beyond any possible bound as n increases, and hence they have no limiting value. To visualize what is happening here, you might plot the points of the sequence on the real line. For both of these sequences, the plotted points will march off to the right without any upper limit. Although a limit does not exist in these cases, we usually write

$$\lim_{n\to\infty} a_n = \infty$$

and

$$\lim_{n\to\infty} b_n = \infty$$

to express the fact that the limits do not exist because the terms in the sequence are eventually always larger than any specified positive bound. On the other hand, if we plot the points of the sequence $\{c_n\}$, as in Figure 1.2.1, we see that although they are always increasing (that is, moving toward the right), nevertheless they never increase beyond 1. Moreover, even though no term in the sequence is ever equal to 1, we can see that the points become arbitrarily close to 1. Hence we say that the limit of the sequence is 1 and we write

$$\lim_{n\to\infty} c_n = 1.$$

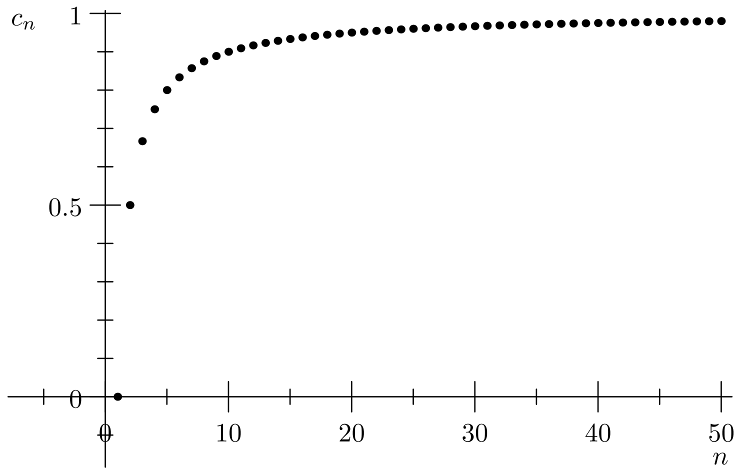

Figure 1.2.2. Plot of $\left(n, 1 - \frac{1}{n}\right)$, $n = 1, 2, 3, \ldots, 50$,

Even though they oscillate between positive and negative values, the terms in the sequence $\{d_n\}$ approach closer and closer to 0 as n increases. Since it is possible to make d_n as close as we like to 0 for all suitably large n, we may write

$$\lim_{n\to\infty} d_n = 0.$$

Finally, for the sequence $\{e_n\}$ there are only two points to plot, alternating between 1 and -1. Since the terms of this sequence oscillate between two numbers, and so do not approach any fixed limiting value, we say that the sequence does not have a limit.

Another approach to visualizing the limiting behavior of a sequence $\{a_n\}$ is to plot the ordered pairs (n, a_n) in the plane for some range of values of n. For example, Figure 1.2.2 shows a plot of the points (n, c_n), $n = 1, 2, 3, \ldots, 50$ for the sequence $\{c_n\}$ given above. Note how the points approach the horizontal line $y = 1$, indicating, as mentioned above, that

$$\lim_{n\to\infty} c_n = 1.$$

Similarly, Figure 1.2.3 shows a plot of the points (n, d_n), $n = 0, 1, 2, \ldots, 10$; here the points approach the horizontal axis, $y = 0$, consistent with our claim that

$$\lim_{n\to\infty} d_n = 0.$$

Figure 1.2.4 shows a plot of (n, e_n), $n = 0, 1, 2, \ldots, 20$. The fact that this sequence does not have a limit is manifest in seeing the vertical coordinate of the points oscillate between 1 and -1.

As the concept of a limit is fundamental to the understanding of calculus, it is important that we make the notion more concrete than we have so far. That is, we need to have a formal definition of limit which exactly captures what we have been discussing intuitively. The idea is that we should say L is the limit of a sequence $\{a_n\}$ if for any open interval I containing L, no matter how small, we can find a point in the sequence beyond which all values of the sequence lie in I. Graphically, this means that if we start plotting the points of the sequence, there will come a time when all points from then on will lie within the interval I. This idea is formalized in the following

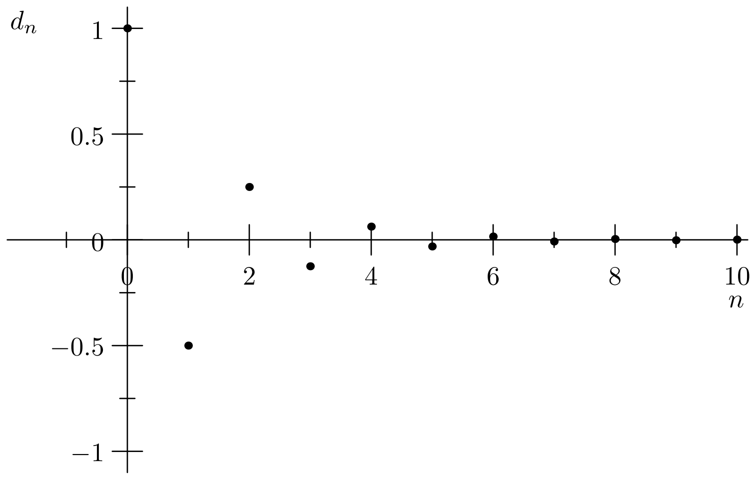

Figure 1.2.3. Plot of $\left(n, \frac{(-1)^n}{2^n}\right)$, $n = 1, 2, 3, \ldots, 10$

definition, where the open interval I is expressed in the form $(L - \epsilon, L + \epsilon)$ and the idea that all values of the sequence beyond a certain point are in this interval is expressed by requiring that $|a_n - L| < \epsilon$, that is, the distance between a_n and L is less than ϵ for all $n > N$.

Definition 1.2.1. We say that the *limit* of the sequence $\{a_n\}$ is L, written

$$\lim_{n\to\infty} a_n = L,$$

if for every $\epsilon > 0$ there exists an integer N such that $|a_n - L| < \epsilon$ whenever $n > N$.

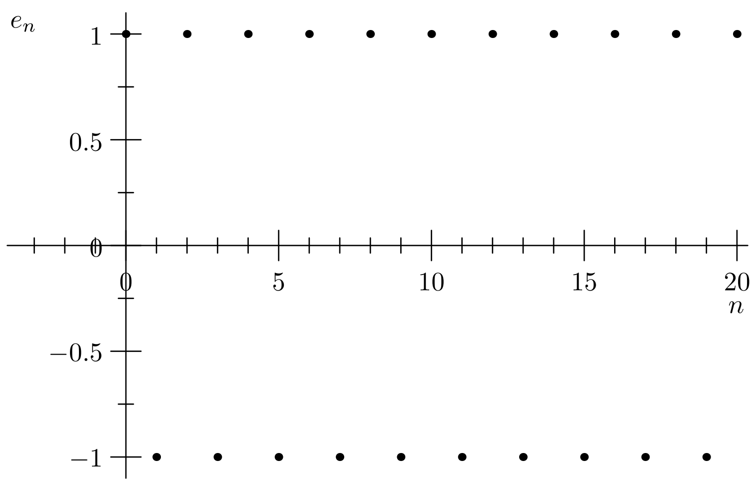

Figure 1.2.4. Plot of $(n, (-1)^n)$, $n = 1, 2, 3, \ldots, 20$

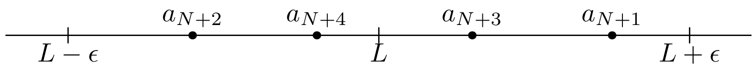

Figure 1.2.5. a_n in $(L-\epsilon, L+\epsilon)$ for $n > N$

Hence to show that the limit of a sequence is a number L, one must show that for any positive number ϵ, it is possible to find an integer N such that the numbers $a_{N+1}, a_{N+2}, a_{N+3}, \dots$ are all in the interval $(L-\epsilon, L+\epsilon)$. See Figure 1.2.5.

Example 1.2.1. We will show that

$$\lim_{n\to\infty} \frac{1}{n} = 0.$$

To do so, we must show that for any given $\epsilon > 0$, we can find an integer N such that

$$\left|\frac{1}{n} - 0\right| < \epsilon$$

whenever $n > N$. Now

$$\left|\frac{1}{n} - 0\right| = \frac{1}{n},$$

so we need only determine the values of n for which

$$\frac{1}{n} < \epsilon.$$

Since

$$\frac{1}{n} < \epsilon \text{ if and only if } n > \frac{1}{\epsilon},$$

it follows that we may take N to be the largest integer less than or equal to $\frac{1}{\epsilon}$. Then whenever $n > N$, we have

$$n > \frac{1}{\epsilon},$$

from which it follows that

$$\frac{1}{n} < \epsilon.$$

This is exactly what we need in order to conclude, by the definition, that

$$\lim_{n\to\infty} \frac{1}{n} = 0.$$

The following definition is useful in situations, such as in the previous example, when we want the largest integer less than or equal to some given value.

Definition 1.2.2. For any real number x, we may define the *floor function*, denoted $\lfloor x \rfloor$, by

$$\lfloor x \rfloor = \text{ the largest integer less than or equal to } x, \tag{1.2.1}$$

and the *ceiling function*, denoted $\lceil x \rceil$, by

$$\lceil x \rceil = \text{ the smallest integer greater than or equal to } x. \tag{1.2.2}$$

For example, $\lfloor 5.3 \rfloor = 5$, $\lceil \pi \rceil = 4$, $\lfloor 3 \rfloor = 3$, and $\lceil 3 \rceil = 3$. With this notation, we could define N in the previous example by

$$N = \left\lfloor \frac{1}{\epsilon} \right\rfloor.$$

Example 1.2.2. We will show that

$$\lim_{n\to\infty} \frac{1}{2^n} = 0.$$

This time we must show that for any $\epsilon > 0$, we can find an integer N such that

$$\left|\frac{1}{2^n} - 0\right| < \epsilon$$

whenever $n > N$. Now

$$\left|\frac{1}{2^n} - 0\right| = \frac{1}{2^n} = \left(\frac{1}{2}\right)^n,$$

so we need to determine the values of n for which

$$\left(\frac{1}{2}\right)^n < \epsilon.$$

We need to solve this inequality for n. Since n is in the exponent, we may use logarithms to simplify the inequality. Although we will not provide a careful treatment of logarithms until Chapter 6, we will assume for the moment some acquaintance with logarithms using base 10. Now

$$\left(\frac{1}{2}\right)^n < \epsilon$$

if and only if

$$\log_{10}\left(\frac{1}{2}\right)^n < \log_{10}(\epsilon).$$

Since

$$\log_{10}\left(\frac{1}{2}\right)^n = n\log_{10}\left(\frac{1}{2}\right),$$

we have

$$\left(\frac{1}{2}\right)^n < \epsilon$$

if and only if

$$n\log_{10}\left(\frac{1}{2}\right) < \log_{10}(\epsilon).$$

Now $\log_{10}\left(\frac{1}{2}\right) < 0$, so

$$n\log_{10}\left(\frac{1}{2}\right) < \log_{10}(\epsilon)$$

if and only if

$$n > \frac{\log_{10}(\epsilon)}{\log_{10}\left(\frac{1}{2}\right)}.$$

Thus if we let

$$N = \left\lfloor \frac{\log_{10}(\epsilon)}{\log_{10}\left(\frac{1}{2}\right)} \right\rfloor,$$

then

$$\left|\frac{1}{2^n} - 0\right| < \epsilon$$

whenever $n > N$. For example, if we take $\epsilon = 0.001$, then, to two decimal places,

$$\frac{\log_{10}(\epsilon)}{\log_{10}\left(\frac{1}{2}\right)} = 9.97,$$

and so we would have

$$N = \lfloor 9.97 \rfloor = 9.$$

This N works because, for $n > 9$,

$$\left|\frac{1}{2^n} - 0\right| = \frac{1}{2^n} \leq \frac{1}{2^{10}} = \frac{1}{1024} < 0.001.$$

Problem 12 at the end of this section will ask you to generalize the previous example to show that

$$\lim_{n\to\infty} r^n = 0$$

whenever $|r| < 1$. This is an important fact that we will make use of later.

In this course we will be concerned more with the development of an intuitive understanding of limits and a computational facility with limits than with the formalism of verifying a specific limit using the above definition. That is not to say that the definition is unimportant; rather a good grasp of the concept in the definition is important for a full understanding of much of what we will do in calculus. In fact, mathematicians of the 19th century arrived at the definition we have stated in their attempts to clarify confusions that had developed in mathematics since the time of Newton and Leibniz. However, for the most part these difficulties are beyond the scope of a text such as this one.

We will see that a few basic properties of limits, combined with a few simple limits like the ones in the previous two examples, will enable us to compute easily a large number of limits. To begin considering these properties, consider the case where we already know that

$$\lim_{n\to\infty} a_n = L \tag{1.2.3}$$

and we want to compute

$$\lim_{n\to\infty} ka_n$$

for some constant $k \neq 0$. Now (1.2.3) tells us that for any $\epsilon > 0$, we may find an integer N such that for $n > N$,

$$|a_n - L| < \frac{\epsilon}{|k|}.$$

It follows that for $n > N$,

$$|ka_n - kL| = |k||a_n - L| < |k|\frac{\epsilon}{|k|} = \epsilon.$$

But this is what it means to say that

$$\lim_{n\to\infty} ka_n = kL. \tag{1.2.4}$$

Note that (1.2.4) is obviously true as well when $k = 0$. Hence we have the following proposition.

