Chapter 2
Functions and Continuity

Abstract Calculus is the study of the rate of change and the total accumulation of processes described by functions. In this chapter we review some familiar notions of function and explore functions that are defined by sequences of functions.

2.1 The Notion of a Function

The idea of a function is the most important concept in mathematics. There are many sources of functions, and they carry information of a special kind. Some are based on observations, like the maximum daily temperature T in your town every day last year:

$$\text{temperature} = T(\text{day}).$$

Some express a causal relation between two quantities, such as the force f exerted by a spring as a function of the displacement:

$$\text{force} = f(\text{displacement}).$$

Research in science is motivated by finding functions to express such causal relations. A function might also express a purely arbitrary relationship like

$$F = \frac{9}{5}C + 32,$$

relating the Fahrenheit temperature scale to the Celsius scale. Or it could express a mathematical theorem:

$$r = -\frac{b}{2} + \frac{\sqrt{b^2-4}}{2},$$

where r is the larger root of the quadratic equation $x^2 + bx + 1 = 0$.

Functions can be represented in different ways. Some of these ways are familiar to you: graphs, tables, and equations. Other methods, such as representing a function

P.D. Lax and M.S. Terrell, *Calculus With Applications*, Undergraduate Texts in Mathematics, DOI 10.1007/978-1-4614-7946-8_2,

through a sequence of functions, or as a solution to a differential equation, are made possible by calculus.

Rather than starting with a definition of function, we shall first give a number of examples and then fit the definition to these.

Example 2.1. The vertical distance h (measured in kilometers) traveled by a rocket depends on the time t (measured in seconds) that has elapsed since the rocket was launched. Figure 2.1 graphically describes the relation between t and h.

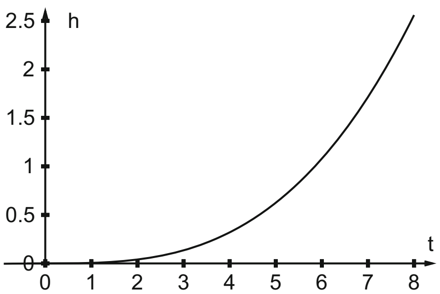

Fig. 2.1 Vertical distance traveled by a rocket. The horizontal axis gives the time elapsed since launch, in seconds. The vertical gives distance traveled, in kilometers

Example 2.2. The graph in Fig. 2.2 shows three related functions: U.S. consumption of oil, the price of oil unadjusted for inflation (the composite price), and the price of oil adjusted for inflation (in 2008 dollars).

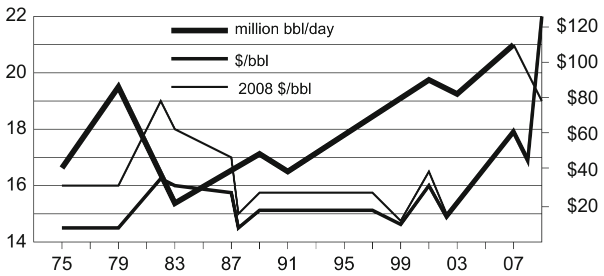

Fig. 2.2 Oil consumption, price, and inflation adjusted price

Example 2.3. The distance d traveled by a body falling freely from rest near the surface of the Earth, measured in meters, and t the time of fall measured in seconds.

$$d = 4.9t^2.$$

Example 2.4. The national debt D in billions of dollars in year y.

y	2004	2005	2006	2007	2008	2009	2010
D	7,354	7,905	8,451	8,951	9,654	10,413	13,954

In contrast, this is the table that appeared in the first edition of this book.

y	1955	1956	1957	1958	1959	1960	1961
D	76	82	85	90	98	103	105

Example 2.5. The volume V of a cube with edge length s is $V = s^3$.

Adjusted gross income	Tax rate (%)
0–8,375	10
8,375–34,000	15
34,000–82,400	25
82,400–171,850	28
171,850–373,650	33
373,650 and above	35

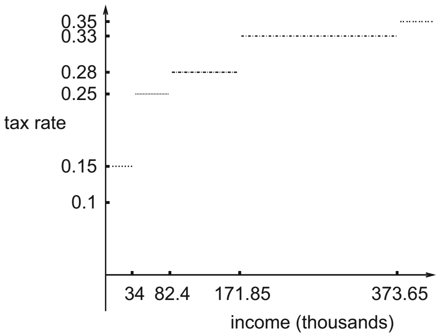

Fig. 2.3 *Left*: a table of tax rates in Example 2.6. *Right*: a graph of the tax rate by income level

Example 2.6. The Internal Revenue Service's 2010 tax rates for single-filing status are given in the table in Fig. 2.3. The tax rates can be described as a function. Let x be adjusted gross income in dollars, and $f(x)$ the rate at which each dollar within that income level is taxed. The function f can be described by the following rule

$$f(x) = \begin{cases} 0.10 & \text{for } 0 \le x \le 8{,}375 \\ 0.15 & \text{for } 8{,}375 < x \le 34{,}000 \\ 0.25 & \text{for } 34{,}000 < x \le 82{,}400 \\ 0.28 & \text{for } 82{,}400 < x \le 171{,}850 \\ 0.33 & \text{for } 171{,}850 < x \le 373{,}650 \\ 0.35 & \text{for } 373{,}650 < x. \end{cases}$$

The graph in Fig. 2.3 makes the jumps in tax rate and the levels of income subject to those rates easier to see. (To compute the tax on 10,000 dollars, for example, the first 8,375 is taxed at 10 %, and the next $10{,}000 - 8{,}375 = 1{,}625$ is taxed at 15 %. So the tax is $(0.15)(1{,}625) + (0.10)(8{,}375) = 1{,}121.25$.)

Fig. 2.4 A function can be thought of as a device in a box, with input and output

We can also think of a function as a box, as in Fig. 2.4. You drop in an input x, and out comes $f(x)$ as the output.

Definition 2.1. A *function* f is a rule that assigns to every number x in a collection D, a number $f(x)$. The set D is called the *domain* of the function, and $f(x)$ is called the *value* of the function at x. The set of all values of a function is called its *range*. The set of ordered pairs $(x, f(x))$ is called the *graph* of f.

When we describe a function by a rule, we assume, unless told otherwise, that the set of inputs is the largest set of numbers for which the rule makes sense. For example, take

$$f(x) = x^2 + 3, \qquad g(x) = \sqrt{x-1}, \qquad h(x) = \frac{1}{x^2-1}.$$

The domain of f is all numbers. The domain of g is $x \geq 1$, and the domain of h is any number other than 1 or -1.

2.1a Bounded Functions

Definition 2.2. We say that a function f is *bounded* if there is a positive number m such that for all values of f, $-m \leq f(x) \leq m$. We say that a function g is *bounded away from* 0 if there is a positive number p such that no value of g falls in the interval from $-p$ to p.

In Fig. 2.5, f is bounded because $-m \leq f(x) \leq m$ for all x, and g is bounded away from 0 because $0 < p \leq g(x)$ for all x.

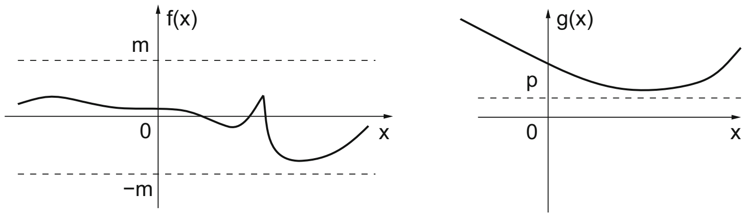

Fig. 2.5 *Left*: f is a bounded function. *Right*: g is bounded away from 0

A function that is not bounded, or is not bounded away from 0, may have one or both of those properties on a subset of its domain.

Example 2.7. Let $h(x) = \dfrac{1}{x^2-1}$. Then h is not bounded. It has arbitrarily large values (both positive and negative) as x tends to 1 or -1. Furthermore, h is not bounded away from 0, because $h(x)$ tends to 0 as x becomes arbitrarily large (positive or negative). However, if we restrict the domain of h to, say, the interval $[-0.8, 0.8]$, then h is both bounded and bounded away from 0 on $[-0.8, 0.8]$. See Fig. 2.6 for the graph of h.

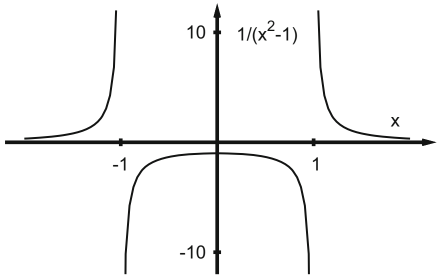

Fig. 2.6 The function $h(x) = \frac{1}{x^2-1}$ is neither bounded nor bounded away from 0

2.1b Arithmetic of Functions

Once you have functions, you can use them to make new functions. The sum of functions f and g is denoted by $f+g$, and the difference by $f-g$:

$$(f+g)(x) = f(x) + g(x), \qquad (f-g)(x) = f(x) - g(x).$$

The product and quotient of functions f and g are denoted by fg and $\dfrac{f}{g}$:

$$(fg)(x) = f(x)g(x), \qquad \frac{f}{g}(x) = \frac{f(x)}{g(x)} \text{ when } g(x) \neq 0.$$

In applications, it makes sense to add or subtract two functions only if their values are measured in the same units. In our example about oil consumption and price, it makes sense to find the difference between the inflation-adjusted price and the nonadjusted price of oil. However, it does not make sense to subtract the price of oil from the number of barrels consumed.

Polynomials. Starting with the simplest functions, the constant functions

$$c(x) = c$$

and the identity function

$$i(x) = x,$$

we can build more complicated functions by forming sums and products. All functions that are obtained from constant and identity functions through repeated additions and multiplications are of the form

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0,$$

where the a's are constants and n is a positive integer. Such a function is called a *polynomial*. A quotient of two polynomials, $\frac{p(x)}{q(x)}$, is called a *rational* function.

Linear Functions. A simple but highly important class of functions is that of *linear* functions. In Chap. 3, we show how to use linear functions to approximate other functions. Every linear function ℓ is of the form

$$\ell(x) = mx + b,$$

where m and b are some given numbers. A linear function is certainly simple from a computational point of view: to evaluate it, we need to perform one multiplication and one addition. Linear functions have the property that

$$\ell(x+h) = m(x+h) + b = \ell(x) + mh,$$

which means that when the input x in increased by h, the output changes by an amount that does not depend on x. The change in the output of a linear function is m times the change in the input (Fig. 2.7),

$$\ell(x+h) - \ell(x) = mh.$$

Example 2.8. In changing temperature from Celsius to Fahrenheit, we use the formula

$$F = \frac{9}{5}C + 32.$$

A change in temperature in degrees Celsius produces a change in temperature in degrees Fahrenheit that is always $\frac{9}{5}$ as large, independent of the temperature.

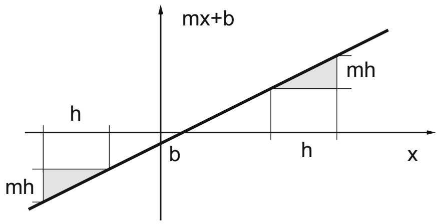

Fig. 2.7 The graph of a linear function $\ell(x) = mx + b$. The change in the output is m times the change in the input

You can completely determine a linear function if you know the function values at two different points. Suppose

$$y_1 = \ell(x_1) = mx_1 + b \quad \text{and } y_2 = \ell(x_2) = mx_2 + b.$$

By subtracting, we see that $y_2 - y_1 = m(x_2 - x_1)$. Solving for m, we get

$$m = \frac{y_2 - y_1}{x_2 - x_1}.$$

The number m is called the *slope* of the line through the points (x_1, y_1) and (x_2, y_2). Then b is also determined by the two points, because

$$b = y_1 - \frac{y_2 - y_1}{x_2 - x_1} x_1 \qquad (x_1 \neq x_2).$$

In addition to visualizing a linear function graphically, we can look at how numbers in the domain are mapped to numbers in the range. In this representation, m can be interpreted as a stretching factor.

Example 2.9. Figure 2.8 shows how the linear function $\ell(x) = 3x - 1$ maps the interval $[0, 2]$ onto the interval $[-1, 5]$, which is three times as long.

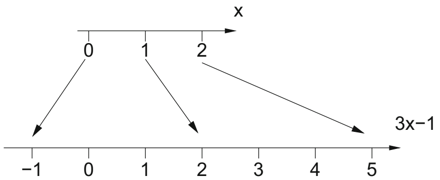

Fig. 2.8 The linear function $\ell(x) = 3x - 1$ as a mapping from $[0, 2]$ to $[-1, 5]$

There are many more important examples of functions to explore. We invite you to work on some in the Problems.

Problems

2.1. For each of these functions, is f bounded? is f bounded away from zero?

(a) $f(x) = x - \frac{1}{x} + 25$
(b) $f(x) = x^2 + 1$
(c) $f(x) = \frac{1}{x^2+1}$
(d) $f(x) = x^2 - 1$

2.2. Plot the national debt as given in Example 2.4 for the years 1955–1961. Is the national debt a linear function of time? Explain.

2.3. Let

$$f(x) = \frac{x^3 - 9x}{x^2 + 3x}, \quad g(x) = \frac{x^2 - 9}{x + 3}, \quad \text{and} \;\; h(x) = x - 3.$$

(a) Show that

$$f(x) = g(x) = h(x) \qquad \text{when } x \neq 0, -3.$$

(b) Find the domains of f, g, and h.
(c) Sketch the graphs of f, g, and h.

2.4. Let $h(x) = \frac{1}{x^2 - 1}$ with domain $[-0.8, 0.8]$. Find bounds p and q on the range of h:

$$p \leq \frac{1}{x^2 - 1} \leq q.$$

2.5. Use the tax table or graph in Example 2.6 to find the total tax on an adjusted gross income of \$200,000.

2.6. The gravitational force between masses M and m with centers separated by distance r is, according to Newton's law,

$$f(r) = \frac{GMm}{r^2}.$$

The value of G depends on the units in which we measure mass, distance, and force. Take the domain to be $r > 0$. Is f rational? bounded? bounded away from 0?

2.7. Here is a less obvious example of a linear function. Imagine putting a rope around the Earth. Make it nice and snug. Now add 20 m to the length of the rope and arrange it concentrically around the Earth. Could you walk under it without hitting your head?