Proposition 1.2.1. *If $\{a_n\}$ is a sequence for which*

$$\lim_{n\to\infty} a_n = L,$$

then for any constant k we have

$$\lim_{n\to\infty} ka_n = k \lim_{n\to\infty} a_n = kL. \tag{1.2.5}$$

Example 1.2.3. Since we have already seen that

$$\lim_{n\to\infty} \frac{1}{n} = 0,$$

it follows that

$$\lim_{n\to\infty} \frac{350}{n} = 350 \lim_{n\to\infty} \frac{1}{n} = (350)(0) = 0.$$

Now suppose we have two sequences $\{a_n\}$ and $\{b_n\}$ with

$$\lim_{n\to\infty} a_n = L \tag{1.2.6}$$

and

$$\lim_{n\to\infty} b_n = M. \tag{1.2.7}$$

Then (1.2.6) and (1.2.7) tell us that for any $\epsilon > 0$, we can find integers N_1 and N_2 such that

$$|a_n - L| < \frac{\epsilon}{2}$$

whenever $n > N_1$ and

$$|b_n - M| < \frac{\epsilon}{2}$$

whenever $n > N_2$. If we let N be the larger of N_1 and N_2, then whenever $n > N$ we will have

$$\begin{aligned} |(a_n + b_n) - (L + M)| &= |(a_n - L) + (b_n - M)| \\ &\le |a_n - L| + |b_n - M| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned} \tag{1.2.8}$$

Note that in (1.2.8) we have used the fact, known as the *triangle inequality*, that for any real numbers x and y,

$$|x + y| \le |x| + |y|. \tag{1.2.9}$$

Thus we have shown

$$\lim_{n\to\infty} (a_n + b_n) = L + M. \tag{1.2.10}$$

Hence we have the following proposition.

Proposition 1.2.2. *If $\{a_n\}$ and $\{b_n\}$ are sequences with*

$$\lim_{n\to\infty} a_n = L$$

and

$$\lim_{n\to\infty} b_n = M,$$

then

$$\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n = L + M. \tag{1.2.11}$$

Example 1.2.4. We have

$$\lim_{n\to\infty} \left(4 + \frac{8}{n}\right) = \lim_{n\to\infty} 4 + \lim_{n\to\infty} \frac{8}{n} = 4 + 8 \lim_{n\to\infty} \frac{1}{n} = 4 + (8)(0) = 4.$$

Note that in the last example we used the fact that if k is a constant and $a_n = k$ for all n, then

$$\lim_{n\to\infty} a_n = k.$$

This follows immediately from the definition since

$$|a_n - k| = 0$$

for all values of n, and so any integer N will work for any value of ϵ.

Again suppose we have two sequences $\{a_n\}$ and $\{b_n\}$ with

$$\lim_{n\to\infty} a_n = L$$

and

$$\lim_{n\to\infty} b_n = M.$$

Then we have

$$\lim_{n\to\infty}(a_n - b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty}(-b_n) = \lim_{n\to\infty} a_n + (-1)\lim_{n\to\infty} b_n = L - M. \quad (1.2.12)$$

Proposition 1.2.3. *If $\{a_n\}$ and $\{b_n\}$ are sequences with*

$$\lim_{n\to\infty} a_n = L$$

and

$$\lim_{n\to\infty} b_n = M,$$

then

$$\lim_{n\to\infty}(a_n - b_n) = \lim_{n\to\infty} a_n - \lim_{n\to\infty} b_n = L - M. \quad (1.2.13)$$

Example 1.2.5. We have

$$\lim_{n\to\infty}\left(\frac{3}{n} - \frac{8}{5^n}\right) = 3\lim_{n\to\infty}\frac{1}{n} - 8\lim_{n\to\infty}\left(\frac{1}{5}\right)^n = (3)(0) - (8)(0) = 0.$$

Note that we have used the result that

$$\lim_{n\to\infty} r^n = 0$$

whenever $|r| < 1$.

We will state three more properties of limits without justifications. Although the reasoning behind these results is similar to the reasoning of the previous three propositions, they require a little more care and are best left to a more advanced course.

Proposition 1.2.4. *If $\{a_n\}$ and $\{b_n\}$ are sequences with*

$$\lim_{n\to\infty} a_n = L$$

and

$$\lim_{n\to\infty} b_n = M,$$

then

$$\lim_{n\to\infty} a_n b_n = (\lim_{n\to\infty} a_n)(\lim_{n\to\infty} b_n) = LM. \quad (1.2.14)$$

Example 1.2.6. We have

$$\lim_{n\to\infty} \frac{1}{n^2} = \left(\lim_{n\to\infty} \frac{1}{n}\right)\left(\lim_{n\to\infty} \frac{1}{n}\right) = (0)(0) = 0.$$

Proposition 1.2.5. *If $\{a_n\}$ and $\{b_n\}$ are sequences with*

$$\lim_{n\to\infty} a_n = L$$

and

$$\lim_{n\to\infty} b_n = M,$$

then

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n} = \frac{L}{M}, \tag{1.2.15}$$

provided $M \neq 0$ and $b_n \neq 0$ for all n.

Example 1.2.7. We have

$$\lim_{n\to\infty} \frac{n-3}{2n+4} = \lim_{n\to\infty} \frac{\frac{n-3}{n}}{\frac{2n+4}{n}} = \lim_{n\to\infty} \frac{1-\frac{3}{n}}{2+\frac{4}{n}} = \frac{\lim_{n\to\infty}\left(1-\frac{3}{n}\right)}{\lim_{n\to\infty}\left(2+\frac{4}{n}\right)} = \frac{1}{2}.$$

Note that we can apply the previous proposition only when both numerator and denominator have a limit. Hence, in this example, we first divided the numerator and denominator by n to put the problem in a form to which we could apply the proposition.

Proposition 1.2.6. *Suppose $\{a_n\}$ is a sequence with*

$$\lim_{n\to\infty} a_n = L.$$

Moreover, suppose p is a rational number, a_n^p is defined for all n, and L^p is defined. Then

$$\lim_{n\to\infty} a_n^p = (\lim_{n\to\infty} a_n)^p = L^p. \tag{1.2.16}$$

Note that the conditions stated in this proposition are necessary to avoid even roots of negative numbers.

Example 1.2.8. We have

$$\lim_{n\to\infty} \sqrt{4-\frac{3}{n}} = \sqrt{\lim_{n\to\infty}\left(4-\frac{3}{n}\right)} = \sqrt{4} = 2.$$

Example 1.2.9. For any rational number $p > 0$, we have

$$\lim_{n\to\infty} \frac{1}{n^p} = \left(\lim_{n\to\infty} \frac{1}{n}\right)^p = 0^p = 0.$$

Example 1.2.10. We have

$$\lim_{n\to\infty}\left(18-\frac{5}{n}+\frac{23}{n^5}\right) = \lim_{n\to\infty} 18 - 5\lim_{n\to\infty}\frac{1}{n} + 23\lim_{n\to\infty}\frac{1}{n^5} = 18-(5)(0)+(23)(0) = 18.$$

Example 1.2.11. Dividing numerator and denominator by n^5, we have

$$\lim_{n\to\infty} \frac{4n^5+5n^2-6}{3n^5+4n-18} = \lim_{n\to\infty} \frac{4+\frac{5}{n^3}-\frac{6}{n^5}}{3+\frac{4}{n^4}-\frac{18}{n^5}} = \frac{\lim_{n\to\infty}\left(4+\frac{5}{n^3}-\frac{6}{n^5}\right)}{\lim_{n\to\infty}\left(3+\frac{4}{n^4}-\frac{18}{n^5}\right)} = \frac{4}{3}.$$

In general, for sequences of the form of the previous example it is useful to divide both numerator and denominator by the highest power of n which occurs in the denominator.

Example 1.2.12. As another illustration of the idea in the previous example, we have

$$\lim_{n\to\infty} \frac{3n^2+2n-1}{2n^3-16n} = \lim_{n\to\infty} \frac{\frac{3}{n}+\frac{2}{n^2}-\frac{1}{n^3}}{2-\frac{16}{n^2}} = \frac{0}{2} = 0.$$

Definition 1.2.3. If $\lim_{n\to\infty} a_n$ exists, we say the sequence $\{a_n\}$ *converges*. If the sequence $\{a_n\}$ does not have a limit, we say the sequence *diverges*.

An important class of divergent sequences are those for which a limit does not exist either because the terms eventually remain greater than any specified positive bound or because they eventually remain less than any specified negative bound, as defined in the following definition.

Definition 1.2.4. A sequence $\{a_n\}$ is said to *diverge to infinity* if for any real number M there exists an integer N such that $a_n > M$ whenever $n > N$, in which case we write

$$\lim_{n\to\infty} a_n = \infty.$$

A sequence $\{a_n\}$ is said to *diverge to negative infinity* if for any real number M there exists an integer N such that $a_n < M$ whenever $n > N$, in which case we write

$$\lim_{n\to\infty} a_n = -\infty.$$

Example 1.2.13. It is easy to see that

$$\lim_{n\to\infty} n^p = \infty$$

for any value of $p > 0$. For given any M, we need only take

$$N = \left\lfloor \sqrt[p]{|M|} \right\rfloor$$

to guarantee that $a_n > M$ whenever $n > N$.

Example 1.2.14. We have

$$\lim_{n\to\infty} 2^n = \infty$$

since, given any M, $2^n > M$ for all n if $M \leq 0$ and $2^n > M$ provided

$$n > \frac{\log_{10}(M)}{\log_{10}(2)}$$

if $M > 0$.

Suppose the sequence $\{a_n\}$ diverges and $k \neq 0$ is a constant. Then the sequence $\{ka_n\}$ must also diverge since if $\{ka_n\}$ converged, then the sequence with nth term

$$\frac{1}{k}(ka_n) = a_n$$

would also converge, contradicting our assumption that $\{a_n\}$ diverges.

Proposition 1.2.7. *If the sequence $\{a_n\}$ diverges and $k \neq 0$ is a constant, then the sequence $\{ka_n\}$ also diverges.*

If the sequence $\{a_n\}$ diverges and the sequence $\{b_n\}$ converges, then the sequence $\{a_n + b_n\}$ also diverges since, if it converged, then the sequence with nth term

$$(a_n + b_n) - b_n = a_n$$

would also converge, contradicting our assumption that $\{a_n\}$ diverges. Similarly, the sequence $\{a_n - b_n\}$ diverges.

Proposition 1.2.8. *If the sequence $\{a_n\}$ diverges and the sequence $\{b_n\}$ converges, then the sequences $\{a_n + b_n\}$ and $\{a_n - b_n\}$ both diverge.*

Suppose the sequence $\{a_n\}$ diverges, the sequence $\{b_n\}$ converges, and

$$\lim_{n\to\infty} b_n \neq 0. \tag{1.2.17}$$

Now (1.2.17) implies that we can find an integer N such that $b_n \neq 0$ for all $n > N$. So if the sequence $\{a_n b_n\}$ converged, then the sequence with, for $n > N$, nth term,

$$\frac{1}{b_n}(a_n b_n) = a_n$$

would also converge, contradicting our assumption that $\{a_n\}$ diverges. Hence $\{a_n b_n\}$ must diverge.

Proposition 1.2.9. *If the sequence $\{a_n\}$ diverges, the sequence $\{b_n\}$ converges, and*

$$\lim_{n\to\infty} b_n \neq 0,$$

then the sequence $\{a_n b_n\}$ diverges

Finally, if the sequence $\{a_n\}$ diverges, the sequence $\{b_n\}$ converges, and $b_n \neq 0$ for all n, then the sequence

$$\left\{\frac{a_n}{b_n}\right\}$$

diverges since, if it converged, the sequence with nth term

$$b_n\left(\frac{a_n}{b_n}\right) = a_n$$

would also converge, contradicting our assumption that $\{a_n\}$ diverges.

Proposition 1.2.10. *If the sequence $\{a_n\}$ diverges, the sequence $\{b_n\}$ converges, and $b_n \neq 0$ for all n, then the sequence*

$$\left\{\frac{a_n}{b_n}\right\}$$

diverges.

Example 1.2.15. Consider

$$\lim_{n\to\infty} \frac{4n^3 + n - 2}{5n^2 - 7n} = \lim_{n\to\infty} \frac{4n + \frac{1}{n} - \frac{2}{n^2}}{5 - \frac{7}{n}}. \tag{1.2.18}$$

Now

$$\lim_{n\to\infty} 4n = \infty$$

and

$$\lim_{n\to\infty} \left(\frac{1}{n} - \frac{2}{n^2}\right) = 0,$$

so

$$\lim_{n\to\infty} \left(4n + \frac{1}{n} - \frac{2}{n^2}\right) = \infty.$$

Moreover,

$$\lim_{n\to\infty} \left(5 - \frac{7}{n}\right) = 5.$$

Thus the numerator in (1.2.18) diverges while the denominator converges. Hence the ratio diverges. In fact, it should be clear that

$$\lim_{n\to\infty} \frac{4n^3 + n - 2}{5n^2 - 7n} = \lim_{n\to\infty} \frac{4n + \frac{1}{n} - \frac{2}{n^2}}{5 - \frac{7}{n}} = \infty.$$

Note that in the previous example it was once again useful to divide numerator and denominator by the highest power of n in the denominator.

Example 1.2.16. We have

$$\lim_{n\to\infty} \frac{15 - 26n^5}{13 + n^2} = \lim_{n\to\infty} \frac{\frac{15}{n^2} - 26n^3}{\frac{13}{n^2} + 1} = -\infty.$$

Example 1.2.17. The absolute values of the terms of the sequence $\{(-2)^n\}$ grow without bound, and so the sequence diverges. However, since the terms alternate in sign, the sequence neither diverges to ∞ nor to $-\infty$.

1.2.1 Monotone sequences.

It is sometimes possible to determine that a given sequence converges without explicitly computing the limit. One important case involves *monotone sequences.*

Definition 1.2.5. We say a sequence $\{a_n\}$ is *monotone increasing* if $a_n \leq a_{n+1}$ for all n. We say a sequence $\{a_n\}$ is *monotone decreasing* if $a_n \geq a_{n+1}$ for all n. We say a sequence is *monotone* if it is either monotone increasing or monotone decreasing.

Now suppose $\{a_n\}$ is a monotone increasing sequence. For such a sequence there either exists a number P such that $a_n \leq P$ for all n or there does not exist such a P. In the latter case, given any real number M, it is then possible to find an integer N such that $a_N > M$. Since the sequence is monotone, it follows that $a_n > M$ for all $n > N$, and so the sequence diverges to infinity. On the other hand, if there does exist a number P such that $a_n \leq P$ for all n, then there in fact exists a number B such that $a_n \leq B$ for all n and $B \leq P$ for any number P with the property that $a_n \leq P$ for all n. The existence of B, known as the *least upper bound* of the sequence $\{a_n\}$, is not at all obvious; indeed, the subtle properties of the real numbers that imply the existence of B were not fully understood until the middle part of the 19th century. However, given the existence of B, it is easy to see that given any $\epsilon > 0$, there exists an integer N for which $a_N > B - \epsilon$

(if not, then $B - \epsilon$ would be an upper bound for the sequence smaller than B). Since the sequence is monotone increasing and $a_n < B$ for all n, it follows that

$$|a_n - B| < \epsilon$$

for all $n > N$. That is, we have shown that the sequence converges and

$$\lim_{n\to\infty} a_n = B.$$

Similar results hold for sequences which are monotone decreasing.

Theorem 1.2.11 (Monotone sequence theorem)**.** *Suppose the sequence $\{a_n\}$ is monotone. If the sequence is monotone increasing and there exists a number P such that $a_n \leq P$ for all n, then the sequence converges. If the sequence is monotone increasing and no such number P exists, then*

$$\lim_{n\to\infty} a_n = \infty.$$

If the sequence is monotone decreasing and there exists a number Q such that $a_n \geq Q$ for all n, then the sequence converges. If the sequence is monotone decreasing and no such number Q exists, then

$$\lim_{n\to\infty} a_n = -\infty.$$

Example 1.2.18. As we shall see in Sections 1.4 and 1.5, we often work with sequences without having an explicit formula for each term in the sequence. For example, suppose all we know about the sequence $\{a_n\}$ is that $a_1 = 4$ and

$$a_{n+1} = \frac{1}{2}a_n$$

for $n = 1, 2, 3, \ldots$. That is, the first term in the sequence is 4 and then each successive term is one-half of its predecessor. Thus

$$\begin{aligned} a_1 &= 4, \\ a_2 &= 2, \\ a_3 &= 1, \\ a_4 &= \frac{1}{2}, \end{aligned}$$

and so on. Hence $\{a_n\}$ is monotone decreasing. Moreover, every term in the sequence is positive, so $a_n \geq 0$ for all n. Thus, by the monotone sequence theorem, $\{a_n\}$ converges. Moreover, note that

$$a_{n+1} = \frac{1}{2}a_n$$

implies that

$$\lim_{n\to\infty} a_{n+1} = \frac{1}{2}\lim_{n\to\infty} a_n. \tag{1.2.19}$$

If we let

$$L = \lim_{n\to\infty} a_n = \lim_{n\to\infty} a_{n+1},$$

then (1.2.19) becomes

$$L = \frac{1}{2}L.$$

Hence $L = 0$. That is,

$$\lim_{n\to\infty} a_n = 0.$$

Problems 1.2

(1) For each of the following, find a general expression for the nth term of a sequence which would yield these values as the first four terms.

(a) $1, \frac{1}{3}, \frac{1}{9}, \frac{1}{27}, \ldots$ (b) $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$

(c) $1, \frac{3}{2}, \frac{5}{3}, \frac{7}{4}, \ldots$ (d) $-\frac{1}{3}, \frac{1}{5}, -\frac{1}{7}, \frac{1}{9}, \ldots$

(2) For each of the following, decide whether the given sequence converges or diverges. If the sequence converges, find its limit.

(a) $a_n = \frac{1}{3^n}, n = 0, 1, 2, \ldots$ (b) $a_n = \pi^n, n = 0, 1, 2, \ldots$

(c) $b_n = \frac{3n-1}{2n+6}, n = 1, 2, 3, \ldots$ (d) $c_n = \cos(\pi n), n = 0, 1, 2, \ldots$

(e) $a_n = \frac{3n^4-6n^3+1}{5n^3+n^2+2}, n = 1, 2, 3, \ldots$ (f) $b_n = \frac{2n^5-3n^2+23}{7n^5+13n^4-12}, n = 1, 2, 3, \ldots$

(g) $c_n = \frac{45-16n^2}{13+5n+6n^3}, n = 1, 2, 3, \ldots$ (h) $b_n = \frac{3n+1}{\sqrt{4n^2+1}}, n = 1, 2, 3, \ldots$

(i) $a_n = (-2)^{2n+1}, n = 1, 2, 3, \ldots$ (j) $a_n = \frac{10-16n^3}{1+n^2}, n = 1, 2, 3, \ldots$

(k) $a_n = \sqrt{\frac{3n^2+n-6}{5n^2+16}}, n = 1, 2, 3, \ldots$ (l) $b_n = \frac{(-1)^n}{5^n}, n = 0, 1, 2, \ldots$

(3) Explain why

$$-\frac{1}{n} \leq \frac{\sin(n)}{n} \leq \frac{1}{n}$$

for $n = 1, 2, 3, \ldots$. What can you conclude about $\lim_{n\to\infty} \frac{\sin(n)}{n}$?

(4) Let $a_n = \left(1 + \frac{1}{n}\right)^n, n = 1, 2, 3, \ldots$.

(a) Compute a_1, a_2, a_3, a_4, and a_5 using a calculator.

(b) Compute values of a_n for $n = 1, 2, 3, \ldots, 200$.

(c) Plot the points (n, a_n) for $n = 1, 2, 3, \ldots, 200$, along with the horizontal line $y = e$.

(d) Does it seem reasonable that $\lim_{n\to\infty} a_n = e$?

(e) What is the smallest value of n for which $a_n > 2.7$?

(f) What is the first value of n for which $|a_n - e| < 0.01$? Recall that $e = 2.71828$ to five decimal places.

(5) Let $a_n = n \sin\left(\frac{1}{n}\right), n = 1, 2, 3, \ldots$.

(a) Compute a_1, a_2, a_3, a_4, and a_5 using a calculator.

(b) Compute values of a_n for $n = 1, 2, 3, \ldots, 200$.

(c) Plot the points (n, a_n) for $n = 1, 2, 3, \ldots, 200$, along with the horizontal line $y = 1$.

(d) Does it seem reasonable that $\lim_{n\to\infty} a_n = 1$?

(e) What is the smallest value of n for which $a_n > 0.999$?

(f) What is the first value of n for which $|a_n - 1| < 0.0001$?

(6) Let $a_n = 1.01^n$ and $b_n = 0.99^n$ for $n = 0, 1, 2, \ldots$. On the same graph, plot the points (n, a_n) and (n, b_n) for $n = 0, 1, 2, \ldots, 200$. How do these two plots compare? Do the sequences converge?

(7) Let $a_n = \frac{10^n}{n!}$ for $n = 1, 2, 3, \ldots$.

(a) Plot the points (n, a_n) for $n = 1, 2, 3, \ldots, 100$.

(b) From the picture in part (a), can you guess $\lim_{n\to\infty} a_n$?

(c) What is the maximum value of a_n for $n = 1, 2, 3, \ldots, 100$?

(d) Can you see why

$$\lim_{n\to\infty} \frac{k^n}{n!} = 0$$

for any constant k?

(8) Consider the sequence $\{a_n\}$ with $a_1 = 10$ and

$$a_{n+1} = \frac{1}{3}a_n$$

for $n = 1, 2, 3, \ldots$. Plot the points (n, a_n) for $n = 1, 2, 3, \ldots, 50$. Do you think this sequence has a limit? Can you verify this?

(9) Consider the sequence $\{a_n\}$ with $a_1 = 2$ and

$$a_{n+1} = 2a_n$$

for $n = 1, 2, 3, \ldots$. Plot the points (n, a_n) for $n = 1, 2, 3, \ldots, 50$. Can you find the limit of this sequence using the same method you used in Problem 8? Does this sequence have a limit?

(10) Consider the sequence $\{a_n\}$ with $a_1 = 0.9$ and

$$a_{n+1} = 2a_n(1 - a_n)$$

for $n = 1, 2, 3, \ldots$. Plot the points (n, a_n) for $n = 1, 2, 3, \ldots, 100$. Do you think this sequence has a limit? If so, can you find it?

(11) In each of the following, for an arbitrary $\epsilon > 0$, find the smallest integer N for which $|a_n - L| < \epsilon$ whenever $n > N$. Verify that your value for N works in the particular case $\epsilon = 0.001$.

(a) $a_n = 1 - \frac{1}{n}, L = 1$ (b) $a_n = 0.98^n, L = 0$

(c) $a_n = \frac{1}{n^2}, L = 0$ (d) $a_n = \frac{3n^3-1}{n^3}, L = 3$

(12) Show that for any $-1 < r < 1$, $\lim_{n\to\infty} r^n = 0$.

(13) Find sequences $\{a_n\}$ and $\{b_n\}$ such that $\{a_n\}$ and $\{b_n\}$ both diverge, but $\{a_n + b_n\}$ converges.

(14) Find sequences $\{a_n\}$ and $\{b_n\}$ such that $\{a_n\}$ diverges, $\{b_n\}$ converges, and $\{a_nb_n\}$ converges.

1.3 The sum of a sequence

This section considers the problem of adding together the terms of a sequence. Of course, this is a problem only if more than a finite number of terms of the sequence are nonzero. In this case, we must decide what it means to add together an infinite number of nonzero numbers. The first example shows how a relatively simple question may lead to such infinite summations.

Example 1.3.1. Suppose a game is played in which a fair coin is tossed until the first time a head appears. What is the probability that a head appears for the first time on an even-numbered toss? To solve this problem, we first need to determine the probability of obtaining a head for the first time on any given even-numbered toss, and then we need to add all these probabilities together. Let P_n denote the probability that the first head appears on the nth toss, $n = 1, 2, 3, \dots$. Then, since the coin is assumed to be fair,

$$P_1 = \frac{1}{2}.$$

Now in order to get a head for the first time on the second toss, we must toss a tail on the first toss and then follow that with a head on the second toss. Since one-half of all first tosses will be tails and then one-half of those tosses will be followed by a second toss of heads, we should have

$$P_2 = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}.$$

Similarly, since one-fourth of all sequences of coin tosses will begin with two tails and then half of these sequences will have a head for the third toss, we have

$$P_3 = \left(\frac{1}{4}\right)\left(\frac{1}{2}\right) = \frac{1}{8}.$$

Continuing in this fashion, it should seem reasonable that, for any $n = 1, 2, 3, \dots$,

$$P_n = \frac{1}{2^n}.$$

Hence we have a sequence of probabilities $\{P_n\}$ for $n = 1, 2, 3, \dots$, and, in order to find the desired probability, we need to add up the even-numbered terms in this sequence. Namely, the probability that a head appears for the first time on an even toss is given by

$$P_2 + P_4 + P_6 + \cdots = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \cdots. \tag{1.3.1}$$

But this involves adding together an infinite number of nonzero values. Is this possible? Can we perform the operation of addition an infinite number of times? In this case the answer is yes, but we will need a few preliminaries before we can finish this particular example.

We begin with a definition of the sum of a sequence $\{a_n\}$. The idea is to create a new sequence by successively adding together the terms of the original sequence. That is, we define a new sequence $\{s_n\}$ where s_n is the sum of the first n terms of the original sequence. If

$$\lim_{n\to\infty} s_n$$

exists, then this indicates that, as we add together more and more terms of $\{a_n\}$, the resulting sums approach a limiting value. It is then reasonable to call this limiting value the sum of the sequence. For example, if

$$a_n = \frac{1}{2^n}$$

for $n = 1, 2, 3, \ldots$, then we would have

$$\begin{aligned} s_1 &= \frac{1}{2}, \\ s_2 &= \frac{1}{2} + \frac{1}{4} = \frac{3}{4}, \\ s_3 &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}, \\ s_4 &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{15}{16}, \end{aligned}$$

and so on. If you plot these points on the real line, you may think of starting at $\frac{1}{2}$, moving $\frac{1}{2}$ the distance to 1 to plot the next point, then $\frac{1}{2}$ the remaining distance to 1 to plot the next point, and so on. After n points, you would be at

$$s_n = 1 - \frac{1}{2^n}. \tag{1.3.2}$$

Clearly,

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} \left(1 - \frac{1}{2^n}\right) = 1,$$

and it would be reasonable to say that the sequence adds up to 1. That is,

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots + \frac{1}{2^n} + \cdots = 1. \tag{1.3.3}$$

This idea is formalized in the following definition.