2.2 Continuity

In this section, we scrutinize the definition of function given in the previous section. According to that definition, a function f assigns a value $f(x)$ to each number x in the domain of f. Clearly, in order to find the value of $f(x)$, we have to know x. But what does knowing x mean? According to Chap. 1, we know x if we are able to produce as close an approximation to x as requested. This means that we never (or hardly ever) know x *exactly*. How then can we hope to determine $f(x)$? A way out of this dilemma is to remember that knowing $f(x)$ means being able to give as close an approximation to $f(x)$ as requested. So we can determine $f(x)$ *if approximate knowledge of x is sufficient for approximate determination of $f(x)$*. The notion of continuity captures this property of a function.

Definition 2.3. We say that a function f is *continuous* at c when:
for any tolerance $\varepsilon > 0$, there is a precision $\delta > 0$ such that $f(x)$ differs from $f(c)$ by less than ε whenever x differs from c by less than δ (Fig. 2.9).

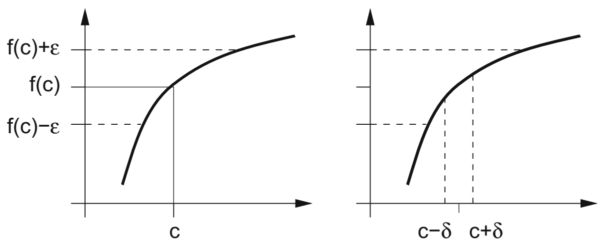

Fig. 2.9 *Left*: for any $\varepsilon > 0$, *Right*: we can find a $\delta > 0$

As a practical matter, f is continuous at c if all the values of f at points near c are very nearly $f(c)$. This leads to a useful observation about continuity: If f is continuous at c and $f(c) < m$, then it is also true that $f(x) < m$ for every x in some sufficiently small interval around c. To see this, take ε to be the distance between $f(c)$ and m, as in Fig. 2.10. Similarly, if $f(c) > m$, there is an entire interval of numbers x around c where $f(x) > m$.

Driver: "But officer, I only hit 90 mph for one instant!"
Officer: "Then you went more than 89 for an entire interval of time!"

Example 2.10. A constant function $f(x) = k$ is continuous at every point c in its domain. Approximate knowledge of c is sufficient for approximate knowledge of $f(c)$ because *all* inputs have the same output, k. As you can see in Fig. 2.11, for every x in the domain, $f(x)$ falls within ε of $f(c)$. No function can be more continuous than that!

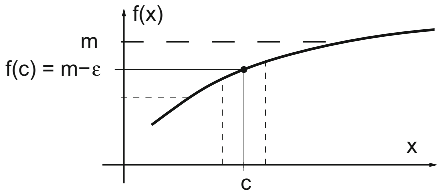

Fig. 2.10 If f is continuous, there will be an entire interval around c in which f is less than m

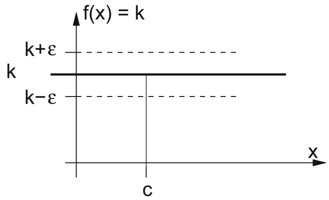

Fig. 2.11 A constant function

Example 2.11. The identity function $f(x) = x$ is continuous at every point c. Because $f(c) = c$, it is clear that approximate knowledge of c is sufficient to determine approximate knowledge of $f(c)$! Figure 2.12 shows that the definition for continuity is satisfied by letting $\delta = \varepsilon$.

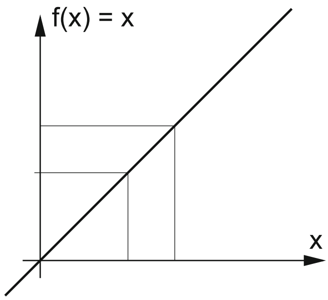

Fig. 2.12 For $f(x) = x$, $\delta = \varepsilon$ will do

A function can be continuous at some points in its domain but not at others.

Example 2.12. The graph of f in Fig. 2.3 shows the IRS 2010 tax rates for single filers. The rate is constant near 82,000. Small changes in income do not change the tax rate near 82,000. Thus f is continuous at 82,000. However, at 82,400, the situation is very different. Knowing that one's income is approximately 82,400 is not sufficient knowledge to determine the tax rate. Near 82,400, small changes in income result in very different tax rates. This is exactly the kind of outcome that continuity prohibits.

Inequalities and absolute values can be used to rewrite the definition of continuity at a point:

Restated definition. We say that a function f is *continuous at* c when: for any tolerance $\varepsilon > 0$, there is a precision $\delta > 0$ such that

$$|f(x) - f(c)| < \varepsilon$$

whenever

$$|x - c| < \delta.$$

The precision δ depends on the tolerance ε.

2.2a Continuity at a Point Using Limits

The concept of the limit of a function gives another way to define continuity at a point.

Definition 2.4. The *limit* of a function $f(x)$ as x tends to c is L,

$$\lim_{x \to c} f(x) = L,$$

when:
for any tolerance $\varepsilon > 0$, there is a precision $\delta > 0$ such that $f(x)$ differs from L by less than ε whenever x differs from c by less than δ, $x \neq c$.

By comparing the definitions of limit as x tends to c and continuity at c, we find a new way to define continuity of f at c.

Alternative definition. We say that a function f is *continuous at* c when:

$$\lim_{x \to c} f(x) = f(c).$$

If f is not continuous at c, we say that f is discontinuous at c.

The limit of $f(x)$ as x tends to c can be completely described in terms of the limits of sequences of numbers. In fact, in evaluating $\lim_{x \to c} f(x)$, we often take a sequence of numbers $x_1, x_2, \ldots, x_n, \ldots$ that tend to c and we see whether the sequence

$f(x_1), f(x_2), \ldots, f(x_n), \ldots$ tends to some number L. In order for $\lim_{x\to c} f(x)$ to exist, we need to know that all sequences $\{x_i\}$ that tend to c result in sequences $\{f(x_i)\}$ that tend to L. In Problem 2.11, we ask you to explore the connection between the limit of a function at a point and limits of sequences of numbers. It will help you see why the next two theorems follow from the laws of arithmetic and the squeeze theorem, Theorems 1.6 and 1.7, for convergent sequences.

Theorem 2.1. *If* $\lim_{x\to c} f(x) = L_1$, $\lim_{x\to c} g(x) = L_2$, *and* $\lim_{x\to c} h(x) = L_3 \neq 0$, *then*

(a) $\lim_{x\to c} \big(f(x)+g(x)\big) = L_1 + L_2$,
(b) $\lim_{x\to c} \big(f(x)g(x)\big) = L_1 L_2$, *and*
(c) $\lim_{x\to c} \dfrac{f(x)}{h(x)} = \dfrac{L_1}{L_3}$.

Theorem 2.2. Squeeze theorem. *If*

$$f(x) \le g(x) \le h(x)$$

for all x in an open interval containing c, except possibly at $x = c$, and if $\lim_{x\to c} f(x) = \lim_{x\to c} h(x) = L$, *then* $\lim_{x\to c} g(x) = L$.

Combining Theorem 2.1 and the limit definition of continuity, one can prove the next theorem, as we ask you to do in Problem 2.12.

Theorem 2.3. *Suppose f, g, and h are continuous at c, and $h(c) \neq 0$. Then $f+g$, fg, and $\dfrac{f}{h}$ are continuous at c.*

We have noted before that any constant function, and the identity function, are continuous at each point c. According to Theorem 2.3, products and sums built from these functions are continuous at each c. Every polynomial

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$$

can be constructed by taking sums and products of functions that are continuous at c. This shows that polynomials are continuous at each c. It also follows from the theorem that a rational function $\dfrac{p(x)}{q(x)}$ is continuous at each number c for which $q(c) \neq 0$.

Example 2.13. Examples 2.10 and 2.11 explain why the constant function 3 and the function x are continuous. So according to Theorem 2.3, the rational function $f(x) = x^2 - \frac{1}{x} - 3 = \frac{x^3 - 1 - 3x}{x}$ is continuous at every point except 0.

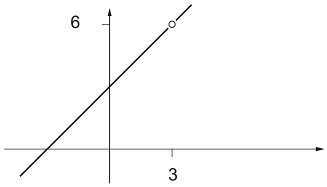

Fig. 2.13 The graph of $f(x) = \frac{x^2-9}{x-3}$

Sometimes a function is undefined at a point c, but the limit of $f(x)$ as x tends to c exists. For example, let

$$f(x) = \frac{x^2 - 9}{x - 3}.$$

Then f is not defined at 3. Notice, however, that

$$\text{for } x \neq 3, \qquad f(x) = \frac{x^2 - 9}{x - 3} = x + 3.$$

The graph of f looks like a straight line with a small hole at the point $x = 3$. (See Fig. 2.13.) The functions $\frac{x^2 - 9}{x - 3}$ and $x + 3$ are quite different at $x = 3$, but they are equal when $x \neq 3$. This means that their limits are the same as x tends to 3:

$$\lim_{x \to 3} \frac{x^2 - 9}{x - 3} = \lim_{x \to 3} (x + 3) = 6.$$

Example 2.14. Let $d(x)$ be defined as follows:

$$d(x) = \begin{cases} x & \text{for } x \leq 1, \\ x - 2 & \text{for } 1 < x. \end{cases}$$

Then d is not continuous at $x = 1$, because $d(1)$ equals 1, yet for x greater than 1 and no matter how close to 1, $d(x)$ is negative. A negative number is not close to 1. See Fig. 2.14.

It is useful to have a way to describe the behavior of $f(x)$ as x approaches c from one side or the other. If $f(x)$ tends to L as x approaches c from the right, $c < x$, we say that the *right-hand limit* of f at c is L, and write

$$\lim_{x \to c+} f(x) = L.$$

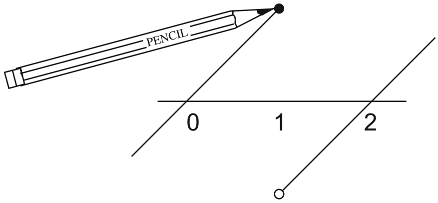

Fig. 2.14 The function $d(x)$ in Example 2.14 is not continuous at $x = 1$

If $f(x)$ tends to L as x approaches c from the left, $x < c$, we say that the *left-hand limit* of f at c is L, and write

$$\lim_{x \to c-} f(x) = L.$$

If $f(x)$ becomes arbitrarily large and positive as x tends to c, we write

$$\lim_{x \to c} f(x) = \infty,$$

and say that $f(x)$ tends to infinity as x tends to c. If $f(x)$ becomes arbitrarily large and negative as x tends to c, we write

$$\lim_{x \to c} f(x) = -\infty,$$

and say that $f(x)$ tends to minus infinity as x tends to c. Neither of these limits exists, but we use the notation to describe the behavior of the function near c. We also use the one-sided versions of these notations, as in Example 2.15.

Example 2.15. Let $f(x) = \dfrac{1}{x}$ for $x \neq 0$. Then $\lim_{x \to 0-} f(x) = -\infty$, $\lim_{x \to 0+} f(x) = \infty$.

It is also useful to have a way to describe one-sided continuity. If $\lim\limits_{x \to c-} f(x) = f(c)$, we say that f is *left continuous* at c. If $\lim\limits_{x \to c+} f(x) = f(c)$, we say that f is *right continuous* at c.

Example 2.16. The function d in Example 2.14 (see Fig. 2.14) is left continuous at 1, and not right continuous at 1:

$$\lim_{x \to 1-} d(x) = 1 = d(1), \qquad \lim_{x \to 1+} d(x) = -1 \neq d(1).$$

Left and right continuity give us a way to describe continuity on an interval that includes endpoints. For example, we say that f is *continuous on* $[a,b]$ if f is continuous at each c in (a,b) as well as right continuous at a, and left continuous at b.

2.2b Continuity on an Interval

Now we return to the question we considered at the start of this section: Is approximate knowledge of x sufficient for approximate knowledge of $f(x)$? We have seen that functions can be continuous at some points and not at others. The most interesting functions are the ones that are continuous at every point on *an interval* where they are defined.

Example 2.17. Let us analyze the continuity of the function $f(x) = x^2$ on the interval $[2,4]$. Let c be any point of this interval; how close must x be to c in order for $f(x)$ to differ from $f(c)$ by less than ε? Recall the identity

$$x^2 - c^2 = (x+c)(x-c).$$

On the left, we have the difference $f(x) - f(c)$ of two values of f. Since both x and c are between 2 and 4, we have $(x+c) \le 8$. It follows that

$$|f(x) - f(c)| = |x+c||x-c| \le 8|x-c|.$$

If we want x^2 to be within ε of c^2, it suffices to take x within $\frac{\varepsilon}{8}$ of c. That is, take $\delta = \frac{\varepsilon}{8}$ or less. This proves the continuity of f on $[2,4]$.

Example 2.18. In Chap. 1, we defined the number e through a sequence of approximations. Our intuition and experience tell us that we should get as good an approximation to e^2 as we desire by squaring a number that is close enough to e. But we do not need to rely on our intuition. Since e is between 2 and 4, the previous example shows that $f(x) = x^2$ is continuous at e. This means that if we want x^2 to be within $\varepsilon = \frac{1}{10^4}$ of e^2, it should suffice to take x within $\delta = \frac{\varepsilon}{8} = \frac{1}{8(10^4)}$ of e; in particular, $\delta < \frac{1}{10^5}$ should suffice. The list below shows squares of successively better decimal approximations to e. It confirms computationally what we proved theoretically.

$$(2.7)^2 = 7.29$$
$$(2.71)^2 = 7.3441$$
$$(2.718)^2 = 7.387524$$
$$(2.7182)^2 = 7.38861124$$
$$(2.71828)^2 = 7.3890461584$$
$$(2.718281)^2 = 7.389051594961$$

Uniform Continuity. In Example 2.17, we showed that the difference between the squares of two numbers in $[2,4]$ will be within ε as long as the two numbers are within $\frac{\varepsilon}{8}$ of each other, no matter which two numbers in $[2,4]$ we are dealing with. Here is the general notion.

Definition 2.5. A function f is called *uniformly continuous* on an interval I if given any tolerance $\varepsilon > 0$, there is a precision $\delta > 0$ such that if x and z are in I and differ by less than δ, then $f(x)$ and $f(z)$ differ by less than ε.

Clearly, a function that is uniformly continuous on an interval is continuous at every point of that interval. It is a surprising mathematical fact that conversely, a function that is continuous at every point of a *closed* interval is uniformly continuous on that interval. We outline the proof of this theorem, Theorem 2.4, in Problem 2.21.