Definition 1.3.1. Given a sequence $\{a_n\}$, $n = 1, 2, 3, \ldots$, we define a new sequence $\{s_n\}$ by letting

$$s_n = a_1 + a_2 + \cdots + a_n \tag{1.3.4}$$

for $n = 1, 2, 3, \ldots$. If the sequence $\{s_n\}$ converges, then we call

$$s = \lim_{n\to\infty} s_n$$

the *sum* of the sequence $\{a_n\}$. The sequence $\{s_n\}$ is called an *infinite series*, or simply a *series*, and an individual term s_n of this sequence is called a *partial sum* of the sequence $\{a_n\}$.

Note that we have assumed that the first term in the sequence $\{a_n\}$ in the definition is a_1. The sequence could just as well start with any other integer index, in which case the sequence of partial sums $\{s_n\}$ would start with the same index. For example, if the first term of the sequence is a_0, then the first partial sum is s_0.

Since summations involving an infinite number, or even a large finite number, of terms are cumbersome to write using the standard plus sign of addition, Σ (the capital Greek *sigma*) is used to denote the process of summation. In particular, we would write

$$s_n = \sum_{j=1}^{n} a_j = a_1 + a_2 + \cdots + a_n \tag{1.3.5}$$

and

$$s = \lim_{n\to\infty} s_n = \lim_{n\to\infty} \sum_{j=1}^{n} a_j = \lim_{n\to\infty} (a_1 + a_2 + \cdots + a_n). \tag{1.3.6}$$

Since (1.3.6) is what we mean by an infinite sum, we will in fact write

$$s = \sum_{j=1}^{\infty} a_j = a_1 + a_2 + \cdots + a_n + \cdots. \tag{1.3.7}$$

For example, in this notation, we may restate our earlier results as

$$\sum_{n=1}^{n} \frac{1}{2^n} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{15}{16},$$

and

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots + \frac{1}{2^n} + \cdots = 1.$$

We should note that, since the sum of a sequence is the limit of another sequence, and not all sequences have limits, there are sequences which do not have sums. For example, the sequence with terms $a_n = 1$ for $n = 1, 2, 3, \ldots$ does not have a sum since

$$s_n = \underbrace{1 + 1 + \cdots + 1}_{n \text{ times}} = n,$$

from which it follows that

$$\lim_{n\to\infty} s_n = \infty.$$

For another example, the sequence $\{(-1)^n\}$, $n = 0, 1, 2, \ldots$, does not have a sum since

$$s_n = \begin{cases} 1, & \text{if } n = 0, 2, 4, \ldots, \\ 0, & \text{if } n = 1, 3, 5, \ldots, \end{cases}$$

a sequence which does not have a limit.

In general it may be difficult to determine the sum of a sequence; in fact, it may be difficult to determine even if the sequence has a sum. We will return to this problem in Chapter 5 when we have more tools at our disposal, as well as more motivation for studying infinite series. For now we will look at an important class of sequences for which the sum is determined with relative ease. These are the sequences for which the terms are in geometric progression; that is, sequences for which successive terms have a common ratio. We call the infinite series which corresponds to such a sequence a *geometric series.*

1.3.1 Geometric series. Suppose $\{a_n\}$ is a sequence with $a_n = cr^{n-1}$, where $c \neq 0$ and r are constants and $n = 1, 2, 3, \ldots$. Then the partial sums are

$$\begin{aligned} s_n &= a_1 + a_2 + a_3 + \cdots + a_n \\ &= c + cr + cr^2 + \cdots + cr^{n-1} \\ &= c(1 + r + r^2 + \cdots + r^{n-1}). \end{aligned} \tag{1.3.8}$$

If $r = 1$, $s_n = nc$ and so $\{s_n\}$ does not converge. If $r \neq 1$, one may show, using long division (or the derivation outlined in Problem 4), that

$$\frac{1 - r^n}{1 - r} = 1 + r + r^2 + \cdots + r^{n-1}. \tag{1.3.9}$$

Hence, if $r \neq 1$,

$$s_n = \frac{c(1-r^n)}{1-r}. \tag{1.3.10}$$

From (1.3.10), it is clear that $\{s_n\}$ does not converge if $|r| \geq 1$. But if $-1 < r < 1$, then

$$\lim_{n\to\infty} r^{n-1} = 0,$$

and so the sequence has the sum

$$s = \lim_{n\to\infty} s_n = \lim_{n\to\infty} \frac{c(1-r^n)}{1-r} = \frac{c}{1-r}. \tag{1.3.11}$$

That is, we have now seen that

$$\sum_{n=1}^{\infty} cr^{n-1} = \frac{c}{1-r} \tag{1.3.12}$$

whenever $-1 < r < 1$.

Example 1.3.2. We have, using (1.3.11) with $c = 1$ and $r = \frac{1}{2}$,

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^{n-1} = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = \frac{1}{1-\frac{1}{2}} = 2.$$

Note that this agrees with our previous result that

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = 1.$$

Example 1.3.3. We have

$$\sum_{n=1}^{\infty} \left(\frac{2}{3}\right)^{n} = \sum_{n=1}^{\infty} \frac{2}{3}\left(\frac{2}{3}\right)^{n-1} = \frac{\frac{2}{3}}{1-\frac{2}{3}} = 2.$$

Example 1.3.4. We have

$$\sum_{n=0}^{\infty} \left(\frac{1}{5^n}\right) = \sum_{n=0}^{\infty} \left(\frac{1}{5}\right)^{n} = \frac{1}{1-\frac{1}{5}} = \frac{5}{4}.$$

Note that in this example the sum starts with $n = 0$ instead of $n = 1$ as in (1.3.12). However, it is the initial power of r in (1.3.12) that is important, not how we write the index. Hence, the sum in this example could be written equally well as

$$\sum_{n=1}^{\infty} \left(\frac{1}{5}\right)^{n-1},$$

or

$$\sum_{n=2}^{\infty} \left(\frac{1}{5}\right)^{n-2},$$

or

$$\sum_{n=100}^{\infty} \left(\frac{1}{5}\right)^{n-100},$$

as well as many other ways. The key in applying (1.3.12) is that we identify c and r so that the first term in the sum is $cr^0 = c$.

Example 1.3.5. We have

$$\sum_{n=2}^{\infty} 4(0.34)^n = \sum_{n=2}^{\infty} 4(0.34)^2(0.34)^{n-2} = \frac{4(0.34)^2}{1-0.34} = 0.7006,$$

where we have used (1.3.12) with $c = 4(0.34)^2$ and $r = 0.34$.

We are now in a position to compute the sum in (1.3.1), and hence complete our first example.

Example 1.3.6. Let P be the probability that, when tossing a fair coin repeatedly, the first head appears on an even toss. Then we have seen that

$$P = P_2 + P_4 + P_6 + \cdots = \sum_{n=1}^{\infty} P_{2n},$$

where

$$P_{2n} = \left(\frac{1}{2}\right)^{2n} = \left(\left(\frac{1}{2}\right)^2\right)^n = \left(\frac{1}{4}\right)^n$$

for $n = 1, 2, 3, \ldots$. Thus

$$P = \sum_{n=1}^{\infty}\left(\frac{1}{4}\right)^n = \sum_{n=1}^{\infty}\frac{1}{4}\left(\frac{1}{4}\right)^{n-1} = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{1}{3}.$$

Example 1.3.7. Economists often talk about the *multiplier effect* of an infusion of money into an economy which results in new spending many times greater than the original amount spent. This is a consequence of the recipients of the money spending a certain percentage of their new money, the recipients of this spending again spending a certain percentage of their gain, and so on. For example, suppose the government spends three million dollars, and suppose that at each stage the recipients spend 90% of the money they receive. Then the first recipients spend $(3)(0.9) = 2.7$ million dollars, the second recipients spend

$$(3)(0.9)(0.9) = (3)(0.9)^2 = 2.43$$

million dollars (that is, 90% of the 2.7 million spent by the first recipients), the third recipients spend

$$(3)(0.9^2)(0.9) = (3)(0.9)^3 = 2.187$$

million dollars, and so on. If we denote the total amount of spending after n transactions by S_n, then, in millions of dollars,

$$\begin{aligned} S_1 &= 3, \\ S_2 &= 3 + 3(0.9), \\ S_3 &= 3 + 3(0.9) + 3(0.9)^2, \\ S_4 &= 3 + 3(0.9) + 3(0.9)^2 + 3(0.9)^3, \end{aligned}$$

and, in general,

$$S_n = 3 + 3(0.9) + 3(0.9)^2 + \cdots + 3(0.9)^{n-1}$$

for $n = 1, 2, 3, \ldots$. Although in actuality there will be only a finite number of transactions, we can see that as n increases the total spending will approach the sum

$$S = \sum_{n=1}^{\infty} 3(0.9)^{n-1} = \frac{3}{1-0.9} = 30$$

million dollars. Thus the initial governmental expenditure of three million dollars results in approximately 30 million dollars, 10 times the initial amount, in new spending in the economy. This partially explains why deficit spending by the government in depressed times can be far more beneficial to the economy than the actual amount spent, and why such spending during other times can be highly inflationary.

Example 1.3.8. This example involves slightly more complicated probabilistic reasoning, as well as some additional algebraic simplification, before the problem is reduced to the summation of a geometric series. Suppose that a certain female animal has a 10% chance of dying during any given year of her life. Moreover, suppose the animal does not reproduce during her first year of life, but every year after has a 20% chance of successfully reproducing. What is the probability that this animal has offspring before dying?

Let P be the probability that the animal has offspring before dying and let P_n be the probability that the animal successfully reproduces for the first time in its nth year. Then

$$P = \sum_{n=1}^{\infty} P_n.$$

Note that our sum extends to infinity even though in reality it is highly unlikely that any such animal would live even to an age of 100 years. We do this because the model we are using, as with all mathematical models, is an idealization of the real situation. In this case, by assuming that a given animal of this species has a constant 10% chance of dying in any given year, we have implicitly assumed that there is no fixed upper bound to its life span. Put another way, we have assigned a positive probability to an animal's living for, as an example, 1000 years, although this probability is very small (namely, $0.9^{1000} \approx 1.748 \times 10^{-46}$) and, hence, is not actually ever going to happen.

Since we have assumed that these animals cannot reproduce in their first year of life, we have $P_1 = 0$. To compute P_2, we note that 90% of all such females will live through their first year and that 20% of these will then have offspring successfully. Hence the proportion of females that successfully reproduce for the first time in their second year is

$$P_2 = (0.9)(0.2).$$

To compute P_3 first we note that the proportion of females living until their third year will be (0.9)(0.9) (that is, 90% of the 90% who lived through their first year). Now 80% of these will not have produced offspring successfully in their second year, so the proportion of females who reach their third year without having reproduced is $(0.9)^2(0.8)$. Finally, 20% of these will have success in reproducing in their third year. Thus

$$P_3 = (0.9)^2(0.8)(0.2).$$

Similar reasoning yields

$$P_4 = (0.9)^3(0.8)^2(0.2)$$

(that is, this represents a female who has lived through three years, did not reproduce in either her second or third year, but did have offspring in her fourth year) and, in general,

$$P_n = (0.9)^{n-1}(0.8)^{n-2}(0.2)$$

for $n = 2, 3, 4, \ldots$. Hence

$$\begin{aligned} P &= \sum_{n=1}^{\infty} P_n \\ &= \sum_{n=2}^{\infty} (0.9)^{n-1}(0.8)^{n-2}(0.2) \\ &= \sum_{n=2}^{\infty} (0.2)(0.9)(0.9)^{n-2}(0.8)^{n-2} \\ &= \sum_{n=2}^{\infty} (0.18)(0.72)^{n-2} \\ &= \frac{0.18}{1-0.72} \\ &= 0.6429, \end{aligned}$$

where the answer has been rounded to four decimal places. Thus we conclude that a given female of this species has just over a 64% chance of reproducing during her lifetime.

1.3.2 The harmonic series. It may happen that a sequence $\{a_n\}$ does not have a sum even though

$$\lim_{n\to\infty} a_n = 0.$$

One important example of this behavior is provided by the sequence $\{a_n\}$ with

$$a_n = \frac{1}{n}$$

for $n = 1, 2, 3, \ldots$. The resulting infinite series with nth partial sum given by

$$s_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \tag{1.3.13}$$

is called the *harmonic series*. Since

$$s_{n+1} = 1 + \frac{1}{2} + \cdots + \frac{1}{n} + \frac{1}{n+1} = s_n + \frac{1}{n+1} > s_n, \tag{1.3.14}$$

the sequence $\{s_n\}$ is monotone increasing. Hence, by the monotone sequence theorem, $\{s_n\}$ either converges or diverges to infinity. Now

$$\begin{aligned} s_1 &= 1, \\ s_2 &= 1 + \frac{1}{2}, \\ s_4 &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} > 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} = 1 + \frac{1}{2} + \frac{1}{2} = 1 + 2\left(\frac{1}{2}\right), \\ s_8 &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &= 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 1 + 3\left(\frac{1}{2}\right), \end{aligned}$$

and

$$s_{16} = s_8 + \sum_{j=9}^{16} \frac{1}{j} > s_8 + \sum_{j=9}^{16} \frac{1}{16} > 1 + 3\left(\frac{1}{2}\right) + \frac{8}{16} = 1 + 4\left(\frac{1}{2}\right).$$

Continuing in this pattern, we can see that, in general,

$$s_{2m} > 1 + \frac{m}{2} \tag{1.3.15}$$

for any $m = 0, 1, 2, \ldots$. Thus, since $\frac{m}{2}$ may be made arbitrarily large, the sequence $\{s_n\}$ does not have an upper bound. Hence we must have

$$\lim_{n\to\infty} s_n = \infty, \tag{1.3.16}$$

and so the harmonic series does not have a sum.