Uniform continuity is a basic notion of calculus.

Theorem 2.4. *If a function f is continuous on $[a,b]$, then f is uniformly continuous on $[a,b]$.*

On a practical level, uniform continuity is a very helpful property for a function to have. When we evaluate a function with a calculator or computer, we round off the inputs, and we obtain outputs that are approximate. If f is uniformly continuous on $[a,b]$, then once we set a tolerance for the output, we can find a single level of precision for *all* the inputs in $[a,b]$, and the approximate outputs will be within the tolerance we have set.

2.2c Extreme and Intermediate Value Theorems

Next, we state and prove two key theorems about continuous functions on a closed interval.

Theorem 2.5. The intermediate value theorem. *If f is a continuous function on a closed interval $[a,b]$, then f takes on all values between $f(a)$ and $f(b)$.*

The theorem says in a careful way that the graph of f does not skip values.

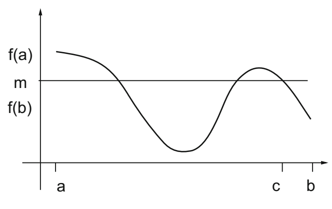

Fig. 2.15 The proof of the intermediate value theorem shows that there exists at least one number c between a and b at which $f(c) = m$

Proof. Let us take the case $f(a) > f(b)$; the opposite case can be treated analogously. Let m be any number between $f(a)$ and $f(b)$, and denote by V the set of those points x in the interval $a < x < b$ where $f(x)$ is greater than m. This set contains the point a, so it is not empty, and it is contained in $[a,b]$, so it is bounded. Denote by c the least upper bound of the set V. We claim that $f(c) = m$ (Fig. 2.15).

Suppose $f(c) < m$. Since f is continuous at c, there is a short interval to the left of c where $f(x) < m$ as well. These points x do not belong to V. And since c is an upper bound for V, no point to the right of c belongs to V. Therefore, every point of this short interval is an upper bound for V, a contradiction to c being the least upper bound.

On the other hand, suppose $f(c) > m$. Since $f(b)$ is less than m, c cannot be equal to b, and is strictly less than b. Since f is continuous at c, there is a short interval to the right of c where $f(x) > m$. But such points belong to V, so c could not be an upper bound for V.

Since according to the two arguments, $f(c)$ can be neither less nor greater than m, $f(c)$ must be equal to m. This proves the intermediate value theorem. □

Example 2.19. One use of the intermediate value theorem is in root-finding. Suppose we want to locate a solution to the equation

$$x^2 - \frac{1}{x} - 3 = 0.$$

Denote the left side by $f(x)$. With some experimentation we find that $f(1)$ is negative and $f(2)$ is positive. The function f is continuous on the interval $[1,2]$. By the intermediate value theorem, there is some number c between 1 and 2 such that $f(c) = 0$. In other words, f has a root in $[1,2]$.

Now let us bisect the interval into two subintervals, $[1,1.5]$ and $[1.5,2]$. We see that $f(1.5) = -1.416\ldots$ is negative, so f has a root in $[1.5,2]$. Bisecting again, we obtain $f(1.75) = -0.508\ldots$, which is again negative, so f has a root in $[1.75,2]$. Continuing in this manner, we can trap the root in an arbitrarily small interval.

Theorem 2.6. The extreme value theorem. *If f is a continuous function on a closed interval $[a,b]$, then f takes on both a maximum value and a minimum value at some points in $[a,b]$.*

One consequence of the extreme value theorem is that every function that is continuous on a closed interval is bounded. Although the extreme value theorem does not tell us how or where to find the bounds, it is still very useful.

Let us look at the graph of f and imagine a line parallel to the x-axis slid vertically upward until it just touches the graph of f at some last point of intersection, which is the maximum. Similarly, slide a line parallel to the x-axis vertically downward. The last point of intersection with the graph of f is the minimum value of f (Fig. 2.16).

We supplant now this intuitive argument by a mathematical proof of the existence of a maximum. The argument for a minimum is analogous.

Proof. Divide the interval $[a,b]$ into two closed subintervals of equal length. We compare the values of f on these two subintervals. It could be the case that there is a point on the first subinterval where the value of f is greater than at any point on the second subinterval. If there is no such point, then it must be the case that for every point x in the first subinterval there is a point z in the second subinterval where the value of f is at least as large as the value of f at x.

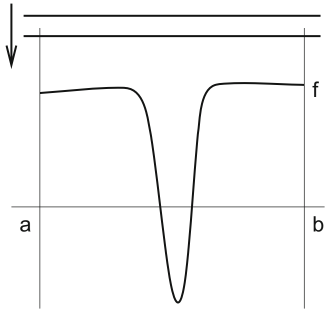

Fig. 2.16 A *horizontal line* moves down, seeking the minimum of a continuous function. The extreme value theorem guarantees that there is a last value where you can stop the moving line, keeping it in contact with the graph

In the first case, we choose the first subinterval, and in the second case, the second subinterval, and denote the chosen subinterval by I_1.

Key property of I_1: for every point x in $[a,b]$ but not in I_1, there is a point in I_1 where f is at least as large as $f(x)$.

Then we repeat the process of subdividing I_1 into two halves and choosing one of the halves according to the principle described above. Call the choice I_2. In this way, we construct a sequence of closed intervals $I_1, I_2, \ldots$, and so on. These intervals are *nested*; that is, the nth interval I_n is contained in the interval I_{n-1}, and its length is one-half of the length of I_{n-1}. Because of the way these intervals were chosen, for every point x in $[a,b]$ and every n, if x is not in I_n, then there is a point z in I_n where the value of f is at least as large as $f(x)$.

We appeal now to the nested interval Theorem 1.19, according to which the subintervals I_n have exactly one point in common; call this point c. We claim that the maximum value of the function f is $f(c)$. For suppose, to the contrary, that there is a point x in $[a,b]$ where the value of f is greater than $f(c)$. Since f is continuous at c, there would be an entire interval $[c-\delta, c+\delta]$ of numbers around c where f is less than $f(x)$. Since the lengths of the intervals I_n tend to zero, it follows that for n large enough, I_n would be contained in the interval $[c-\delta, c+\delta]$, so the value of f at every point of I_n would be smaller than $f(x)$. We can also take n sufficiently large that x is not in I_n. But this contradicts the key property of the intervals I_n established above. □

The extreme value theorem can be extended to open intervals in two special cases.

Corollary 2.1. *If f is continuous on an open interval (a,b) and $f(x)$ tends to infinity as x tends to each of the endpoints, then f has a minimum value at some point in (a,b).*

Similarly, if $f(x)$ tends to minus infinity as x tends to each of the endpoints, then f has a maximum value at some point in (a,b).

We invite you to prove this result in Problem 2.18.

Problems

2.8. Evaluate the following limits.

(a) $\lim_{x\to 4}\left(2x^3+3x+5\right)$

(b) $\lim_{x\to 0}\dfrac{x^2+2}{x^3-7}$

(c) $\lim_{x\to 5}\dfrac{x^2-25}{x-5}$

2.9. Evaluate the following limits.

(a) $\lim_{x\to 0}\dfrac{x^3-9x}{x^2+3x}$

(b) $\lim_{x\to -3}\dfrac{x^3-9x}{x^2+3x}$

(c) $\lim_{x\to 1}\dfrac{x^3-9x}{x^2+3x}$

2.10. Let $f(x)=\dfrac{|x|}{x}$ when $x\neq 0$, and $f(0)=1$.

(a) Sketch the graph of f.
(b) Is f continuous on $[0,1]$?
(c) Is f continuous on $[-1,0]$?
(d) Is f continuous on $[-1,1]$?

2.11. The limit of a function can be completely described in terms of the limits of sequences. To do this, show that these two statements are true:

(a) If $\lim_{x\to c} f(x)=L$ and x_n is any sequence tending to c, then $\lim_{n\to\infty} f(x_n)=L$.
(b) If $\lim_{n\to\infty} f(x_n)=L$ for every sequence x_n tending to c, then $\lim_{x\to c} f(x)=L$.

Conclude that in the discussion of continuity, $\lim_{x\to c} f(x)=f(c)$ is equivalent to $\lim_{n\to\infty} f(x_n)=f(c)$ for every sequence x_n tending to c.

2.12. Suppose that functions f, g, and h are each defined on an interval containing c, that they are continuous at c, and that $h(c) \neq 0$. Show that $f+g$, fg, and $\dfrac{f}{h}$ are continuous at c.

2.13. Let $f(x) = \dfrac{x^{32}+x^{10}-7}{x^2+2}$ on the interval $[-20,120]$. Is f bounded? Explain.

2.14. Show that the equation

$$\frac{x^6+x^4-1}{x^2+1} = 2$$

has a solution on the interval $[-2,2]$.

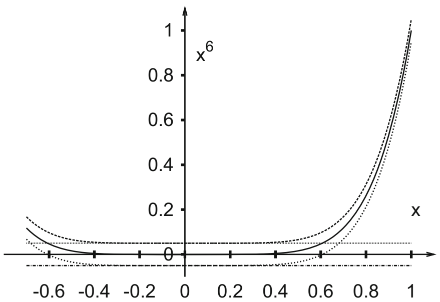

Fig. 2.17 Graphs are shown for x^6, $x^6+\frac{1}{20}$, $x^6-\frac{1}{20}$, and for the constant functions $\pm\frac{1}{20}$ on $[-0.7,1]$. See Problem 2.15

2.15. In Fig. 2.17, estimate the largest interval $[a,b]$ such that x^6 and 0 differ by less than $\frac{1}{20}$ on $[a,b]$.

2.16. Let $f(x) = \dfrac{1}{x}$. Show that on the interval $[3,5]$, $f(x)$ and $f(c)$ do not differ by more than $\frac{1}{9}|x-c|$. Copy the definition of uniform continuity onto your paper, and then explain why f is uniformly continuous on $[3,5]$.

2.17. You plan to compute the squares of numbers between 9 and 10 by squaring truncations of their decimal expansions. If you truncate after the eighth place, will this ensure that the outputs are within 10^{-7} of the true value?

2.18. Prove the first statement in Corollary 2.1, that if f is continuous on (a,b) and $f(x)$ tends to infinity as x tends to each of a and b, then f has a minimum value at some point in (a,b).

2.19. Explain why the function $f(x) = x^2 - \dfrac{1}{x} - 3 = \dfrac{x^3-1-3x}{x}$ is uniformly continuous on every interval $[a,b]$ not containing 0.

2.20. Let $f(x) = 3x + 5$.

(a) Suppose each domain value x is rounded to x_{approx} and $|x - x_{\text{approx}}| < \frac{1}{10^m}$. How close is $f(x_{\text{approx}})$ to $f(x)$?
(b) If we want $|f(x) - f(x_{\text{approx}})| < \frac{1}{10^7}$, how close should x_{approx} be to x?
(c) On what interval can you use the level of precision you found in part (b)?

2.21. Explain the following steps to show that a function that is continuous at every point of a closed interval is uniformly continuous on that interval. It will be a proof by contradiction, so we assume that f is continuous, but not uniformly continuous, on $[a,b]$.

(a) There must be some $\varepsilon > 0$ and for each $n = 1, 2, 3, \ldots$, two numbers x_n, y_n in $[a,b]$ for which $|x_n - y_n| < \frac{1}{n}$ and $|f(x_n) - f(y_n)| \geq \varepsilon$.
(b) Use Lemma 1.1 and monotone convergence to show that a subsequence of the x_n (that is, a sequence consisting of some of the x_n) converges to some number c in $[a,b]$.
(c) To simplify notation, we can now take the symbols x_n to mean the subsequence, and y_n corresponding. Use the fact that $|x_n - y_n| < \frac{1}{n}$ to conclude that the y_n also converge to c.
(d) Use continuity of f and Problem 2.11 to show that $\lim_{n\to\infty} f(x_n) = f(c)$.
(e) Show that $\lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} f(y_n)$, and that this contradicts our assumption that $|f(x_n) - f(y_n)| \geq \varepsilon$.

2.3 Composition and Inverses of Functions

In Sect. 2.1, we showed how to build new functions out of two others by adding, multiplying, and dividing them. In this section, we describe another way.

2.3a Composition

We start with a simple example:

A rocket is launched vertically from point L. The distance (in kilometers) of the rocket from the launch point at time t is $h(t)$. An observation post O is located 1 km from the launch site (Fig. 2.18). To determine the distance d of the rocket from the observation post as a function of time, we can use the Pythagorean theorem to express d as a function of h,

$$d(h) = \sqrt{1 + h^2}.$$

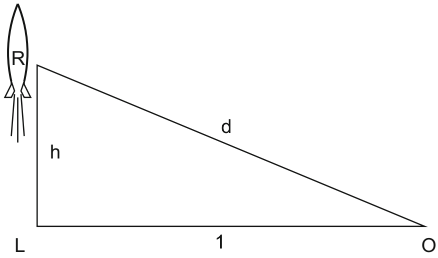

Fig. 2.18 Tracking the rocket from the observation post

Therefore, the distance from R to O at time t is

$$d(h(t)) = \sqrt{1+(h(t))^2}.$$

The process that builds a new function in this way is called *composition*; the resulting function is called the *composition* of the two functions.

Definition 2.6. Let f and g be two functions, and suppose that the range of g is included in the domain of f. Then the composition of f with g, denoted by $f \circ g$, is defined by

$$(f \circ g)(x) = f(g(x)).$$

We also say that we have *composed* the functions.

The construction is well described by Fig. 2.19.

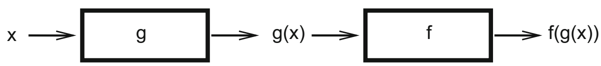

Fig. 2.19 Composition of functions, using the box picture of Fig. 2.4

Example 2.20. Let g and f be the linear functions $y = g(x) = 2x+3$, and $z = f(y) = 3y+1$. The composition $z = f(g(x)) = 3(2x+3)+1 = 6x+10$ is illustrated in Fig. 2.20.

We saw in Fig. 2.8 that the linear function $mx+b$ stretches every interval by a factor of $|m|$. In Fig. 2.20, we see that when the linear functions are composed, these stretching factors are multiplied.

Example 2.21. The effect of composing a function f with $g(x) = x+1$ depends on the order of composition. For example $f(g(x)) = f(x+1)$ shifts the graph of

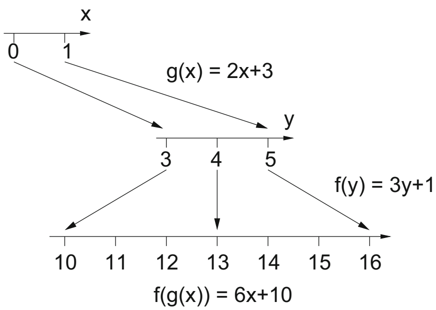

Fig. 2.20 A composition of two linear functions

f one unit to the left, since the output of f at x is the same as the output of $f \circ g$ at $x-1$. On the other hand, $g(f(x)) = f(x)+1$ shifts the graph of f up one unit. See Fig. 2.21.

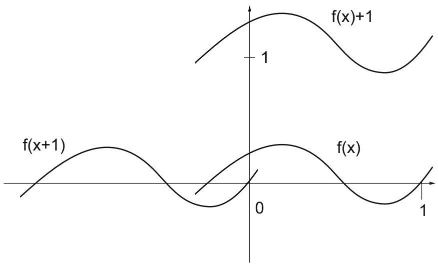

Fig. 2.21 Composition with the translation $x+1$, in Example 2.21. It makes a difference which function is applied first

Example 2.22. Let $h(x) = 3x$. The graph of $f(h(x))$ looks as though the domain of f has been compressed by a factor of 3. This is because the output of f at x is the same as the output of $f \circ h$ at $\frac{x}{3}$. If we compose f and h in the opposite order, the graph of $h(f(x)) = 3f(x)$ is the graph of f stretched by a factor of three in the vertical direction. See Fig. 2.22.

Example 2.23. Let $h(x) = -x$. The graph of $h(f(x)) = -f(x)$ is the reflection of the graph of f across the x-axis, while the graph of $f(h(x)) = f(-x)$ is the reflection of the graph of f across the y-axis.

Example 2.24. If $f(x) = \dfrac{1}{x+1}$ and $g(x) = x^2$, then

$$(f \circ g)(x) = \frac{1}{x^2+1} \quad \text{and} \quad (g \circ f)(x) = \left(\frac{1}{x+1}\right)^2 = \frac{1}{x^2+2x+1}.$$

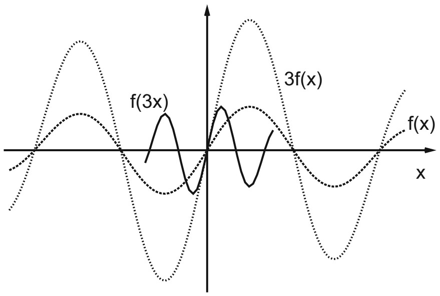

Fig. 2.22 Composition with multiplication $3x$ results in stretching or compressing the graph. See Example 2.22

Notice that $f \circ g$ and $g \circ f$ are quite different functions. Thus *composition is not a commutative operation*. This is not surprising: using the output of g as input for f is quite different from using the output of f as input for g.

Theorem 2.7. *The composition of two continuous functions is continuous.*

Proof. We give an intuitive proof of this result. We want to compare the values of $f(g(x))$ with those of $f(g(z))$ as the numbers x and z vary. Since f is continuous, these values will differ by very little when the numbers $g(x)$ and $g(z)$ are close. But since g is also continuous, those values $g(x)$ and $g(z)$ will be close whenever x and z are sufficiently close. □

Here is a related theorem about limits, which we show you how to prove in Problem 2.33.

Theorem 2.8. *Suppose* $f \circ g$ *is defined on an interval containing* c, *that* $\lim_{x \to c} g(x) = L$, *and that* f *is continuous at* L. *Then* $\lim_{x \to c}(f \circ g)(x) = f(L)$, *that is,*

$$\lim_{x \to c} f\big(g(x)\big) = f\big(\lim_{x \to c} g(x)\big).$$

2.3b Inverse Functions

We look at some examples of compositions of functions that undo each other.

Example 2.25. For $f(x) = 2x + 3$ and $g(x) = \frac{1}{2}x - \frac{3}{2}$, we see that

$$f(g(x)) = 2\left(\frac{1}{2}x - \frac{3}{2}\right) + 3 = x \quad \text{and } g(f(x)) = \frac{1}{2}(2x+3) - \frac{3}{2} = x.$$

Example 2.26. Let $f(x) = \dfrac{1}{x+1}$ when $x \neq -1$, and $g(x) = \dfrac{1-x}{x}$ when $x \neq 0$. Then if $x \neq 0$, we have

$$f(g(x)) = \frac{1}{\left(\frac{1-x}{x}\right)+1} = \frac{1}{\frac{1-x}{x}+\frac{x}{x}} = x.$$

You may also check that when $x \neq -1$, we have $g(f(x)) = x$.

In both of the examples above, we see that f applied to the output of g returns the input of g, and similarly, g applied to the output of f returns the input of f. We may ask the following question about a function: if we know the output, can we determine the input?

Definition 2.7. If a function g has the property that different inputs always lead to different outputs, i.e., if $x_1 \neq x_2$ implies $g(x_1) \neq g(x_2)$, then we can determine its input from the output. Such a function g is called *invertible*; its *inverse* f is defined in words: the domain of f is the range of g, and $f(y)$ is defined as the number x for which $g(x) = y$. We denote the inverse of g by g^{-1}.

By the way in which it is defined, we see that g^{-1} undoes, or reverses, g: it works backward from the output of g to the input. If g is invertible, then g^{-1} is also invertible, and its inverse is g. Furthermore, the composition of a function and its inverse, in either order, is the identity function:

$$(g \circ g^{-1})(y) = y \quad \text{and} \quad (g^{-1} \circ g)(x) = x.$$

Here is another example:

Example 2.27. Let $g(x) = x^2$, and restrict the domain of g to be $x \geq 0$. Since the squares of two different nonnegative numbers are different, g is invertible. Its inverse is $g^{-1}(x) = \sqrt{x}$. Note that if we had defined $g(x) = x^2$ and taken its domain to be all numbers, not just the nonnegative ones, then g would not have been invertible, since $(-x)^2 = x^2$. Thus, invertibility depends crucially on what we take to be the domain of the function (Fig. 2.23).

Monotonicity. The graph of a function can be very helpful in determining whether the function is invertible. If lines parallel to the x-axis intersect the graph in at most one point, then different domain values are assigned different range values, and the function is invertible. Two kinds of functions that pass this "horizontal line test" are the *increasing* functions and the *decreasing* functions.

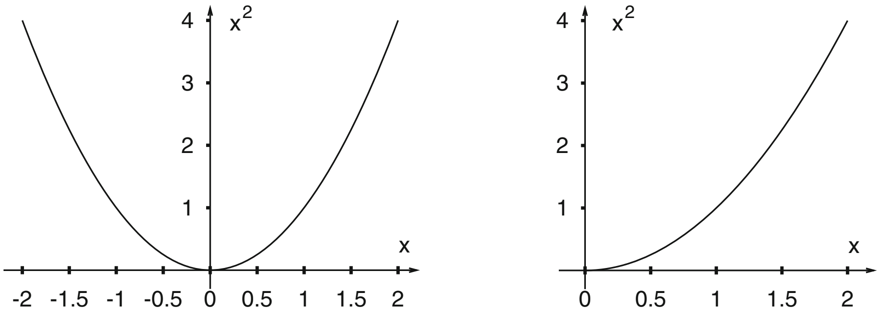

Fig. 2.23 *Left*: x^2 is plotted with the domain all numbers. *Right*: the domain is the positive numbers. Only one of these functions is invertible

Definition 2.8. An *increasing* function is one for which $f(a) < f(b)$ whenever $a < b$. A *decreasing* function is one for which $f(a) > f(b)$ whenever $a < b$. A *nondecreasing* function is one for which $f(a) \leq f(b)$ whenever $a < b$. A *nonincreasing* function is one for which $f(a) \geq f(b)$ whenever $a < b$.

Example 2.28. Suppose f is increasing and $f(x_1) > f(x_2)$. Which of the following is true?

(a) $x_1 = x_2$
(b) $x_1 > x_2$
(c) $x_1 < x_2$

Item (a) is certainly not true, because then we would have $f(x_1) = f(x_2)$. Item (b) is consistent with f increasing, but this does not resolve the question. If item (c) were true, then $f(x_1) < f(x_2)$, which is not possible. So it is (b) after all.

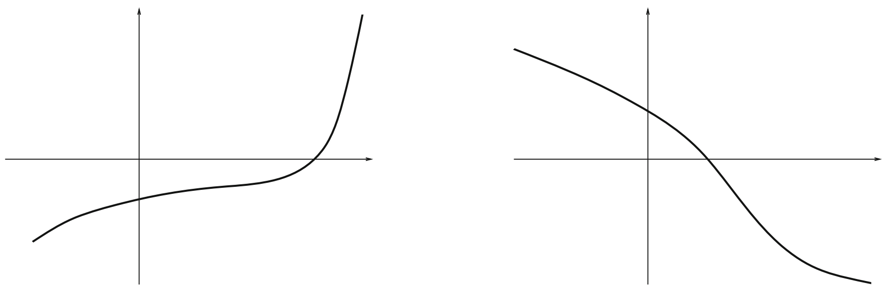

Fig. 2.24 Two graphs of monotonic functions. *Left*: increasing, *Right*: decreasing

Figure 2.24 shows the graphs of an increasing function and a decreasing function. Both pass the horizontal line test and both are invertible.

Definition 2.9. Functions that are either increasing or decreasing are called *strictly monotonic*. Function that are either nonincreasing or nondecreasing are called *monotonic*.

If f is strictly monotonic, then the graph of its inverse is simply the reflection of the graph of f across the line $y = x$ (Fig. 2.25).

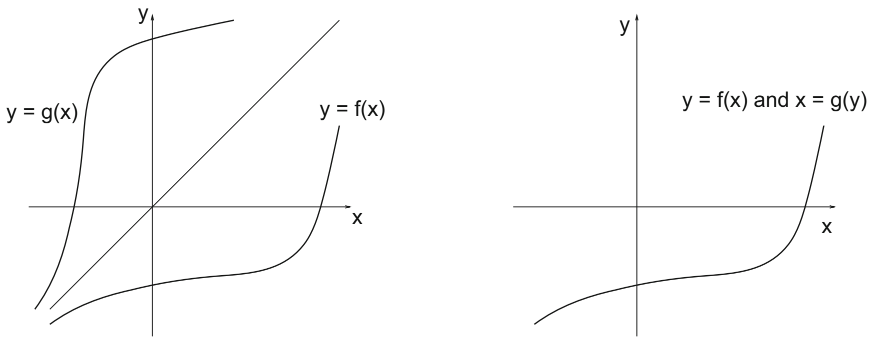

Fig. 2.25 *Left*: graphs of an increasing function f and its inverse g. *Right*: if you write $f(x) = y$ and $x = g(y)$, then the graph of $f(x) = y$ is also the graph of $g(y) = x$

The Inversion Theorem. The graphs suggest the following theorem:

Theorem 2.9. Inversion theorem. *Suppose that f is a continuous and strictly monotonic function defined on an interval $[a,b]$. Then its inverse g is a continuous strictly monotonic function defined on the closed interval between $f(a)$ and $f(b)$.*

Proof. A strictly monotonic function is invertible, because different inputs always result in different outputs. The inverse is strictly monotonic, as we ask you to show in Problem 2.30.

What remains to be shown is that the domain of the inverse function is precisely the closed interval between $f(a)$ and $f(b)$, no more no less, and that f^{-1} is continuous. According to the intermediate value theorem, for every m between $f(a)$ and $f(b)$, there is a number c such that $m = f(c)$. Thus every number between $f(a)$ and $f(b)$ is in the domain of the inverse function. On the other hand, the value $f(c)$ of a strictly monotonic function at the point c between a and b must lie between $f(a)$ and $f(b)$. This shows that the domain of f^{-1} is precisely the closed interval between $f(a)$ and $f(b)$.

Next, we show that f^{-1} is continuous. Let ε be any tolerance. Divide the interval $[a,b]$ into n subintervals of length less than $\frac{\varepsilon}{2}$, with endpoints $a = a_0, a_1, \ldots, a_n = b$.

The values $f(a_i)$ divide the range of f into an equal number of subintervals. Denote by δ the length of the smallest of these. See Fig. 2.26. Let y_1 and y_2 be numbers in the range that are within δ of each other. Then y_1 and y_2 are in either the same or adjacent subintervals of the range. Correspondingly, $f^{-1}(y_1)$ and $f^{-1}(y_2)$ lie in the same or adjacent subintervals of $[a,b]$. Since the lengths of the subintervals of $[a,b]$ were made less than $\frac{\varepsilon}{2}$, we have

$$|f^{-1}(y_1) - f^{-1}(y_2)| < \varepsilon.$$

Thus we have shown that given any tolerance ε, there is a δ such that if y_1 and y_2 differ by less than δ, then $f^{-1}(y_1)$ and $f^{-1}(y_2)$ differ by less than ε. This shows that f is uniformly continuous on $[a,b]$, hence continuous. □

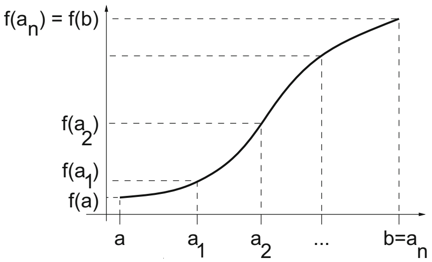

Fig. 2.26 The inverse of a continuous strictly monotonic function is continuous

As an application of the inversion theorem, take $f(x) = x^n$, n any positive integer. Then f is continuous and increasing on every interval $[0,b]$, so it has an inverse g. The value of g at a is the nth root of a and is written with a fractional exponent:

$$g(a) = a^{1/n}.$$

By the inversion theorem, the nth-root function is continuous and strictly monotonic. Then powers of such functions, such as $x^{2/3} = (x^{1/3})^2$, are continuous and strictly monotonic on $[0,b]$. Figure 2.27 shows some of these functions and their inverses.

We shall see later that many important functions can be defined as the inverse of a strictly monotonic continuous function and that we can make important deductions about a function f from properties of its inverse f^{-1}.