Although the partial sums of the harmonic series diverge to infinity, they grow very slowly. For example, if $n = 500,000,000$, then s_n is between 20 and 21. That is,

$$20 < 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{500{,}000{,}000} < 21. \tag{1.3.17}$$

Problems 1.3

(1) Find the sum of each of the following infinite series which has a sum.

(a) $\sum_{n=1}^{\infty} \left(\frac{1}{3}\right)^{n-1}$ (b) $\sum_{n=1}^{\infty} 4(0.21)^{n-1}$

(c) $\sum_{n=1}^{\infty} \frac{2}{5^n}$ (d) $\sum_{n=0}^{\infty} \frac{2}{7^n}$

(e) $\sum_{n=1}^{\infty} 7\left(\frac{1}{3}\right)^n \left(\frac{2}{5}\right)^{n-1}$ (f) $\sum_{n=3}^{\infty} \left(\frac{2}{3}\right)^n$

(g) $\sum_{n=1}^{\infty} \left(\frac{3}{2}\right)^{n-1}$ (h) $\sum_{n=1}^{\infty} 0.99999^n$

(i) $\sum_{n=1}^{\infty} 1.00001^n$ (j) $\sum_{n=30}^{\infty} 5\left(\frac{3^n}{7^{n-1}}\right)$

(k) $\sum_{n=100}^{\infty} \left(\frac{91}{89}\right)^{n-1}$ (l) $\sum_{n=1}^{\infty} \left(\frac{\pi}{4}\right)^n$

(m) $\sum_{n=1}^{\infty} \sin(\pi n)$ (n) $\sum_{n=1}^{\infty} \cos(\pi n)$

(2) Consider the infinite series with nth partial sum

$$s_n = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \cdots + \frac{1}{n!} = \sum_{j=0}^{\infty} \frac{1}{j!}.$$

Note that, by definition, $0! = 1$.

(a) Show that $s_n < 3$ for all values of n. Hint: Note that

$$n! = (1)(2)(3)\cdots(n) \geq (1)(2)(2)\cdots(2) = 2^{n-1}$$

for $n = 1, 2, 3, \ldots$.

(b) Combine (a) with the fact that $s_{n+1} > s_n$ for all n to conclude that

$$\sum_{j=1}^{\infty} \frac{1}{j!}$$

exists and is less than 3.

(c) In fact,

$$\sum_{j=1}^{\infty} \frac{1}{j!} = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \cdots$$

is a well-known irrational number. Add up a sufficient number of terms to enable you to guess the value of the sum. How many terms did it take?

(d) How many terms are necessary to obtain a partial sum that is within 0.000001 of the sum?

(3) The sum

$$4 \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = 4\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots\right)$$

is a well-known irrational number.

(a) Add up a sufficient number of terms to enable you to guess the value of the sum. How many terms did it take?

(b) How many terms are necessary to obtain a partial sum that is within 0.01 of the sum?

(4) This problem outlines an alternative method for deriving the result of (1.3.9). Suppose $r \neq 1$. For $n = 1, 2, 3, \ldots$, let

$$s_n = 1 + r + r^2 + \cdots + r^{n-1}.$$

Show that $s_n - rs_n = 1 - r^n$ and conclude that

$$s_n = \frac{1 - r^n}{1 - r}.$$

(5) Using the model we used for the multiplier effect, find the total amount of new spending resulting from each of the following.

(a) The government spends 2 billion dollars; each recipient spends 80% of what he or she receives.

(b) The government spends 250 million dollars; each recipient spends 95% of what he or she receives.

(c) The government spends A dollars; each recipient spends $100r\%$, $0 < r < 1$, of what he or she receives.

(6) Government regulations specify that a bank may not loan 100% of its deposits; the bank must keep a certain percentage of its deposits in reserve. For example, if a bank must keep 15% of its deposits in reserve, then it may loan out \$850 from a \$1000 deposit. Typically, this \$850 will again be deposited in a bank, and that bank may loan out 85% of it. Again, this money will be deposited and 85% of it given out in loans. As this will continue indefinitely, the multiplier effect comes into play and the total amount of money in all the deposits resulting from the initial \$1000 deposit can be computed in the same manner as in our example.

(a) Compute the total amount of the deposits resulting from the initial \$1000 deposit.

(b) How would the answer in (a) change if the reserve rate was changed from 15% to 20%?

(c) How would the answer in (a) change if the reserve rate was changed from 15% to 10%?

(7) A ball is dropped from a height of 10 meters. Suppose that every time it strikes the ground, it bounces back to a height which is 75% of the height of the previous bounce. Assuming an infinite number of bounces (again, an idealized mathematical model), how far does the ball travel before it comes to rest? What would happen if it rebounded to only 25% of its initial height?

(8) Suppose the animal in our final example above could not produce offspring for its first three years of life. How would this change the probability of a female's reproducing before dying?

(9) Suppose the animal in our final example above has only an 80% chance of living through a given year. How does this change the probability of a female's reproducing before dying?

(10) Suppose a female animal of the type discussed in the final example above has a $100r\%$, $0 \le r \le 1$, chance of reproducing each year after its first year.

(a) Find the probability P of a female's reproducing before dying.

(b) Plot P as a function of r for $0 \le r \le 1$.

(c) Find the value of r for which $P = 0.5$.

(11) How many terms of the harmonic series are needed to obtain a partial sum larger than 5? How many terms are needed to obtain a partial sum larger than 10?

(12) Plot the points (n, s_n), where s_n is the nth partial sum of the harmonic series, for $n = 1, 2, 3, \ldots, 1000$. What does this show you about the rate of growth of the partial sums?

(13) The first example of this section is a particular case of the more general problem of computing probabilities associated with the waiting time for some event to occur. As another example, suppose that an electronic switch works with probability p and fails with probability $q = 1 - p$. Then, using reasoning analogous to that used in the coin tossing example, the probability that the first failure will occur on the nth use of the switch is $p^{n-1}q$, $n = 1, 2, 3, \ldots$.

(a) Can you justify this probability?

(b) The reliability of the switch is given by the function

$$R(n) = \text{probability that the switch does not fail until after the } n\text{th use}$$
$$= \sum_{j=n+1}^{\infty} p^{j-1}q.$$

Show that $R(n) = p^n$, $n = 1, 2, 3, \ldots$.

(c) Find a way to show that $R(n) = p^n$ directly without using an infinite series.

1.4 Difference equations

At this point almost all of our sequences have had explicit formulas for their terms. That is, we have looked mainly at sequences for which we could write the nth term as $a_n = f(n)$ for some known function f. For example, if

$$a_n = \frac{n+1}{n^2+3},$$

then it is an easy matter to compute explicitly, say, $a_{10} = \frac{11}{103}$ or $a_{100} = \frac{101}{10003}$. In such cases we are able to compute any given term in the sequence without reference to any other terms in the sequence. However, it is often the case in applications that we do not begin with an explicit formula for the terms of a sequence; rather, we may know only some relationship between the various terms. An equation which expresses a value of a sequence as a function of the other terms in the sequence is called a *difference equation*. In particular, an equation which expresses the value a_n of a sequence $\{a_n\}$ as a function of the term a_{n-1} is called a *first-order difference equation*. If we can find a function f such that $a_n = f(n)$, $n = 1, 2, 3, \ldots$, then we will have solved the difference equation. In this section we will consider a class of difference equations that are solvable in this sense; in the next section we will discuss an example where an explicit solution is not possible.

Example 1.4.1. Suppose a certain population of owls is growing at the rate of 2% per year. If we let x_0 represent the size of the initial population of owls and x_n the number of owls n years later, then

$$x_{n+1} = x_n + 0.02x_n = 1.02x_n \tag{1.4.1}$$

for $n = 0, 1, 2, \ldots$. That is, the number of owls in any given year is equal to the number of owls in the previous year plus 2% of the number of owls in the previous year. Equation (1.4.1) is an example of a first-order difference equation; it relates the number of owls in a given year with the number of owls in the previous year. Hence we know the value of a specific x_n once we know the value of x_{n-1}. To get the sequence started we have to know the value of x_0. For example, if initially we have a population of $x_0 = 100$ owls and we want to know what the population will be after four years, we may compute

$$\begin{aligned} x_1 &= 1.02x_0 = (1.02)(100) = 102, \\ x_2 &= 1.02x_1 = (1.02)(102) = 104.04, \\ x_3 &= 1.02x_2 = (1.02)(104.04) = 106.1208, \end{aligned}$$

and

$$x_4 = 1.02x_3 = (1.02)(106.1208) = 108.243216.$$

Thus we would expect about 108 owls in the population after four years. Note that although it is not possible to have a fractional part of an owl, it is nevertheless important to keep the fractional part in intermediary calculations.

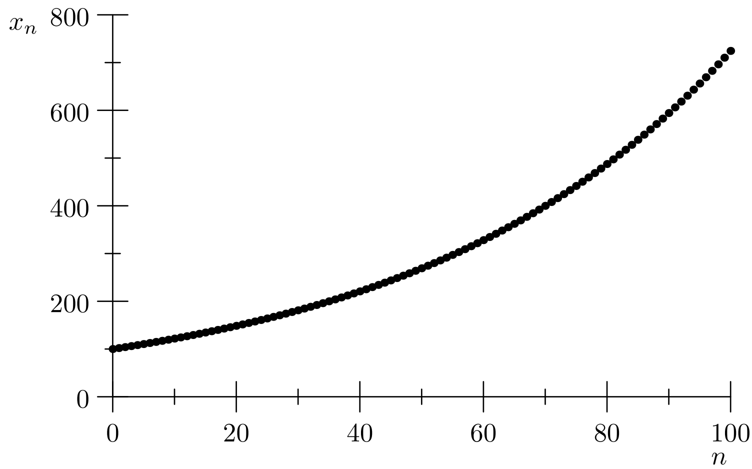

Figure 1.4.1. Plot of (n, x_n), where $x_{n+1} = 1.02x_n$

We may work backwards to find x_4 explicitly in terms of x_0:

$$\begin{aligned} x_4 &= 1.02x_3 \\ &= (1.02)(1.02)x_2 \\ &= (1.02)(1.02)(102)x_1 \\ &= (1.02)(1.02)(1.02)(1.02)x_0 \\ &= (1.02)^4 x_0. \end{aligned}$$

This is interesting because it indicates that we can compute x_4 without reference to the values of x_1, x_2, and x_3, provided, of course, that we know the value of x_0. If we do this in general, then we have solved the difference equation $x_{n+1} = 1.02x_n$. Namely, we have, for any $n = 1, 2, 3, \ldots,$

$$x_n = 1.02x_{n-1} = (1.02)^2 x_{n-2} = (1.02)^3 x_{n-3} = \cdots = (1.02)^n x_0. \tag{1.4.2}$$

For example, if $x_0 = 100$ as above, then we can compute

$$x_{20} = (1.02)^{20}(100) \approx 149,$$

or even

$$x_{150} = (1.02)^{150}(100) \approx 1{,}950,$$

without having to compute any intermediate values.

For a geometric feeling of how the population is changing with time, Figure 1.4.1 shows a plot of the points (n, x_n) for $n = 0, 1, 2, \ldots, 100$. Of course, whether or not our model will provide an accurate prediction of the owl population 100 or 200 years into the future is an entirely different question. Frequently, a simple population model like this will be valid only for a short span of time during which the rate of growth of population remains stable.

By replacing 1.02 with an arbitrary constant α in (1.4.2), we arrive at the general result that the solution of the difference equation

$$x_{n+1} = \alpha x_n, \tag{1.4.3}$$

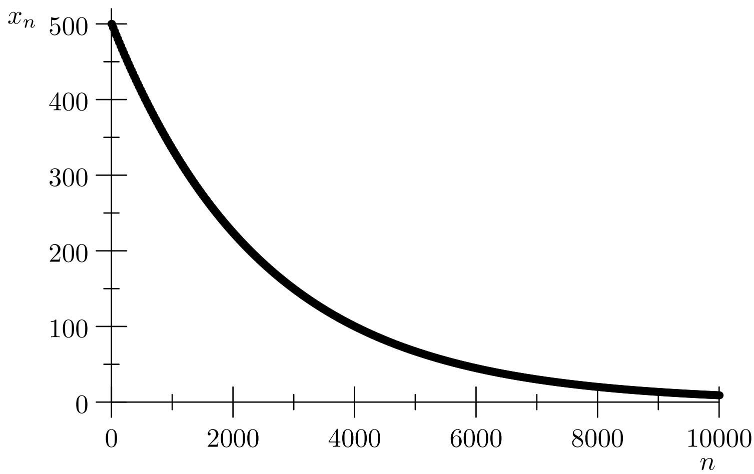

Figure 1.4.2. Amount of radium versus years

$n = 0, 1, 2, \ldots$, is given by

$$x_n = \alpha^n x_0, \tag{1.4.4}$$

$n = 0, 1, 2, \ldots$. Note that this difference equation and its solution are useful whenever we are interested in a sequence of numbers where the $(n + 1)$st term is a constant multiple of the nth term. Our first example, where a population was assumed to grow at a constant rate, is a common example of this type of behavior. Another common example is when a quantity decreases at a constant rate over time. This behavior is discussed in the next example in the context of radioactive decay.