Problems

2.22. Find the inverse function of $f(x) = x^5$. Sketch the graphs of f and f^{-1}.

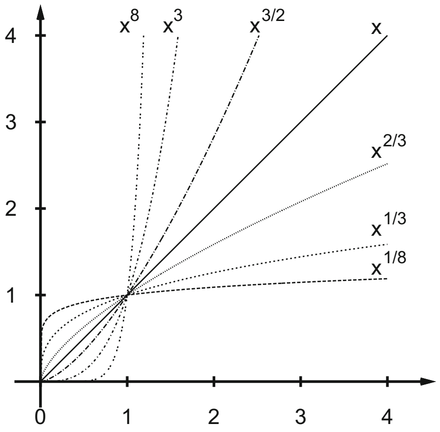

Fig. 2.27 The power functions

2.23. The volume of water V in a bottle is a function of the height H of the water, say $V = f(H)$. See Fig. 2.28. Similarly, the height of the water is a function of the volume of water in the bottle, say $H = g(V)$. Show that f and g are inverse functions.

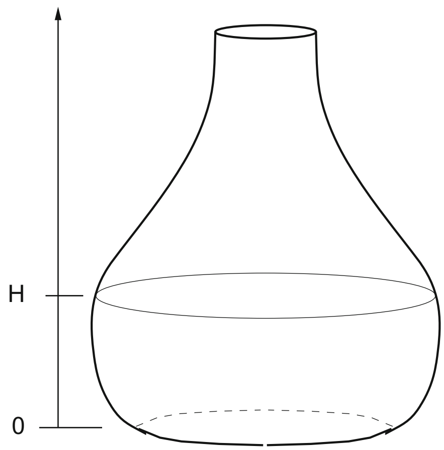

Fig. 2.28 A bottle of water for Problem 2.23

2.24. Let $f(x) = x$, $g(x) = x^2$, $h(x) = x^{1/5}$, and $k(x) = x^2 + 5$. Find formulas for the compositions

(a) $(h \circ g)(x)$
(b) $(g \circ h)(x)$
(c) $(f \circ g)(x)$
(d) $(k \circ h)(x)$

(e) $(h \circ k)(x)$
(f) $(k \circ g \circ h)(x)$

2.25. Is there a function $f(x) = x^a$ that is its own inverse function? Is there more than one such function?

2.26. Show that the function $f(x) = x - \dfrac{1}{x}$, on domain $x > 0$, is increasing by explaining each of the following items.

(a) The sum of two increasing functions is increasing.
(b) The functions x and $-\dfrac{1}{x}$ are increasing.

2.27. Tell how to compose some of the functions defined in Problem 2.24 to produce the functions

(a) $(x^2+5)^2+5$
(b) $(x^2+5)^2$
(c) x^4+5

2.28. The graph of a function f on $[0,a]$ is given in Fig. 2.29. Use the graph of f to sketch the graphs of the following functions.

(a) $f(x-a)$
(b) $f(x+a)$
(c) $f(-x)$
(d) $-f(x)$
(e) $f(-(x-a))$

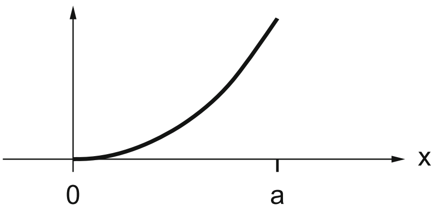

Fig. 2.29 The graph of the function f in Problem 2.28

2.29. Use the intermediate value theorem to show that the equation

$$\sqrt{x^2+1} = \sqrt[3]{x^5+2}$$

has a solution in $[-1,0]$.

2.30. (a) Show that the inverse of an increasing function is increasing. (b) Then state the analogous result for decreasing functions.

2.31. (a) Suppose f is increasing. Is $f \circ f$ increasing? Give a proof or a counterexample. (b) Suppose f is decreasing. Is $f \circ f$ decreasing? Give a proof or a counterexample.

2.32. Assume that functions f and g are increasing. Is fg increasing? If so give a proof, and if not, explain why not.

2.33. Prove Theorem 2.8 by explaining the following.

(a) Given any $\varepsilon > 0$, there is a $\delta > 0$ such that if $|z-L| < \delta$, then $|f(z)-f(L)| < \varepsilon$.
(b) For the δ in part (a), there is an $\eta > 0$ such that if $|x-c| < \eta$, then $|g(x)-L| < \delta$.
(c) Given any $\varepsilon > 0$, there is an $\eta > 0$ such that if $|x-c| < \eta$, then $|f(g(x)) - f(L)| < \varepsilon$.
(d) $\lim_{x\to c} f(g(x)) = f(L)$.

2.4 Sine and Cosine

It is often asserted that the importance of trigonometry lies in its usefulness for surveying and navigation. Since the proportion of our population engaged in these pursuits is rather small, one wonders what kind of stranglehold surveyors and navigators have over professional education to be able to enforce the universal teaching of this abstruse subject. Or is it merely inertia? The answer, of course, is that the importance of trigonometry lies elsewhere: in the description of *rotation* and *vibration*. It is an astonishing fact of mathematical physics that the vibration of as diverse a collection of objects as:

springs
strings
airplane wings
steel beams
light beams
and water streams
building sways
ocean waves
and sound waves ...

and many others are described in terms of trigonometric functions. That such diverse phenomena can be treated with a common tool is one of the most striking successes of calculus. Some simple and some not so simple examples will be discussed in the next chapters.

One also learns from older texts that there are six trigonometric functions:

sine, cosine, tangent, cotangent
secant, and cosecant.

This turns out to be a slight exaggeration. There are only two basic functions, sine and cosine; all the others can be defined in terms of them, when necessary.

Furthermore, sine and cosine are so closely related that each can be expressed in terms of the other; so one can say that there is really only one trigonometric function.

Geometric Definition. We shall describe the functions sine and cosine geometrically, using the circle of radius 1 in the Cartesian (x,y)-plane centered at the origin, which is called the *unit circle* (Fig. 2.30).

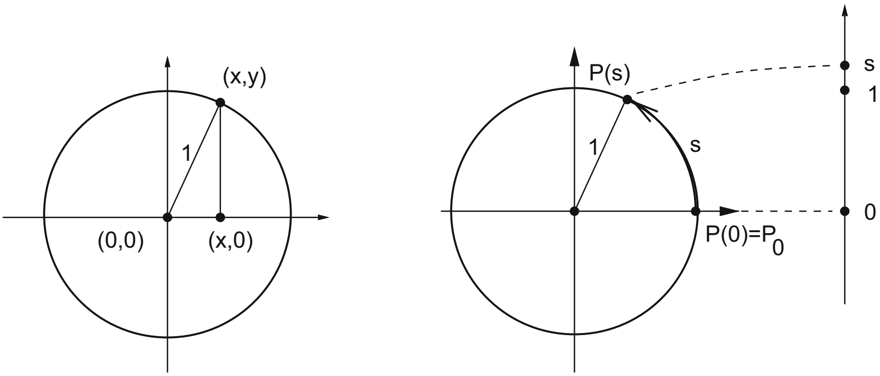

Fig. 2.30 *Left*: the unit circle. *Right*: measuring along the circumference using the same scale as on the axes is called radian measure

Let (x,y) be any point on the unit circle. The triangle with vertices $(0,0)$, $(x,0)$, and (x,y) is a right triangle. By the Pythagorean theorem,

$$x^2+y^2=1.$$

Let P_0 be the point $(1,0)$ on the unit circle. Let $P(s)$ be that point on the unit circle whose distance measured from P_0 counterclockwise along the arc of the unit circle is s.

You can imagine this distance along the arc with the aid of a very thin string of length s. Fasten one of its ends to the point P_0, and wrap the string counterclockwise around the circle. The other end of the string is at the point $P(s)$.

The two rays from the origin through the points P_0 and $P(s)$ form an angle. We define the size of this angle to be s, the length of the arc connecting P_0 and $P(s)$. Measuring along the circumference of the unit circle using the same scale as on the axes is called radian measure (Fig. 2.30). An angle of length 1 therefore has measure equal to one radian, and the radian measure of a right angle is $\frac{\pi}{2}$.

Definition 2.10. Denote the x- and y-coordinates of $P(s)$ by $x(s)$ and $y(s)$. We define

$$\cos s = x(s), \quad \sin s = y(s).$$

One immediate consequence of the definition is that $\cos s$ and $\sin s$ are continuous functions: The length of the chord between $P(s)$ and $P(s+\varepsilon)$ is less than ε. The differences Δx and Δy in the coordinates between $P(s)$ and $P(s+\varepsilon)$ are each less than the length of the chord. But these differences are also the changes in the cosine and sine:

$$|\Delta x| = |\cos(s+\varepsilon) - \cos s| < \varepsilon, \qquad |\Delta y| = |\sin(s+\varepsilon) - \sin s| < \varepsilon.$$

See Fig. 2.31.

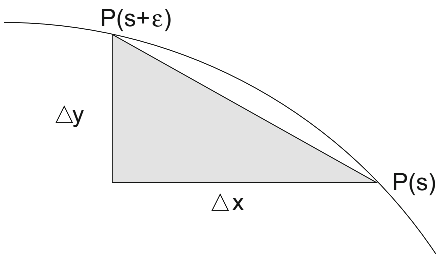

Fig. 2.31 A small arc of the unit circle with corresponding x and y increments. Observe that the x and y increments are smaller than the arc increment

We list this fact of continuity together with some other properties:

(a) The cosine and sine functions are continuous.
(b) $\cos^2 s + \sin^2 s = 1$. This is because the cosine and sine are the coordinates of a point of the unit circle, where $x^2 + y^2 = 1$.
(c) Since the circumference of the whole unit circle is 2π, when a string of length $s+2\pi$ is wrapped around the unit circle in the manner described before, the endpoint $P(s+2\pi)$ coincides with the point $P(s)$. Therefore,

$$\cos(s+2\pi) = \cos s, \quad \sin(s+2\pi) = \sin s.$$

This property of the functions sine and cosine is called "periodicity," with period 2π. See Fig. 2.32.
(d) $\cos 0 = 1$, and the value $\cos s$ decreases to -1 as s increases to π. Then $\cos s$ increases again to 1 at $s = 2\pi$. $\sin 0 = 0$ and $\sin s$ also varies from -1 to 1.
(e) $P\left(\frac{\pi}{2}\right)$ lies one quarter of the circle from P_0. Therefore, $P\left(\frac{\pi}{2}\right) = (0,1)$, and

$$\cos\left(\frac{\pi}{2}\right) = 0, \qquad \sin\left(\frac{\pi}{2}\right) = 1.$$

(f) The point $P\left(\frac{\pi}{4}\right)$ is halfway along the arc between P_0 and $P\left(\frac{\pi}{2}\right)$. By symmetry, we see that $x\left(\frac{\pi}{4}\right) = y\left(\frac{\pi}{4}\right)$. By the Pythagorean theorem, $\left(x\left(\frac{\pi}{4}\right)\right)^2 + \left(y\left(\frac{\pi}{4}\right)\right)^2 = 1$. It follows that $(\cos\frac{\pi}{4})^2 = (\sin\frac{\pi}{4})^2 = \frac{1}{2}$, and so

$$\cos\left(\frac{\pi}{4}\right) = \sqrt{\frac{1}{2}}, \qquad \sin\left(\frac{\pi}{4}\right) = \sqrt{\frac{1}{2}}.$$

(g) For angles s and t, there are addition formulas

$$\cos(s+t) = \cos s \cos t - \sin s \sin t,$$
$$\sin(s+t) = \sin s \cos t + \cos s \sin t,$$

which will be discussed later.

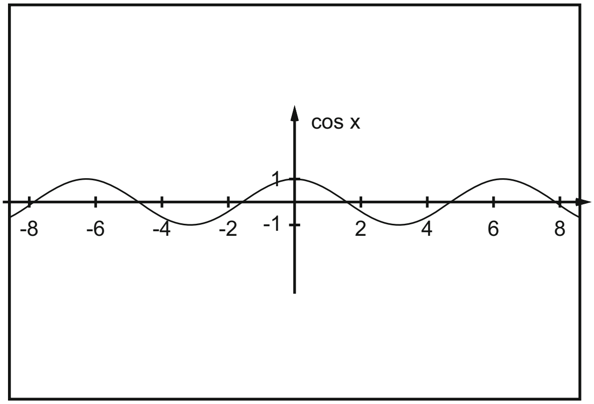

Fig. 2.32 Part of the graph of the cosine

Problems

2.34. On a sketch of the unit circle, mark the circumference at six equally spaced points. Are these subdivisions more, or less, than one radian each?

2.35. Which of the following pairs of numbers could be the cosine and sine of some angle?

(a) $(0.9, 0.1)$
(b) $(\sqrt{0.9}, \sqrt{0.1})$

2.36. Sketch the unit circle, and on it, mark the approximate location of points having angles of 1, 2, 6, 2π, and -0.6 from the horizontal axis.

2.37. The ancient Babylonians measured angles in degrees. They divided the full circle into 360 angles of equal size, each called one degree. So the size of a right angle in Babylonian units is 90 degrees. Since its size in modern units is $\frac{\pi}{2}$ radians, it follows that one radian equals $\frac{90}{\frac{\pi}{2}} = 57.295\ldots$ degrees. Let $c(x) = \cos\left(\frac{x}{57.295\ldots}\right)$ which is the cosine of an angle of x degrees. Sketch the graph of c as nearly as you can to scale, and explain how it differs from the graph of the cosine.

2.38. Which of the following functions are bounded, and which are bounded away from 0?

(a) $f(x) = \sin x$
(b) $f(x) = 5\sin x$
(c) $f(x) = \dfrac{1}{\sin x}$ for $x \neq n\pi,\ n = 0, \pm 1, \pm 2, \ldots$

2.39. A weight attached to a Slinky (a weak spring toy) oscillates up and down. Its position at time t is $y = 1 + 0.2\sin(3t)$ meters from the floor. What is the maximum height reached, and how much time elapses between successive maxima?

2.40. Use the intermediate value theorem to prove that the equation

$$x = \cos x$$

has a solution on the interval $[0, \frac{\pi}{2}]$.

2.41. Show that $\sin s$ is an increasing function on $[-\frac{\pi}{2}, \frac{\pi}{2}]$, and therefore has an inverse. Its inverse is denoted by $\sin^{-1}$.

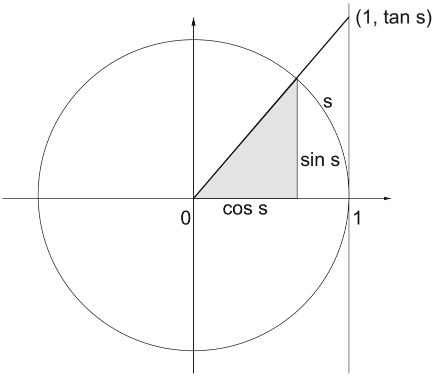

Fig. 2.33 The tangent of s. See Problems 2.42 and 2.43

2.42. Define the tangent function by $\tan s = \dfrac{\sin s}{\cos s}$ whenever the denominator is not 0. Refer to Fig. 2.33 to show that $\tan s$ is an increasing function on $(-\frac{\pi}{2}, \frac{\pi}{2})$. Show that $\tan x$ has a continuous inverse on $(-\infty, \infty)$. Its inverse is denoted by $\tan^{-1}$.

2.43. Set $z = \tan s$ and $y = \sin s$ in Fig. 2.33.

(a) Show that $\sin(\tan^{-1}(z)) = \dfrac{z}{\sqrt{1+z^2}}$.
(b) Express $\cos(\sin^{-1}(y))$ without using any trigonometric functions.

2.5 Exponential Function

We present two examples of naturally occurring functions f that arise in modeling growth and decay and satisfy the relation

$$f(t+s) = f(t)f(s). \tag{2.1}$$

We shall then show that all continuous functions satisfying this relation are exponential functions. Further natural examples of exponential functions are given in Chap. 10.

2.5a Radioactive Decay

Radioactive elements are not immutable. With the passage of time, they change into other elements. It is important to know how much of a given amount is left after time t has elapsed. To express this problem mathematically, we describe the decay by the following function:

Let $M(t)$ denote the fraction of material of a unit mass remaining after the elapse of time t. Assume that M is a continuous function of time, $M(0) = 1$, and that $0 < M(t) < 1$ for $t > 0$.

How much will be left of an initial supply of mass A after the elapse of time t? The number of atoms present does not affect the likelihood of any individual atom decaying. A solitary atom is as likely to decay as one buried among thousands of other atoms. Since $M(t)$ is the fraction of material left of a unit mass after time t, $AM(t)$ is the amount left after time t if we start with mass A:

$$(\text{amount left at time } t) = AM(t) \tag{2.2}$$

How much will be left of a mass A of material after time $s+t$ has elapsed? By definition of the function M, the amount left is $AM(s+t)$. But there is another way of answering this question. Observe that after time s has elapsed, the remaining mass is $AM(s)$, and then after an additional time t has elapsed, the amount left is $\big(AM(s)\big)M(t)$. These two answers must be the same, and therefore,

$$M(s+t) = M(s)M(t). \tag{2.3}$$

Since $M(s)$ and $M(t)$ are less than 1, M is decreasing, and $M(t)$ tends to zero as t tends to infinity. We assumed that M is a continuous function, and $M(0) = 1$. According to the intermediate value theorem, there is a number h for which $M(h) = 1/2$. Since M is decreasing, there is only one such number. Setting $s = h$ in relation (2.3), we get that

$$M(h+t) = \frac{1}{2}M(t).$$

In words: starting at any time t, let additional time h elapse, then the mass of the material is halved. The number h is called the *half-life* of the radioactive material.

For example, the half-life of radium-226 is about 1601 years, and the half-life of carbon-14 is about 5730 years.

2.5b Bacterial Growth

We turn to another example, the growth of a colony of bacteria. We describe the growth by the following function:

Let $P(t)$ be the size of the bacterial population of initial unit size after it has grown for time t. Assume that P is a continuous function of time, $P(0)=1$, and that $P(t)>1$ for $t>0$.

If we supply ample nutrients so that the bacteria do not have to compete with each other, and if there is ample room for growth, then it is reasonable to conclude that the size of the colony at any time t is proportional to its initial size A, whatever that initial size is:

$$(\text{size at time } t) = AP(t), \tag{2.4}$$

What will be the size of a colony, of initial size A, after it has grown for time $s+t$? According to Eq. (2.4), the size will be $AP(s+t)$. But there is another way of calculating the size of the population. After time s has elapsed, the population size has grown to $AP(s)$. After an additional time t elapses, the size of the population will, according to Eq. (2.4), grow to $\big(AP(s)\big)P(t) = AP(s)P(t)$. The two answers must be the same, and therefore,

$$P(s+t) = P(s)P(t). \tag{2.5}$$

Since $P(t)>1$, P is an increasing function, and $P(t)$ tends to infinity as t increases. We assumed that P is continuous and $P(0)=1$, so by the intermediate value theorem, there is a value d for which $P(d)=2$. Since P is an increasing function, there is only one such value. Setting $s=d$ in Eq. (2.5) gives

$$P(d+t) = 2P(t);$$

d is called the *doubling time* for the bacterial colony. Starting from any time t, the colony doubles after additional time d elapses.

2.5c Algebraic Definition

Next we show that every continuous function f that satisfies

$$f(x+y) = f(x)f(y) \quad \text{and} \quad a = f(1) > 0,$$

must be an exponential function $f(x)=a^x$. For example, $P(t)$ and $M(t)$ in the last section are such functions.

The relation $f(x+y) = f(x)f(y)$ is called the functional equation of the exponential function. If $y = x$, the equation gives

$$f(x+x) = f(2x) = f(x)f(x) = (f(x))^2 = f(x)^2,$$

where in the last form we have omitted unnecessary parentheses. When $y = 2x$, we get

$$f(x+2x) = f(x)f(2x) = f(x)f(x)^2 = f(x)^3.$$

Continuing in this fashion, we get

$$f(nx) = f(x)^n. \tag{2.6}$$

Take $x = 1$. Then

$$f(n) = f(n1) = f(1)^n = a^n.$$

This proves that $f(x) = a^x$ when x is any positive integer. Take $x = \dfrac{1}{n}$ in Eq. (2.6). We get $f(1) = a = f\big(\dfrac{1}{n}\big)^n$. Take the nth root of both sides. We get $f\big(\frac{1}{n}\big) = a^{1/n}$. This proves that $f(x) = a^x$ when x is any positive integer reciprocal. Next take $x = \dfrac{1}{p}$ in Eq. (2.6); we get

$$f\Big(\frac{n}{p}\Big) = f\Big(\frac{1}{p}\Big)^n = \big(a^{1/p}\big)^n = a^{n/p}.$$

So we have shown that for all positive rational numbers $r = \dfrac{n}{p}$,

$$f(r) = a^r.$$

In Problem 2.52, we ask you to show that $f(0) = 1$ and that $f(r) = a^r$ for all negative rational numbers r. Assume that f is continuous. Then it follows that $f(x) = a^x$ for irrational x as well, since x can be approximated by rational numbers.

The algebraic properties of the exponential functions a^x extend to all numbers x as well, where $a > 0$:

- $a^x a^y = a^{x+y}$
- $(a^x)^n = a^{nx}$
- $a^0 = 1$
- $a^{-x} = \dfrac{1}{a^x}$
- $a^x > 1$ for $x > 0$ and $a > 1$
- $a^x < 1$ for $x > 0$ and $0 < a < 1$

We can use these properties to show that for $a > 1$, $f(x) = a^x$ is an increasing function. Suppose $y > x$. Then $y - x > 0$, and $a^{y-x} > 1$. Since $a^{y-x} = \dfrac{a^y}{a^x}$, it follows that $a^y > a^x$. By a similar argument when $0 < a < 1$, we can show that a^x is decreasing.

2.5d Exponential Growth

Though it has a precise mathematical meaning, the phrase "exponential growth" is often used as a metaphor for any extremely rapid increase. Here is the mathematical basis of this phrase:

Theorem 2.10. Exponential growth. *For $a > 1$, the function a^x grows faster than x^k as x tends to infinity, no matter how large the exponent $k = 0, 1, 2, 3 \dots$.*

In other words, the quotient $\dfrac{a^x}{x^k}$ tends to infinity as x tends to infinity (Fig. 2.34).

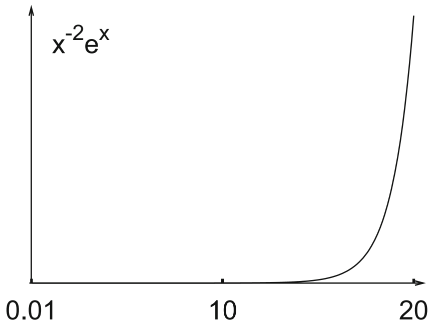

Fig. 2.34 The function $\dfrac{e^x}{x^2}$ plotted on $[0.01, 20]$. The *vertical scale* is compressed by a factor of 100,000

Proof. We first consider the case $k = 0$: that a^x tends to infinity for all a greater than 1. This is certainly true for $a = 10$, because $10^2 = 100$, $10^3 = 1000$, etc., clearly tend to infinity. It follows that a^x tends to infinity for all a greater than 10.

Consider the set of all numbers a for which a^x is bounded for all positive x. The set is not empty, because, for example, $a = 1$, and $a = \dfrac{1}{2}$ have this property. The set has an upper bound, because every number larger than 10 is *not* in the set. So the set of such a has a least upper bound. Denote the least upper bound by c. Since $a = 1$ lies in the set, c is not less than 1. We claim that c is 1. For suppose that c were greater than 1. Then b, the square root of c, and d, the square of c, would satisfy the inequalities

$$b < c < d.$$

Since d is greater than the least upper bound c, d^x tends to infinity with x. Since by definition, d is b^4, $b^{4x} = d^x$ tends to infinity with x. But since b is less than the least upper bound c, its powers remain bounded. This is a contradiction, so c must be 1. Therefore, a^x tends to infinity for all a greater than 1.

Next we consider the case $k=1$: $\dfrac{a^x}{x}$ tends to infinity as x tends to infinity. Denote the function $\dfrac{a^x}{x}$ by $f(x)$. Then

$$f(x+1) = \frac{a^{x+1}}{x+1} = \frac{a^x}{x}\frac{a}{1+\frac{1}{x}} = f(x)\frac{a}{1+\frac{1}{x}}. \tag{2.7}$$

We claim that for large x, the factor $\dfrac{a}{1+\frac{1}{x}}$ is larger than 1: we know that $a>1$, so in fact, $a > 1+\dfrac{1}{m}$ for some integer m. Write $b = \dfrac{a}{1+\frac{1}{m}}$. Then for all $x \geq m$,

$$\frac{a}{1+\frac{1}{x}} \geq \frac{a}{1+\frac{1}{m}} = b > 1,$$

as claimed. Then by Eq. (2.7),

$$f(x+1) \geq f(x)b,$$
$$f(x+2) \geq f(x)b^2,$$

and continuing in this way, we see that

$$f(x+n) \geq f(x)b^n$$

for each positive integer n. Every large number X can be represented as some number x in $[m,m+1]$ plus a large positive integer n. Denote by M the minimum value of f in $[m,m+1]$. Then

$$f(X) = f(x+n) \geq f(x)b^n \geq Mb^n.$$

Since $b>1$, this shows that $f(X)$ tends to infinity as X does.

In the cases $k>1$, we argue as follows. Using the rules for the exponential function, we see that

$$\frac{a^x}{x^k} = \left(\frac{s^x}{x}\right)^k, \qquad \text{where} \quad s^k = a. \tag{2.8}$$

Since a is greater than 1, so is s. As we have already shown, $\dfrac{s^x}{x}$ tends to infinity as x does. Then so does its kth power. □

Later, in Sect. 4.1b, we shall give a much simpler proof of the theorem on exponential growth using calculus.

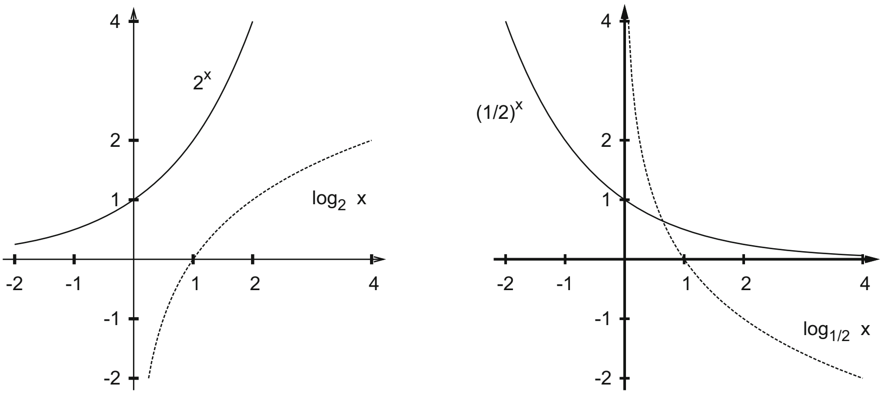

Fig. 2.35 *Left*: graphs of 2^x and $\log_2 x$. *Right*: graphs of $\left(\frac{1}{2}\right)^x$ and $\log_{1/2} x$

2.5e Logarithm

For a greater than 1, a^x is an increasing continuous function, and for $0 < a < 1$, a^x is decreasing. Hence for $a \neq 1$, a^x has a continuous inverse function, called the logarithm to the base a, which is defined by

$$\log_a y = x \quad \text{when} \quad y = a^x.$$

If $a > 1$, $\log_a$ is an increasing function. If $0 < a < 1$, $\log_a$ is a decreasing function. In either case, the domain of $\log_a$ is the range of a^x, all positive numbers (Fig. 2.35).

The exponential function is characterized by

$$a^x a^y = a^{x+y}.$$

Applying the function $\log_a$, we get

$$\log_a(a^x a^y) = x + y.$$

Take any two positive numbers u and v and denote their logarithms by x and y:

$$x = \log_a u, \qquad a^x = u, \qquad y = \log_a v, \qquad a^y = v. \tag{2.9}$$

We get

$$\log_a(uv) = \log_a u + \log_a v. \tag{2.10}$$

Calculations. The logarithm was invented by the Scottish scientist John Napier and expounded in a work published in 1614. Napier's logarithm was to the base e. In English, this is called called the "natural logarithm," a phrase that will be explained in the next section.

The base-10 logarithm, called the "common logarithm" in English, was introduced by Henry Briggs in 1617, based on Napier's logarithm. The significance of base ten is this: every positive number a can be written as $a = 10^n x$ (recall scientific notation $a = x \times 10^n$, where n is an integer and x is a number between 1 and 10). Then $\log_{10} a = n + \log_{10} x$. Therefore, the base-ten logarithms for numbers between 1 and 10 are sufficient to determine the base-ten logarithms for all positive numbers.