Example 1.4.2. Radium is a radioactive element which decays at a rate of 1% every 25 years. This means that the amount left at the beginning of any given 2 year period is equal to the amount at the beginning of the previous 25-year period minus 1% of that amount. That is, if x_0 is the initial amount of radium and x_n is the amount of radium still remaining after $25n$ years, then

$$x_{n+1} = x_n - 0.01x_n = 0.99x_n \tag{1.4.5}$$

for $n = 0, 1, 2, \ldots$. Since this is a difference equation of the form of (1.4.3) with $\alpha = 0.99$, we know that the solution is of the form (1.4.4). Namely,

$$x_n = (0.99)^n x_0 \tag{1.4.6}$$

for $n = 0, 1, 2, \ldots$. For example, the amount left after 100 years is given by

$$x_4 = (0.99)^4 x_0 = 0.9606x_0,$$

where we have rounded the answer to four decimal places. That is, approximately 96% of the initial amount of radium will be left after 100 years. A plot of the amount of radium left versus number of years, assuming an initial amount of 500 grams, is given in Figure 1.4.2.

The half-life of a radioactive element is the number of years required for one-half of an initial amount to decay. Suppose that, for this example, N is the smallest integer for which x_N is less than one-half of the initial amount of radium. This would mean that

$$\frac{1}{2}x_0 \geq (0.99)^N x_0,$$

which implies that

$$\frac{1}{2} \geq (0.99)^N.$$

Taking logarithms, we have

$$\log_{10}\left(\frac{1}{2}\right) \geq \log_{10}\left((0.99)^N\right),$$

which implies that

$$\log_{10}\left(\frac{1}{2}\right) \geq N\log_{10}(0.99).$$

Solving for N, and remembering that $\log_{10}(0.99) < 0$, we have

$$N \geq \frac{\log_{10}\left(\frac{1}{2}\right)}{\log_{10}(0.99)} = 68.98,$$

rounding to two decimal places. Hence, since N must be an integer, we have $N = 69$. Recalling that we are working with 25-year units of time, this shows that the half-life of radium is approximately $(25)(69) = 1725$ years. For example, this means that if we started with an initial amount of 100 grams of radium, after 1725 years we would still have 50 grams left. It would then take an additional 1725 years until the remaining amount would be reduced to 25 grams.

Although we have stated the results of the preceding example in discrete time units, namely, units of 25 years each, later we will see that the results hold for continuous time as well. In other words, although the difference equation (1.4.5) has been set up for nonnegative integer values of n, the solution (1.4.6) is valid for arbitrary nonnegative values of n. We will hold off discussion of these ideas until we consider differential equations, the continuous time versions of difference equations, in Chapter 6.

It is interesting to compare the plots in Figures 1.4.1 and 1.4.2. The first is an example of *exponential growth*, whereas the second is an example of *exponential decay*. In the first, the steepness of the graph increases with time; in the second, the graph flattens out over time. The difference equation (1.4.3) will always lead to the first behavior when $\alpha > 1$ and to the second when $0 < \alpha < 1$.

1.4.1 First-order linear difference equations. Given constants α and β, a difference equation of the form

$$x_{n+1} = \alpha x_n + \beta, \tag{1.4.7}$$

$n = 0, 1, 2, \ldots$, is called a *first-order linear difference equation*. Note that the difference equation (1.4.3) is of this form with $\beta = 0$. A procedure analogous to the method we used to solve (1.4.3) will enable us to solve this equation as well. Namely,

$$\begin{aligned} x_n &= \alpha x_{n-1} + \beta \\ &= \alpha(\alpha x_{n-2} + \beta) + \beta \\ &= \alpha^2 x_{n-2} + \beta(\alpha + 1) \\ &= \alpha^2(\alpha x_{n-3} + \beta) + \beta(\alpha + 1) \\ &= \alpha^3 x_{n-3} + \beta(\alpha^2 + \alpha + 1) \\ &\ \ \vdots \\ &= \alpha^n x_0 + \beta(\alpha^{n-1} + \alpha^{n-2} + \cdots + \alpha^2 + \alpha + 1). \end{aligned}$$

Note that if $\alpha = 1$, this gives us

$$x_n = x_0 + n\beta, \tag{1.4.8}$$

$n = 0, 1, 2, \ldots$, as the solution of the difference equation $x_{n+1} = x_n + \beta$. If $\alpha \neq 1$, we know from Section 1.3 that

$$\alpha^{n-1} + \alpha^{n-2} + \cdots + \alpha^2 + \alpha + 1 = \frac{1 - \alpha^n}{1 - \alpha}.$$

Hence

$$x_n = \alpha^n x_0 + \beta\left(\frac{1 - \alpha^n}{1 - \alpha}\right), \tag{1.4.9}$$

$n = 0, 1, 2, \ldots$, is the solution of the first-order linear difference equation $x_{n+1} = \alpha x_n + \beta$ when $\alpha \neq 1$.

We have seen examples of first-order linear equations in the population growth and radioactive decay examples above. Another interesting example arises in modeling the change in temperature of an object placed in an environment held at some constant temperature, such as a cup of tea cooling to room temperature or a glass of lemonade warming to room temperature. If T_0 represents the initial temperature of the object, S the constant temperature of the surrounding environment, and T_n the temperature of the object after n units of time, then the change in temperature over one unit of time is given by

$$T_{n+1} - T_n = k(T_n - S), \tag{1.4.10}$$

$n = 0, 1, 2, \ldots$, where k is a constant which depends upon the object. This difference equation is known as *Newton's law of cooling*. The equation says that the change in temperature over a fixed unit of time is proportional to the difference between the temperature of the object and the temperature of the surrounding environment. That is, large temperature differences result in a faster rate of cooling (or warming) than do small temperature differences. If S is known and enough information is given to determine k, then this equation may be rewritten in the form of a first-order linear difference equation and, hence, solved explicitly. The next example shows how this may be done.

Example 1.4.3. Suppose a cup of tea, initially at a temperature of 180°F, is placed in a room which is held at a constant temperature of 80°F. Moreover, suppose that after one minute the tea has cooled to 175°F. What will the temperature be after 20 minutes?

If we let T_n be the temperature of the tea after n minutes and we let S be the temperature of the room, then we have $T_0 = 180$, $T_1 = 175$, and $S = 80$. Newton's law of cooling states that

$$T_{n+1} - T_n = k(T_n - 80), \tag{1.4.11}$$

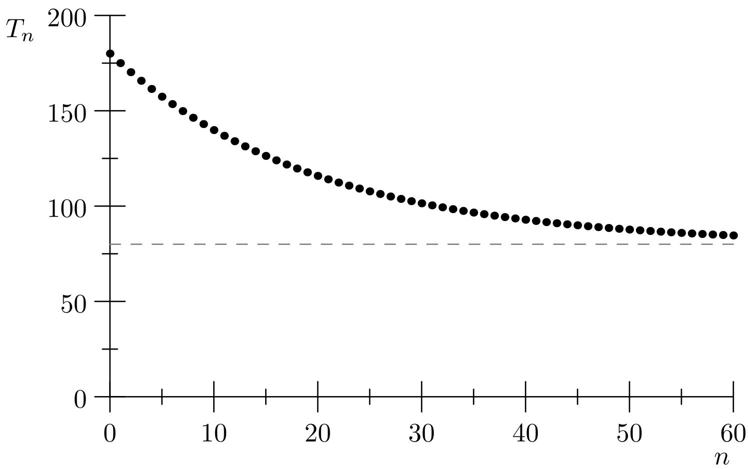

Figure 1.4.3. Temperature decreases toward room temperature

$n = 0, 1, 2, \ldots$, where k is a constant which we will have to determine. To do so, we make use of the information given about the change in the temperature of the tea during the first minute. Namely, applying (1.4.10) with $n = 0$, we must have

$$T_1 - T_0 = k(T_0 - 80).$$

That is,

$$175 - 180 = k(180 - 80).$$

Hence

$$-5 = 100k,$$

and so

$$k = -\frac{5}{100} = -0.05.$$

Thus (1.4.10) becomes

$$T_{n+1} - T_n = -0.05(T_n - 80) = -0.05T_n + 4.$$

Hence

$$T_{n+1} = T_n - 0.05T_n + 4 = 0.95T_n + 4 \tag{1.4.12}$$

for $n = 0, 1, 2, \ldots$. Now (1.4.12) is in the standard form of a first-order linear difference equation, so from (1.4.9) we know that the solution is

$$\begin{aligned} T_n &= (0.95)^n(180) + 4\left(\frac{1 - (0.95)^n}{1 - 0.95}\right) \\ &= 180(0.95)^n + 80(1 - (0.95)^n) \\ &= 80 + 100(0.95)^n \end{aligned}$$

for $n = 0, 1, 2, \ldots$. In particular,

$$T_{20} = 80 + 100(0.95)^{20} = 115.85,$$

where we have rounded the answer to two decimal places. Hence after 20 minutes the tea has cooled to just under 116°F. Also, since

$$\lim_{n\to\infty} (0.95)^n = 0,$$

we see that

$$\lim_{n\to\infty} T_n = \lim_{n\to\infty} (80 + 100(0.95)^n) = 80. \tag{1.4.13}$$

That is, as we would expect, the temperature of the tea will approach an equilibrium temperature of 80°F, the room temperature. In Figure 1.4.3 we have plotted temperature T_n versus time n for $n = 0, 1, 2, \ldots, 60$, along with the horizontal line $T = 80$. As indicated by (1.4.13), we can see that T_n decreases asymptotically toward 80°F as n increases.

Problems 1.4

(1) Compute the next five terms of each of the following sequences from the given information.

(a) $x_0 = 10$, $x_{n+1} = x_n + 4$ (b) $y_0 = -1$, $y_{n+1} = \frac{1}{y_n}$

(c) $x_0 = 40$, $x_{n+1} = 2x_n - 20$ (d) $z_0 = 2$, $z_{n+1} = z_n^2 - z_n$

(e) $x_0 = 2$, $x_1 = 3$, $x_{n+2} = x_{n+1} + x_n$ (f) $x_0 = 15$, $x_n = \frac{1}{3}x_{n-1} + 2$

(2) Solve the following difference equations with the given initial condition. Use your solution to find x_{10}.

(a) $x_{n+1} = 2x_n$, $x_0 = 5$ (b) $x_{n+1} = \frac{3}{4}x_n$, $x_0 = 100$

(c) $x_{n+1} = 1.8x_n + 10$, $x_0 = 20$ (d) $4x_{n+1} - 2x_n = 12$, $x_0 = 6$

(e) $x_{n+1} - x_n = 3x_n + 4$, $x_0 = 2$ (f) $5x_{n+1} - 3x_n = 2x_{n+1} - x_n$, $x_0 = 100$

(3) A population of weasels is growing at a rate of 3% per year. Let w_n be the number of weasels n years from now and suppose that there are currently 350 weasels.

(a) Write a difference equation which describes how the population changes from year to year.

(b) Solve the difference equation of part (a). If the population growth continues at the rate of 3%, how many weasels will there be 15 years from now?

(c) Plot w_n versus n for $n = 0, 1, 2, \ldots, 100$.

(d) How many years will it take for the population to double?

(e) Find $\lim_{n\to\infty} w_n$. What does this say about the long-term size of the population? Will this really happen?

(4) If the rate of growth of the weasel population in Problem 3 was 5% instead of 3%, how many years would it take for the population to double?

(5) Suppose that the weasel population of Problem 3 would grow at a rate of 3% a year if left to itself, but poachers kill six weasels every year for their fur.

(a) Write a difference equation which describes how the population changes from year to year.

(b) Solve the difference equation of part (a). How many weasels will there be in 15 years?

(c) Find $\lim_{n\to\infty} w_n$. What does this say about the long-term size of the population?

(d) Will the population eventually double? If so, how long will this take?

(e) Plot w_n versus n for $n = 0, 1, 2, \ldots, 100$.

(6) Suppose that the weasel population of Problem 3 would grow at a rate of 3% a year if left to itself, but poachers kill 15 weasels every year for their fur.

(a) Write a difference equation which describes how the population changes from year to year.

(b) Solve the difference equation of part (a). How many weasels will there be in 15 years?

(c) Find $\lim_{n\to\infty} w_n$. What does this say about the long-term size of the population?

(d) Will the population eventually double? If so, how long will this take?

(e) Will the population eventually die out? If so, how long will this take?

(f) Plot w_n versus n for $n = 0, 1, 2, \ldots, 100$.

(7) A radioactive element is known to decay at the rate of 2% every 20 years.

(a) If initially you had 165 grams of this element, how much would you have in 60 years?

(b) What is the half-life of this element?

(c) Suppose that the bones of a certain animal maintain a constant level of this element while the animal is living, but the element begins to decay as soon as the animal dies. If a bone of this animal is found and is determined to have only 10% of its original level of this element, how old is the bone?

(8) Repeat Problem 7 if the element decays at the rate of 3% every 10 years.

(9) A cup of coffee has an initial temperature of 165°F, but cools to 155°F in one minute when placed in a room with a temperature of 70°F. Let T_n be the temperature of the coffee after n minutes.