Table 2.1 is part of a traditional table of base-10 logarithms. It shows numbers 1.000 through 9.999, the last digit being read across the top row. We know that $\log_{10}(9.999)$ is nearly $\log_{10}(10) = 1$, and this tells us how to read the table: the entry in the lower right-hand corner must mean that $\log_{10}(9.999) = 0.99996$.

We illustrate multiplication by an example:

Example 2.29. What is the product of $a = 4279$ and $b = 78{,}520$? Write $a = 4.279 \times 10^3$. According to Table 2.1,

$$\log_{10}(4.279) = 0.63134.$$

Therefore, $\log_{10} a = 3.63134$. Similarly, $b = 7.852 \times 10^4$. According to the table, then,

$$\log_{10}(7.852) = 0.89498,$$

and therefore, $\log_{10} b = 4.89498$. To multiply a and b we use the fundamental property (2.10) of logarithms to write

$$\log_{10} ab = \log_{10} a + \log_{10} b = 3.63164 + 4.89498 = 8.52632.$$

By the table, the number whose base-10 logarithm is 0.52632 is, within a tolerance of 2×10^{-4}, equal to 3.360. This shows that the product ab is approximately 336,000,000, within a tolerance of 2×10^4.

Using a calculator, we get $ab = 335{,}987{,}080$, which is quite close to our approximate value calculated using base-10 logarithms.

No.	0	1	2	3	4	5	6	7	8	9
100	00000	00043	00087	00130	00173	00217	00260	00303	00346	00389
...	–	–	–	–	–	–	–	–	–	–
335	52504	52517	52530	52543	52556	52569	52582	52595	52608	52621
336	52634	52647	52660	52673	52686	52699	52711	52724	52737	52750
...	–	–	–	–	–	–	–	–	–	–
427	63043	63053	63063	63073	63083	63094	63104	63114	63124	63134
428	63144	63155	63165	63175	63185	63195	63205	63215	63225	63236
...	–	–	–	–	–	–	–	–	–	–
526	72099	72107	72115	72123	72132	72140	72148	72156	72165	72173
...	–	–	–	–	–	–	–	–	–	–
785	89487	89492	89498	89504	89509	89515	89520	89526	89531	89537
...	–	–	–	–	–	–	–	–	–	–
999	99957	99961	99965	99970	99974	99978	99981	99987	99991	99996
No.	0	1	2	3	4	5	6	7	8	9

Table 2.1 Excerpt from the $\log_{10}$ tables in Bowditch's practical navigator, 1868. We read, for example, $\log_{10}(3.358) = 0.52608$ from row no. 335, column 8

Division is carried out the same way, except we subtract the logarithms instead of adding them.

One cannot exaggerate the historical importance of being able to do arithmetic with base-ten logarithms. Multiplication and division by hand is a time-consuming, frustrating activity, prone to error.[1] For 350 years, no scientist, no engineer, no office, no laboratory was without a table of base-ten logarithms. Because of the force of habit, most scientific calculators have the base-ten logarithm available, although the main use of those logarithms is to perform multiplication and division. Of course, these arithmetic operations are performed by a calculator by pressing a button. The button labeled "log" often means $\log_{10}$. In the past, the symbol $\log x$, without any subscript, denoted the logarithm to base ten; the natural log of x was denoted by $\ln x$. Since in our time, multiplication and division are done by calculators, the base-ten logarithm is essentially dead, and rather naturally, $\log x$ has come to denote the natural logarithm of x.

Why Is the Natural Logarithm Natural? The explanation you will find in the usual calculus texts is that the inverse of the base-e logarithm, the base-e exponential function, is the most natural of all exponential functions because it has special properties related to calculus. Since Napier did not know what the inverse of the natural logarithm was, nor did he know calculus (he died about 25 years before Newton was born), his motivation must have been different. Here it is:

Suppose f and g are functions inverse to each other. That is, if $f(x) = y$, then $g(y) = x$. Then if we have a list of values $f(x_j) = y_j$ for the function f, it is also a list of values $x_j = g(y_j)$ for the function g. As an example, take the exponential function $f(x) = (10)^x = y$. Here is a list of its values for $x = 0, 1, 2, \ldots, 10$:

x	0	1	2	...	9	10
y	1	10	100	...	1,000,000,000	10,000,000,000

The inverse of the function $(10)^x = y$ is the base-10 logarithm, $\log_{10} y = x$. We have listed above its values for $y = 1, 10, 100, \ldots, 10,000,000,000$. The trouble with this list is that the values y for which $\log_{10} y$ is listed are very far apart, so we can get very little information about $\log_{10} y$ for values of y in between the listed values.

Next we take the base-2 exponential function $f(x) = 2^x = y$. Here is a list of its values for $x = 0, 1, 2, \ldots, 10$:

x	0	1	2	...	9	10
y	1	2	4	...	512	1024

The inverse of the function $2^x = y$ is the base-2 logarithm, $\log_2 y = x$. Here the values y for which the values of $\log_2 y$ are listed are not so far apart, but they are still quite far apart.

[1] There is a record of an educational conference in the Middle Ages on the topic, "Can one teach long division without flogging?"

Clearly, to make the listed values of the exponential function lie close together, we should choose the base small, but still greater than 1. So let us try the base $a = 1.01$. Here is a list of the values of $y = (1.01)^x$ for $x = 0, 1 \ldots 100$. Note that the evaluation of this exponential function for integer values of x requires just one multiplication for each value of x:

x	0	1	2	...	99	100
y	1	1.01	1.0201	...	2.6780	2.7048

The inverse of the function $(1.01)^x = y$ is the base-1.01 logarithm, $\log_{1.01} y = x$. The listed values y of the base-1.01 logarithm are close to each other, but the values of the logarithms are rather large: $\log_{1.01} 2.7048 = 100$. There is an easy trick to fix this. Instead of using 1.01 as the base, use $a = (1.01)^{100}$. Then

$$\left((1.01)^{100}\right)^x = (1.01)^{100x}.$$

We list values of a^x now for $x = 0, 0.01, 0.02, \ldots, 1.00$, which gives a table almost identical to the previous table:

x	0	0.01	0.02	...	0.99	1.00
y	1	1.01	1.0201	...	2.6780	2.7048

To further improve matters, we can take powers of numbers even closer to 1 as a base: Take as base $1 + \frac{1}{n}$ raised to the power n, where n is a large number. As n tends to infinity, $\left(1 + \dfrac{1}{n}\right)^n$ tends to e, the base of the natural logarithm.

Problems

2.44. Use the property $e^{x+y} = e^x e^y$ to find the relation between e^z and e^{-z}.

2.45. Suppose f is a function that satisfies the functional equation $f(x+y) = f(x)f(y)$, and suppose c is any number. Define a function $g(x) = f(cx)$. Explain why $g(x+y) = g(x)g(y)$.

2.46. A bacteria population is given by $p(t) = p(0)a^t$, where t is in days since the initial time. If the population was 1000 on day 3, and 200 on day 0, what was it on day 1?

2.47. A population of bacteria is given by $p(t) = 800(1.023)^t$, where t is in hours. What is the initial population? What is the doubling time for this population? How long will it take to quadruple?

2.48. Let P_0 be the initial principal deposited in an account. Write an expression for the account balance after 1 year in each of the following cases.

(a) 4 % simple interest,

(b) 4 % compounded quarterly (4 periods per year),
(c) 4 % compounded daily (365 periods per year),
(d) 4 % compounded continuously (number of periods tends to infinity),
(e) x % compounded continuously.

2.49. Calculate the product ab by hand, where a and b are as in Example 2.29.

2.50. Solve $e^{-x^2} = \frac{1}{2}$ for x.

2.51. Suppose $f(x) = ma^x$, and we know that

$$f\left(x+\frac{1}{2}\right) = 3f(x).$$

Find a.

2.52. Use the functional equation $f(x+y) = f(x)f(y)$ and $f(1) = a \neq 0$ to show that

(a) $f(0) = 1$,
(b) $f(r) = a^r$ for negative rational numbers r.

2.53. Suppose P satisfies the functional equation $P(x+y) = P(x)P(y)$, and that N is any positive integer. Prove that

$$P(0)+P(1)+P(2)+\cdots+P(N)$$

is a finite geometric series.

2.54. If b is the arithmetic mean of a and c, prove that e^b is the geometric mean of e^a and e^c.

2.55. Knowing that $e > 2$, explain why

(a) $e^{10} > 1000$,
(b) $\log 1000 < 10$,
(c) $\log 1{,}000{,}000 < 20$.

2.56. Let a denote a number greater than 1, $a = 1+p$, where p is positive. Show that for all positive integers n, $a^n > 1+pn$.

2.57. We know that $\frac{e^x}{x^2}$ tends to infinity as x does. In particular, it is eventually more than 1. Substitute $y = x^2$ and derive that

$$\log y < \sqrt{y}$$

for large y.

2.58. Use the relation $\log(uv) = \log u + \log v$ to show that $\log\left(\frac{x}{y}\right) = \log x - \log y$.

2.6 Sequences of Functions and Their Limits

We saw in Chap. 1 that we can only rarely present numbers exactly. In general, we describe them as limits of infinite sequences of numbers. What is true of numbers is also true of functions; we can rarely describe them exactly. We often describe them as limits of sequences of functions. It is not an exaggeration to say that almost all interesting functions are defined as limits of sequences of simpler functions. Therein lies the importance of the concept of a convergent sequence of functions.

Since most of the functions we shall study are continuous, we investigate next what convergence means for sequences of continuous functions. It turns out that with the right definition of convergence, continuity is preserved under the operation of taking limits.

First we look at some simple examples.

2.6a Sequences of Functions

Example 2.30. Consider the functions

$$f_0(x) = 1,\ f_1(x) = x,\ f_2(x) = x^2,\ f_3(x) = x^3,\ \dots\ , f_n(x) = x^n,\ \dots$$

on $[0,1]$. For each x in $[0,1]$, we get the following limits as n tends to infinity:

$$\lim_{n\to\infty} f_n(x) = \begin{cases} 0 & 0 \le x < 1, \\ 1 & x = 1. \end{cases}$$

Define f to be the function on $[0,1]$ given by $f(x) = \lim_{n\to\infty} f_n(x)$. The sequence of functions f_n converges to f, a discontinuous function. See Fig. 2.36.

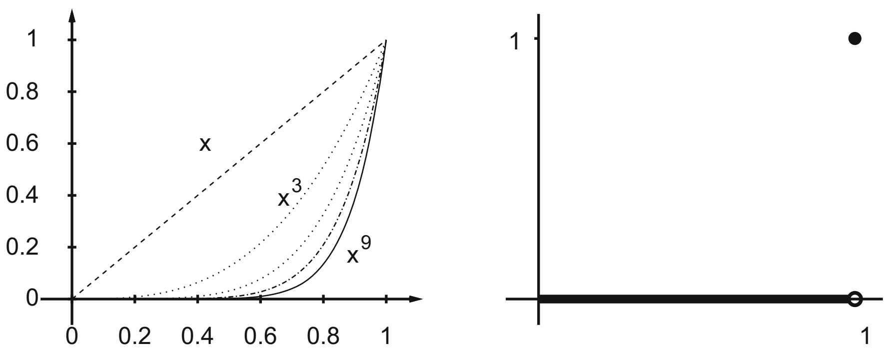

Fig. 2.36 *Left*: the functions $f_n(x) = x^n$ for $n = 1, 3, 5, 7$, and 9 are graphed on the interval $[0,1]$. *Right*: the discontinuous limit f. See Example 2.30

Example 2.30 shows that a sequence of continuous functions can converge to a discontinuous function. This is an undesirable outcome that we would like to avoid.

Example 2.31. Consider the functions $g_n(x) = x^n$ on $[0, \frac{1}{2}]$. The functions g_n are continuous on $[0, \frac{1}{2}]$ and converge to the constant function $g(x) = 0$, a continuous function.

These examples prompt us to make two definitions for sequence convergence. For a sequence of continuous functions $f_1, f_2, f_3, \ldots$ to converge to f, we certainly should require that for each x in their common domain, $\lim_{n\to\infty} f_n(x) = f(x)$.

Definition 2.11. A *sequence* of functions simply means a list $f_1, f_2, f_3, \ldots$ of functions with a common domain D. The sequence is said to *converge pointwise* to a function f on D if

$$\lim_{n\to\infty} f_n(x) = f(x) \qquad \text{for each } x \text{ in } D.$$

Uniform Convergence. We saw in Example 2.30 that a sequence of continuous functions may converge pointwise to a limit that is not continuous. We define a stronger form of convergence that avoids this trouble.

Definition 2.12. A sequence of functions $f_1, f_2, f_3, \ldots$ defined on a common domain D is said to *converge uniformly* on D to a limit function f if given any tolerance $\varepsilon > 0$, no matter how small, there is a whole number N depending on ε such that for all $n > N$, $f_n(x)$ differs from $f(x)$ by less than ε for all x in D.

To illustrate some benefits of uniform convergence, consider the problem of evaluating $f(x) = \cos x$. For instance, how would you compute $\cos(0.5)$ without using a calculator? We will see in Chap. 4 that one of the important applications of calculus is a method to generate a sequence of polynomial functions

$$p_n(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + k_n \frac{x^n}{n!} \qquad (k_n = 0,\ n \text{ odd, and } k_n = (-1)^{n/2},\ n \text{ even})$$

that converges uniformly to $\cos x$ on every closed interval $[-c, c]$. This means that once you set c and the tolerance ε, there is a polynomial p_n such that

$$|\cos x - p_n(x)| < \varepsilon \qquad \text{for all } x \text{ in } [-c, c].$$

In Chap. 4 we will see that we can get $|\cos x - p_n(x)| < \varepsilon$ for all x in $[-1, 1]$ by taking n such that $n! > \dfrac{1}{\varepsilon}$. For example, $\cos(0.3)$, $\cos(0.5)$, and $\cos(0.8)$ can each be approximated using $p_4(x) = 1 - \dfrac{x^2}{2!} + \dfrac{x^4}{4!}$, and since the convergence is uniform, the error in doing so will be less than $\frac{1}{24}$ in all cases. Evaluating $\cos x$ at an irrational number in $[-1,1]$ introduces an interesting complication. For example, $\cos\left(\frac{e}{3}\right)$ is approximated by

$$p_4\left(\frac{e}{3}\right) = 1 - \frac{1}{2}\left(\frac{e}{3}\right)^2 + \frac{1}{24}\left(\frac{e}{3}\right)^4.$$

Now we need to approximate $p_4\left(\frac{e}{3}\right)$ using some approximation to $\frac{e}{3}$, such as 0.9060939, which will introduce some error. Thinking ahead, there are many irrational numbers in $[-1,1]$ at which we would like to evaluate the cosine. Happily, p_4 is uniformly continuous on $[-1,1]$. We can find a single level of precision δ for the inputs, so that if z is within δ of x, then $p_4(z)$ is within ε of $p_4(x)$.