(a) Write a difference equation, in standard first-order linear form, which describes the change in temperature of the coffee from minute to minute.

(b) Solve the difference equation from part (a).

(c) Find the temperature of the coffee after 25 minutes.

(d) Find $\lim_{n\to\infty} T_n$.

(e) Plot T_n versus n for $n = 0, 1, 2, \ldots, 120$.

(f) Does the temperature ever reach 70°F?

(10) A glass of lemonade, initially at a temperature of 42°F, is placed in a room with a temperature of 78°F. If the lemonade warms to 45°F in 30 seconds, what will its temperature be in 10 minutes?

(11) An iron ingot, heated to a temperature of 300°C, is placed in a liquid bath held at a constant temperature of 90°C. If the ingot cools to 250°C in two minutes, what will its temperature be in 20 minutes?

(12) A glass of ginger ale is left in a room. Initially, the ginger ale has a temperature of 45°F, but after one minute the temperature has increased to 50°F and after two minutes it has increased to 54°F. What is the temperature of the room?

(13) In his book *Liber Abaci* (*Book of the Abacus*), Leonardo of Pisa, also know as Fibonacci, posed the following question: How many pairs of rabbits will be produced in a year, beginning with a single pair, if in every month each pair bears a new pair which becomes productive from the second month on? (See *A History of Mathematics* by Carl B. Boyer, Princeton University Press, 1985, page 281.)

(a) Let f_n be the number of pairs of rabbits in the nth month. Explain why $f_1 = 1$ and $f_2 = 1$.
(b) Explain why $f_{n+2} = f_{n+1} + f_n$ for $n = 1, 2, 3, \ldots$.
(c) Compute f_n for $n = 3, 4, 5, 6, 7, 8$ by hand.
(d) Compute f_n for $n = 1, 2, 3, \ldots, 100$.
(e) What is $\lim_{n\to\infty} f_n$?
(f) Compute

$$r_n = \frac{f_n}{f_{n+1}}$$

for $n = 1, 2, 3, \ldots, 100$. Do you think $\lim_{n\to\infty} r_n$ exists? If so, what is a good approximation for this limit to five decimal places?

(g) Show that

$$r_{n+1} = \frac{1}{1 + r_n}.$$

(h) Using (g) and assuming that $\lim_{n\to\infty} r_n$ exists, show that

$$\lim_{n\to\infty} r_n = \frac{\sqrt{5} - 1}{2},$$

the *golden section ratio*.

(14) Given $x_0 = 0$ and $x_{10} = 20$, show that $x_n = 2n$ satisfies the difference equation

$$x_n = \frac{x_{n-1} + x_{n+1}}{2}$$

for $n = 1, 2, 3, \ldots, 9$. This difference equation is a discrete model for the equilibrium heat distribution along a straight piece of wire running from 0 to 10 with the temperature at 0 held at 0° and the temperature at 10 held at 20°.

(15) How would the solution to Problem 14 change if we changed the boundary conditions to $x_0 = 10$ and $x_{10} = 50$?

(16) An approximate solution of a two-dimensional version of the model in Problem 14 may be found using a spreadsheet. For example, you might set cells A1–A20 and H1–H20 equal to 10 and cells B1–G1 and B20–G20 equal to 0. This would represent a flat rectangular piece of metal with the temperature along the vertical sides held fixed at 10° and the temperature along the horizontal sides held fixed at

0°. Now set the value of every cell inside the rectangle to be equal to the average of the values of its four neighboring cells. For example, you would put the formula (A2+C2+B1+B3)/4 in cell B2 and then copy this cell to all the cells in the block from B2 to G19. Now have the spreadsheet repeatedly compute the values of the cells until they stabilize (that is, until they no longer change values when you recompute). If you format the cell values so that they are all integers, this should not take too long. What you have now is the equilibrium heat distribution for the metal plate. Now try different boundary conditions to obtain different equilibrium heat distributions.

1.5 Nonlinear difference equations

In Section 1.4 we discussed the difference equation

$$x_{n+1} = \alpha x_n, \tag{1.5.1}$$

$n = 0, 1, 2, \ldots$, as a model for either growth or decay and we saw that its solution is given by

$$x_n = \alpha^n x_0,$$

$n = 0, 1, 2, \ldots$. Now

$$\lim_{n\to\infty} \alpha^n = \begin{cases} 0, & \text{if } 0 < \alpha < 1, \\ 1, & \text{if } \alpha = 1, \\ \infty, & \text{if } \alpha > 1, \end{cases} \tag{1.5.2}$$

from which it follows that if $\{x_n\}$ is a solution of (1.5.1) with $x_0 > 0$, then

$$\lim_{n\to\infty} x_n = x_0 \lim_{n\to\infty} \alpha^n = \begin{cases} 0, & \text{if } 0 < \alpha < 1, \\ x_0, & \text{if } \alpha = 1, \\ \infty, & \text{if } \alpha > 1. \end{cases} \tag{1.5.3}$$

These limiting values are consistent with our radioactive decay example since, in that case, $0 < \alpha < 1$ and we would expect the amount of a radioactive element to decline toward 0 over time. The case $0 < \alpha < 1$ also may make sense for a population model if the population is declining and heading toward extinction. However, the unbounded growth indefinitely into the future implied by the case $\alpha > 1$ is very unlikely for a population model: eventually ecological or even sociological problems come to the forefront, such as when the population begins to overreach the resources available to it, and the rate of growth of the population changes. Even for bacteria growing in a Petri dish, diminishing food and space eventually cause a change in the rate of growth. Hence the equation

$$x_{n+1} = \alpha x_n \tag{1.5.4}$$

for $n = 0, 1, 2, \ldots$ and $\alpha > 1$, called the *uninhibited*, or *natural*, *growth model*, although often accurate as a model of population growth over short periods of time, is usually too simplistic for predictions over long time spans.

1.5.1 The inhibited growth model. Suppose we wish to model the growth of a certain population which, without ecological constraints, would grow at a rate of $100\beta\%$ per unit of time. That is, if x_n represents the size of the population after n units of time and there are no constraints on the size of the population, then

$$x_{n+1} - x_n = \beta x_n \tag{1.5.5}$$

for $n = 0, 1, 2, \ldots$. However, suppose that, because of the limitation of resources, the population will begin to decline if it ever has more than M individuals. We call M the *carrying capacity* of the available resources, the maximum population which is sustainable over time. Then it would be reasonable to modify our model by forcing the amount of increase over a unit of time to decrease as the size of the population approaches M and to become negative if the size of the population ever exceeds M. One way to accomplish this is to multiply the term βx_n in (1.5.5) by

$$\frac{M - x_n}{M},$$

a ratio which is close to 1 when x_n is small, close to 0 when x_n is close to M, and negative when x_n exceeds M. This leads us to the difference equation

$$x_{n+1} - x_n = \beta x_n \left(\frac{M - x_n}{M} \right),$$

$n = 0, 1, 2, \ldots$, or, equivalently,

$$x_{n+1} = x_n + \frac{\beta}{M} x_n (M - x_n), \tag{1.5.6}$$

$n = 0, 1, 2, \ldots$, which we call the *inhibited growth model*, also known as the *discrete logistic equation*. This is an example of a nonlinear difference equation because if we multiply out the right-hand side of the equation we have a quadratic term, namely, $\frac{\beta}{M} x_n^2$. Such equations are, in general, far more difficult to solve than the linear difference equations we considered in Section 1.4; in fact, many nonlinear difference equations are not solvable in terms of the elementary functions of calculus. Hence we will not consider any methods for solving such equations, relying instead on computing specific solutions by iterating the equation using a calculator or, preferably, a computer.

Example 1.5.1. Suppose a population of owls, currently numbering 100, has a natural growth rate of 4%, but, because of the limited resources of their natural habitat, can sustain a population of no more than 500. If we let x_n represent the size of the population n years from now, then, using the inhibited growth model, we should have

$$x_{n+1} = x_n + \frac{0.04}{500} x_n (500 - x_n) = x_n + 0.00008 x_n (500 - x_n)$$

for $n = 0, 1, 2, \ldots$. Using this equation we are able to compute, for example, the predicted size of the population for the next 10 years:

Year	Population
0	100.0
1	103.2
2	106.5
3	109.8
4	113.3
5	116.8
6	120.3
7	124.0
8	127.8
9	131.5
10	135.4

Here, and in subsequent tables, we have rounded our results to the first decimal place. It is interesting to compare these results to the corresponding results for the uninhibited growth model.

If we let y_n be the predicted population n years from now using the uninhibited growth model, then we would have

$$y_{n+1} = y_n + 0.04y_n = 1.04y_n,$$

$n = 0, 1, 2, \ldots$, which has the exact solution

$$y_n = 100(1.04)^n$$

for $n = 0, 1, 2, \ldots$. From this model we obtain the following predicted population sizes:

Year	Population
0	100.0
1	104.0
2	108.2
3	112.5
4	117.0
5	121.7
6	126.5
7	131.6
8	136.9
9	142.3
10	148.0

As we would expect, the population is growing more slowly under the inhibited population growth model than under the uninhibited model. Moreover, this difference will become more pronounced over time. For example, after 150 years we would have $x_{150} = 495.4$ and $y_{150} = 35{,}892$, illustrating how the inhibited growth model is constrained by the carrying capacity of 500 while the uninhibited growth model will have unbounded growth. Figures 1.5.1 and 1.5.2 provide a graphical comparison of the two models for $n = 0, 1, 2, \ldots, 150$. Note that it appears that

$$\lim_{n\to\infty} x_n = 500,$$

while

$$\lim_{n\to\infty} y_n = \infty.$$

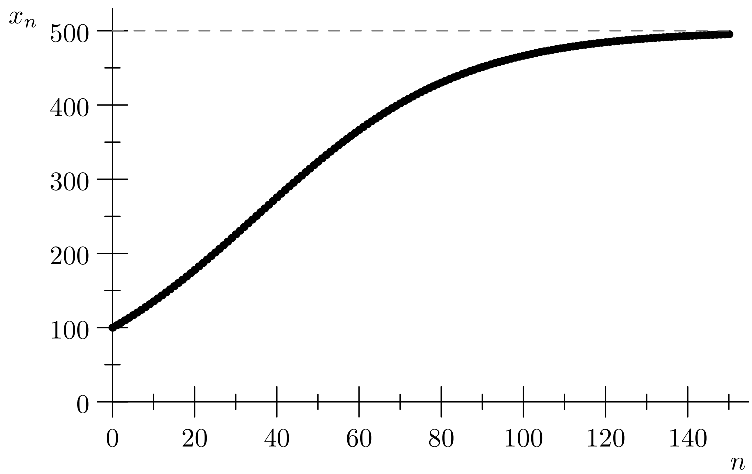

Figure 1.5.1. Inhibited growth with $\beta = 0.04$

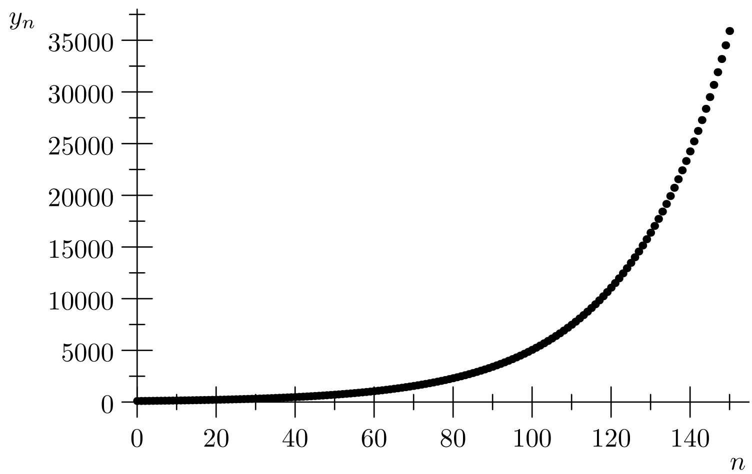

Figure 1.5.2. Uninhibited growth with $\beta = 0.04$

With the inhibited growth model, if $0 < \beta < 1$ and $x_n < M$, then

$$\beta \frac{x_n}{M} < 1,$$

so

$$x_{n+1} = x_n + \beta \frac{x_n}{M}(M - x_n) < x_n + (M - x_n) = M \tag{1.5.7}$$

for $n = 0, 1, 2, \ldots$. Thus if $0 < \beta < 1$, and we start with $x_0 < M$, then $x_n < M$ for all n. Moreover, since

$$\beta \frac{x_n}{M} > 0,$$

we have

$$x_{n+1} = x_n + \beta \frac{x_n}{M}(M - x_n) > x_n \tag{1.5.8}$$

for all n. Hence the sequence $\{x_n\}$ is monotone increasing and bounded, and so must have a limit. In Problem 8 you will be asked to verify that this limit is in fact M, as appeared to be the case in the previous example.

If $\beta > 1$, it may be the case that there are values of n for which $x_n > M$, in which case

$$\beta \frac{x_n}{M}(M - x_n) < 0$$

and, as a consequence, $x_{n+1} < x_n$.

Example 1.5.2. Suppose $x_0 = 100$ and $M = 500$ as in the previous example, but now let $\beta = 1.5$. That is,

$$x_{n+1} = x_n + \frac{1.5}{500} x_n(M - x_n) = x_n + 0.003 x_n(500 - x_n)$$

for $n = 0, 1, 2, \ldots$. This equation generates the following values:

Year	Population
0	100.0
1	220.0
2	404.8
3	520.4
4	448.5
5	505.3
6	497.2
7	501.4
8	499.3
9	500.3
10	499.8

Note how the values increase rapidly (as we should expect with such a large value for β) to above the carrying capacity of 500, but then oscillate about 500, with the oscillations diminishing in size. In fact it may be shown that it is also true in this case that

$$\lim_{n\to\infty} x_n = 500.$$

See Figure 1.5.3.

It is possible to show that, for the inhibited growth model of (1.5.6),

$$\lim_{n\to\infty} x_n = M$$

whenever $0 < \beta \leq 2$. However, there are other possible behaviors when $\beta > 2$.

Example 1.5.3. Suppose $x_0 = 100$ and $M = 500$, as in the previous examples, but now let $\beta = 2.3$. That is,

$$x_{n+1} = x_n + \frac{2.3}{500} x_n(M - x_n) = x_n + 0.0046 x_n(500 - x_n)$$

for $n = 0, 1, 2, \ldots$. This equation generates the following values:

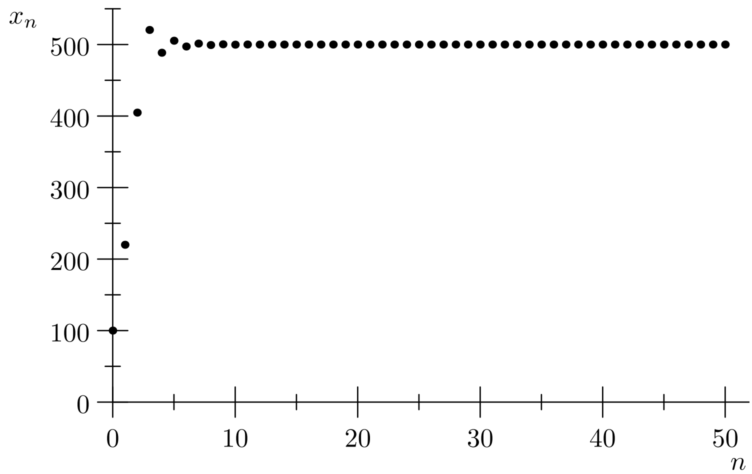

Figure 1.5.3. Inhibited growth with $\beta = 1.5$

Year	Population	Year	Population
0	100.0	11	343.8
1	284.0	12	590.8
2	566.2	13	344.0
3	393.8	14	590.9
4	586.2	15	343.9
5	353.8	16	590.8
6	591.7	17	343.9
7	342.0	18	590.8
8	590.6	19	343.9
9	344.5	20	590.8
10	590.9		

Notice that instead of approaching a single limiting value, the population is settling down to an oscillation between 344 and 591. We say that the sequence $\{x_n\}$ is approaching a *limiting cycle of period 2*, as shown in Figure 1.5.4.

It is possible to obtain limiting cycles of longer periods by increasing β. For example, Figure 1.5.5 shows the effect of letting $\beta = 2.48$. Note that $\{x_n\}$ appears to be approaching a limiting cycle of period 4.

With appropriate choices for β and x_0, it is in fact possible for the inhibited growth model to exhibit limiting cycles of any given period. This is related to the fact that it is possible for this model to behave chaotically. Intuitively, a sequence is *chaotic* if it displays erratic behavior which, although in theory completely determined by a difference equation such as (1.5.6), is in practice unpredictable because small changes in the initial value x_0 yield strikingly different sequences. For example, Figures 1.5.6 and 1.5.7 illustrate the differing behavior of the inhibited growth model with $\beta = 2.95$, first for an initial population of 100 and then for an initial population of 101.

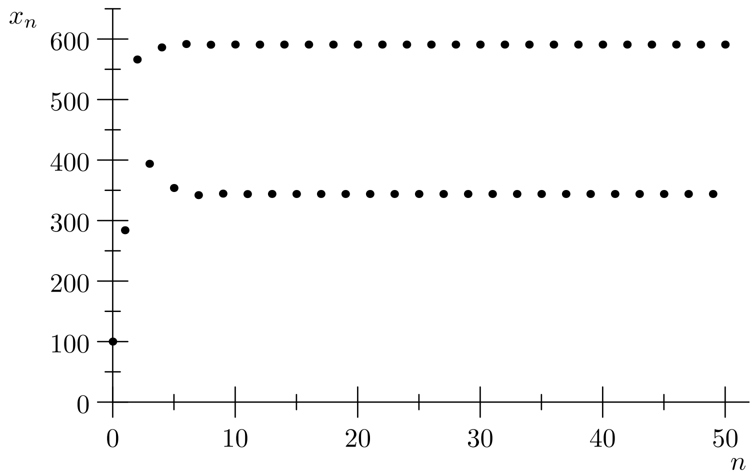

Figure 1.5.4. Inhibited growth with $\beta = 2.3$

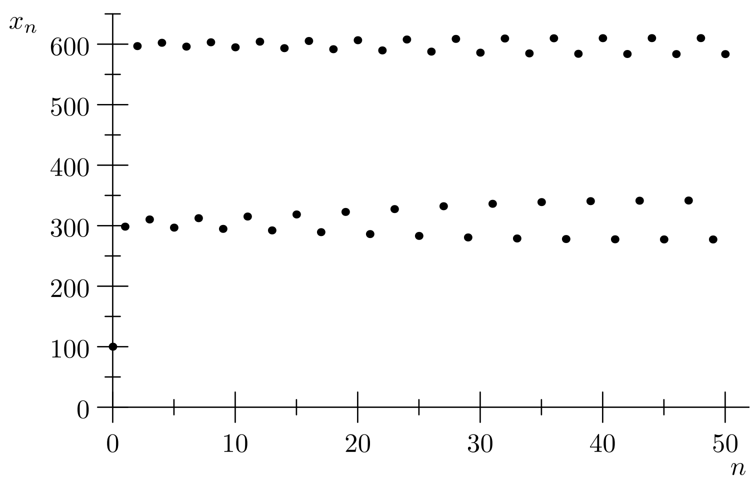

Figure 1.5.5. Inhibited growth with $\beta = 2.48$

Problems 1.5

(1) A population of weasels has a natural growth rate of 3% per year. Let w_n be the number of weasels n years from now and suppose there are currently 300 weasels.

(a) Suppose the carrying capacity of the weasel's habitat is 1000. Using an inhibited growth model, write a difference equation which describes how the population changes from year to year.

(b) Using the difference equation from (a), compute w_n for $n = 1, 2, \ldots, 150$.

(c) How many years will it take for the population to double? To triple?

(d) Plot w_n versus n for $n = 0, 1, 2, \ldots, 150$. From the plot, guess $\lim_{n\to\infty} w_n$.

(e) Compare your answers with those to Problem 3 in Section 1.4.

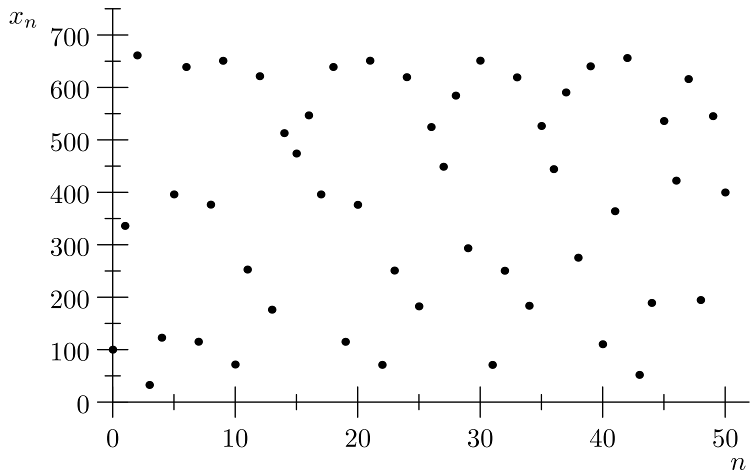

Figure 1.5.6. Inhibited growth with $\beta = 2.95$ and $x_0 = 100$

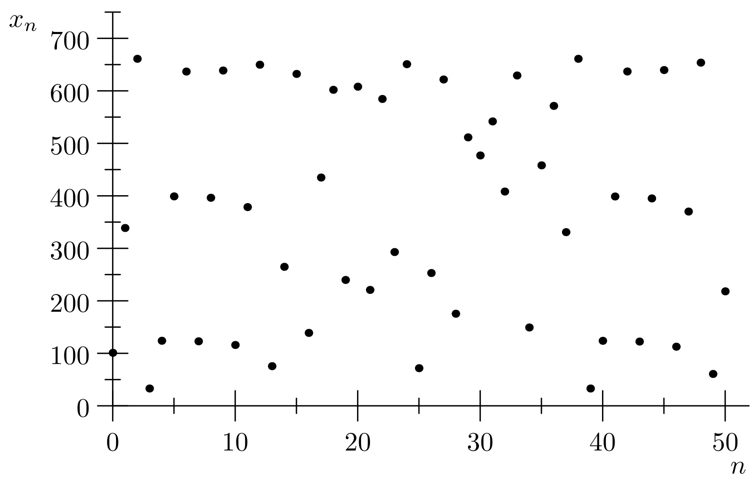

Figure 1.5.7. Inhibited growth with $\beta = 2.95$ and $x_0 = 101$

(2) Suppose a population of northern pike in a lake in Montana has a natural growth rate of 4.5% per year, but the lake can support no more than 10,000 pike. Let p_n be the number of pike n years from now and suppose $p_0 = 1000$.

(a) Use the inhibited growth model to write a difference equation which describes how the population changes from year to year.

(b) Using the difference equation from (a), compute p_n for $n = 1, 2, 3, \ldots, 50$.

(c) How many years will it take for the population to double? To triple?

(d) Plot p_n versus n for $n = 0, 1, 2, \ldots, 200$. From the plot, guess $\lim_{n\to\infty} p_n$.

(e) How many years will it take for the population to reach 9500?

(3) Do Problem 2 assuming an uninhibited growth model and no restrictions on the number of pike that the lake can support.

(4) Suppose r_n represents the number of snowshoe rabbits in a certain National Forest in Alaska after n years with an initial value of $r_0 = 5000$. Moreover, suppose the forest can support no more than 10,000 rabbits and $\{r_n\}$ satisfies the inhibited growth model

$$r_{n+1} = r_n + \frac{\beta}{10,000} r_n(10,000 - r_n)$$

for $n = 0, 1, 2, \ldots$. For each of the following values for β, plot r_n versus n for $n = 0, 1, 2, \ldots, 100$ and comment on the behavior of the sequence, in particular noting any limiting values or limiting cycles.

(a) $\beta = 0.5$ (b) $\beta = 1.5$

(c) $\beta = 2.4$ (d) $\beta = 2.5$

(e) $\beta = 2.56$ (f) $\beta = 2.9$

(5) Using an initial value of $x_0 = 0.5$, let $\{x_n\}$ be the sequence which satisfies the difference equation

$$x_{n+1} = \mu x_n(1 - x_n),$$

$n = 0, 1, 2, \ldots$. Plot x_n versus n for the following values of μ and comment on the behavior of the sequence, in particular noting any limiting values or limiting cycles.

(a) $\mu = 0.9$ (b) $\mu = 1.0$

(c) $\mu = 1.5$ (d) $\mu = 2.0$

(e) $\mu = 2.5$ (f) $\mu = 3.0$

(g) $\mu = 3.1$ (h) $\mu = 3.5$

(i) $\mu = 3.57$ (j) $\mu = 1 + \sqrt{8}$

(k) $\mu = 3.99$ (l) $\mu = 4.0$

(6) Repeat Problem 5 starting with an initial value of $x_0 = 0.6$.

(7) If f is any function defined for real numbers, then the difference equation

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

$n = 0, 1, 2, \ldots$, is called a *discrete dynamical system*. For any given initial condition x_0, the sequence $\{x_n\}$ which satisfies this equation is called an orbit of f. Note that an orbit of f is simply the sequence of points

$$x_0, f(x_0), f(f(x_0)), f(f(f(x_0))), \ldots.$$

For example, the difference equation in Problem 5 is an example of a discrete dynamical system with $f(x) = \mu x(1 - x)$. For each of the following, compute 50 terms of the given orbit and discuss its behavior.

(a) $x_0 = 10$, $f(x) = 2x$ (b) $x_0 = 100$, $f(x) = 0.8x$

(c) $x_0 = 2$, $f(x) = \cos(x)$ (d) $x_0 = 2$, $f(x) = \sin(x)$

(e) $x_0 = 5$, $f(x) = \frac{1}{2}\left(x + \frac{2}{x}\right)$ (f) $x_0 = 1$, $f(x) = \frac{2x^2}{3x^2-5}$

(g) $x_0 = 0$, $f(x) = x^2 + 1.0$ (h) $x_0 = 0$, $f(x) = x^2 - 0.5$

(i) $x_0 = 0$, $f(x) = x^2 - 0.8$ (j) $x_0 = 0$, $f(x) = x^2 - 1.0$

(k) $x_0 = 0$, $f(x) = x^2 - 1.9$ (l) $x_0 = 0$, $f(x) = x^2 - 2.0$

(8) Assuming that the sequence $\{x_n\}$ satisfying the inhibited growth model equation

$$x_{n+1} = x_n + \frac{\beta}{M} x_n (M - x_n)$$

has a limit, show that $\lim_{n\to\infty} x_n = M$.