Looking at the big picture, we conclude that for a given tolerance ε, we can find n so large that $|\cos x - p_n(x)| < \frac{\varepsilon}{2}$ for all x in $[-1,1]$. Then we can find a precision δ such that if x and z are in $[-1,1]$ and differ by less than δ, then $p_n(x)$ and $p_n(z)$ will differ by less than $\frac{\varepsilon}{2}$. Using the triangle inequality, we get

$$|\cos x - p_n(z)| \le |\cos x - p_n(x)| + |p_n(x) - p_n(z)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Finding the right n and δ to meet a particular tolerance can be complicated, but we know in theory that it can be done. In short, approximate knowledge of the inputs and approximate knowledge of the function can be used to determine the function values within any given tolerance.[2] This is good news for computing.

Knowing that a sequence of continuous functions converges uniformly on $[a,b]$ guarantees that its limit function is continuous on $[a,b]$.

Theorem 2.11. *Let $\{f_n\}$ be a sequence of functions, each continuous on the closed interval $[a,b]$. If the sequence converges uniformly to f, then f is continuous on $[a,b]$.*

Proof. If f_n converges uniformly, then for n large enough,

$$|f_n(x) - f(x)| < \varepsilon$$

for all x in $[a,b]$. Since f_n is continuous on $[a,b]$, f_n is uniformly continuous on $[a,b]$ by Theorem 2.4. So for x_1 and x_2 close enough, say

$$|x_1 - x_2| < \delta,$$

2 There once was a function named g,
approximated closely by p.
When we put in x nearly,
we thought we'd pay dearly,
but $g(x)$ was as close as can be. –Anon.

This limerick expresses that $|g(x) - p(x_{\text{approx}})| \le |g(x) - p(x)| + |p(x) - p(x_{\text{approx}})|$.

$f_n(x_1)$ and $f_n(x_2)$ will differ by less than ε. Next we see a nice use of the triangle inequality (Sect. 1.1b). The argument is that you can control the difference of function values at two points x_1 and x_2 by writing

$$f(x_1) - f(x_2) = f(x_1) - f_n(x_1) + f_n(x_1) - f_n(x_2) + f_n(x_2) - f(x_2)$$

and grouping these terms cleverly. We have, then, by the triangle inequality that

$$|f(x_1) - f(x_2)| \le |f(x_1) - f_n(x_1)| + |f_n(x_1) - f_n(x_2)| + |f_n(x_2) - f(x_2)|$$

Each of these terms is less than ε if $|x_1 - x_2| < \delta$. This proves the uniform continuity of f on $[a,b]$. □

We now present examples of uniformly convergent sequences of continuous functions.

Example 2.32. The sequence of functions $f_n(x) = x^n$ on $[-c,c]$, where c is a positive number less than 1, converges pointwise to the function $f(x) = 0$, because for each x in $[-c,c]$, x^n tends to 0 as n tends to infinity. To see why the sequence converges uniformly to f, look at the difference between $f_n(x) = x^n$ and 0 on $[-c,c]$. For any tolerance ε, we can find a whole number N such that $c^N < \varepsilon$, and hence $c^n < \varepsilon$ for every $n > N$ as well. Let x be any number between $-c$ and c. Then

$$|f_n(x) - 0| = |x^n| \le c^n < \varepsilon.$$

Therefore, the difference between x^n and 0 is less than ε **for all** x in $[-c,c]$. That is, the sequence of functions converges uniformly. Note that the limit function, $f(x) = 0$, is continuous, as guaranteed by the theorem (Fig. 2.37).

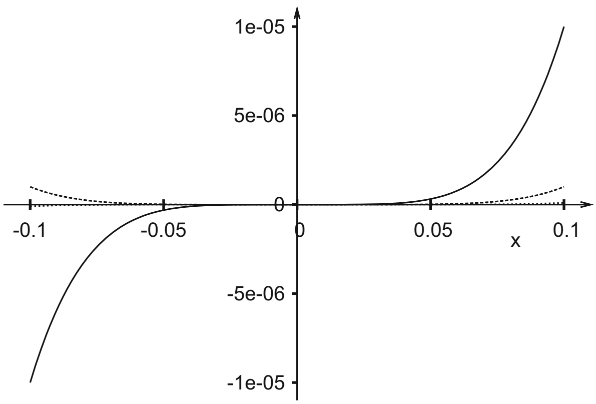

Fig. 2.37 The functions $f_n(x) = x^n$ for $n = 5$, 6, and 7 are graphed on the interval $[-0.1, 0.1]$. Note that the graph of f_7 is indistinguishable from the x-axis

Geometric Series. Consider the sequence of functions $\{f_n\}$ given by

$$f_n(x) = 1 + x + x^2 + \cdots + x^{n-1},$$

where x is in the interval $[-c, c]$ and $0 < c < 1$. The sum defining $f_n(x)$ is also given by the formula

$$f_n(x) = \frac{1 - x^n}{1 - x}.$$

For each x in $(-1, 1)$, $f_n(x)$ tends to $f(x) = \frac{1}{1-x}$, so the sequence f_n converges pointwise to f. To see why the f_n converge uniformly in $[-c, c]$, form the difference of $f_n(x)$ and $f(x)$. We get

$$f(x) - f_n(x) = \frac{x^n}{1 - x}.$$

For x in the interval $[-c, c]$, $|x|$ is not greater than c, and $|x^n|$ is not greater than c^n. It follows that

$$|f(x) - f_n(x)| = \frac{|x|^n}{1 - x} \leq \frac{c^n}{1 - c} \qquad \text{for all } x \text{ in } [-c, c].$$

Since c^n tends to zero, we can choose N so large that for n greater than N, $\frac{c^n}{1-c}$ is less than ε, and hence $f(x)$ differs from $f_n(x)$ by less than ε for all x in $[-c, c]$. This proves that f_n tends to f uniformly on the interval $[-c, c]$, $c < 1$. Note that $\frac{1}{1-x}$ is continuous on $[-c, c]$, as guaranteed by the theorem (Fig. 2.38).

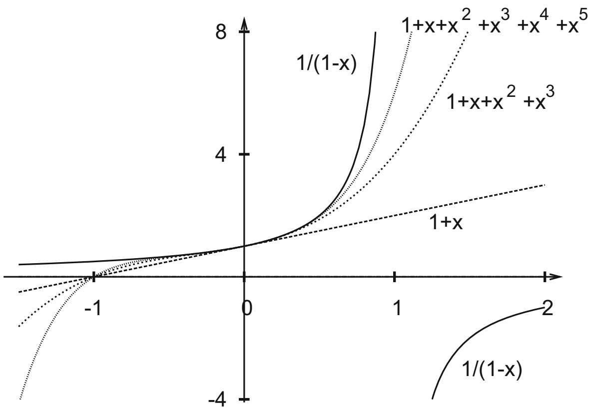

Fig. 2.38 The sequence of functions $f_n(x) = 1 + x + \cdots + x^n$ converges uniformly to $\frac{1}{1-x}$ on $[-c, c]$ when $c < 1$

Operations on Convergent Sequences of Functions. We can combine uniformly convergent sequences of continuous functions.

Theorem 2.12. *Suppose f_n and g_n are uniformly convergent sequences of continuous functions on $[a,b]$, converging to f and g. Then*

(a) $f_n + g_n$ converges uniformly to $f + g$.
(b) $f_n g_n$ converges uniformly to fg.
(c) If $f \neq 0$ on $[a,b]$, then for n large enough, $f_n \neq 0$ and $\dfrac{1}{f_n}$ tends to $\dfrac{1}{f}$ uniformly.
(d) If h is a continuous function with range contained in $[a,b]$, then $g_n \circ h$ converges uniformly to $g \circ h$.
(e) If k is a continuous function on a closed interval that contains the range of each g_n and g, then $k \circ g_n$ converges uniformly to $k \circ g$.

Proof. We give an outline of the proof of this theorem. For (a), use the triangle inequality:

$$|(f(x)+g(x)) - (f_n(x)+g_n(x))| \leq |f(x) - f_n(x)| + |g(x) - g_n(x)|.$$

For all x in $[a,b]$, the terms on the right are smaller than any given tolerance, provided that n is large enough. Figure 2.39 shows the idea. We guide you through the details of proving part (a) in Problem 2.61.

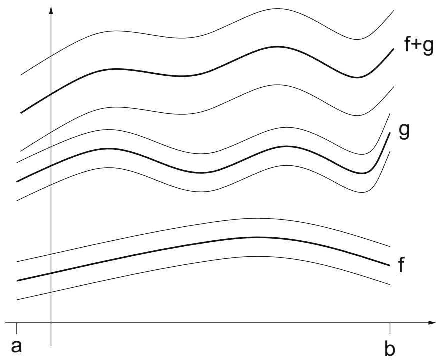

Fig. 2.39 Functions f_n are within ε of f for $n > N_1$, and the g_n are within ε of g for $n > N_2$. The sums $f_n + g_n$ are then within 2ε of $f + g$ for n larger than both N_1 and N_2

For (b), use

$$|f(x)g(x) - f_n(x)g_n(x)| = |(f(x) - f_n(x))g(x) + f_n(x)(g(x) - g_n(x))|$$
$$\leq |f(x) - f_n(x)||g(x)| + |f_n(x)||g(x) - g_n(x)|.$$

We can make the factors $|f(x) - f_n(x)|$ and $|g(x) - g_n(x)|$ small by taking n large. We check the factor $|f_n(x)|$. By the extreme value theorem (Theorem 2.6), $|f|$ has a maximum value M, so $-M \leq f(x) \leq M$. Since the f_n converge to f uniformly, they are within distance 1 of f for large n, and $-M-1 \leq f_n(x) \leq M+1$ for all x. Thus we have

$$|f(x)g(x) - f_n(x)g_n(x)| \leq |f(x) - f_n(x)||g(x)| + (M+1)|g(x) - g_n(x)|$$

for large n, and this can be made arbitrarily small by taking n sufficiently large.

For (c): If f is not zero on an interval, then it is either positive at every point or negative at every point. For if it were positive at some point c and negative at another point d, then according to Theorem 2.5, the intermediate value theorem, $f(x)$ would be zero at some point x between c and d, contrary to our assumption about f. Take the case that f is positive. According to Theorem 2.6, the extreme value theorem, $f(x)$ takes on its minimum at some point of the closed interval $[a,b]$. This minimum is a positive number m, and $f(x) \geq m$ for all x in the interval. Since $f_n(x)$ tends uniformly to $f(x)$ on the interval, it follows that for n greater than some number N, $f_n(x)$ differs from $f(x)$ by less than $\frac{1}{2}m$. Since $f(x) \geq m$, $f_n(x) \geq \frac{1}{2}m$. We use

$$\frac{1}{f_n(x)} - \frac{1}{f(x)} = \frac{f(x) - f_n(x)}{f_n(x)f(x)}.$$

The right-hand side is not more than $\dfrac{|f(x) - f_n(x)|}{\left(\frac{1}{2}m\right)m}$ in absolute value, from which the result follows.

For (d) we use that $g(y) - g_n(y)$ is uniformly small for all y, and then take $y = h(x)$ to see that $g(h(x)) - g_n(h(x))$ is uniformly small for all x.

For (e) we use that $g(x) - g_n(x)$ is uniformly small for all x, and then use uniform continuity of k to see that $k(g(x)) - k(g_n(x))$ is uniformly small for all x.

This completes the outline of the proof. □

The beauty of Theorem 2.12 is that it allows us to construct a large variety of uniformly convergent sequences of functions. Here are a few examples.

Example 2.33. Let $g_n(x) = 1 + x + x^2 + \cdots + x^n$, and let $h(u) = -u^2$, where u is in $[-c,c]$, and $0 < c < 1$. Then

$$g_n(h(u)) = 1 - u^2 + u^4 - u^6 + \cdots + (-u^2)^n$$

converges uniformly in $[-c,c]$ to $\dfrac{1}{1+u^2}$.

Example 2.34. Let $r > 0$, a any number, and set

$$k_n(x) = 1 + \frac{x-a}{r} + \cdots + \left(\frac{x-a}{r}\right)^n.$$

Then $k_n(x) = g_n\left(\frac{x-a}{r}\right)$, where g_n is as in Example 2.33. The k_n converge uniformly to

$$\frac{1}{1 - \frac{x-a}{r}} = \frac{r}{r-x+a}$$

on every closed interval contained in $(a-r, a+r)$. This is true by part (d) of Theorem 2.12.

Example 2.35. Let $h(t) = \frac{1}{2}\cos t$, where $g_n(x)$ is as in Example 2.33. Then

$$g_n(h(t)) = 1 + \frac{1}{2}\cos t + \left(\frac{1}{2}\cos t\right)^2 + \cdots + \left(\frac{1}{2}\cos t\right)^n$$

converges uniformly to $\dfrac{2}{2-\cos t}$ for all t.

2.6b Series of Functions

Definition 2.13. The sequence of functions $\{f_n\}$ can be added to make a new sequence $\{s_n\}$, called the sequence of *partial sums* of $\{f_n\}$:

$$s_n = f_0 + f_1 + f_2 + \cdots + f_n = \sum_{j=0}^{n} f_j.$$

The sequence of functions $\{s_n\}$ is called a *series* and is denoted by

$$\sum_{j=0}^{\infty} f_j.$$

If $\lim_{n\to\infty} s_n(x)$ exists, denote it by $f(x)$, and we say that the series converges to $f(x)$ at x. We write

$$\sum_{j=0}^{\infty} f_j(x) = f(x).$$

If the sequence of partial sums converges uniformly on D, we say that the series converges uniformly on D.

We saw earlier that the sequence of partial sums of the geometric series

$$s_n(x) = 1 + x + x^2 + \cdots + x^n = \frac{1 - x^{n+1}}{1 - x}$$

converges uniformly to $\frac{1}{1-x}$ on every interval $[-c, c]$, if $0 < c < 1$. We often write

$$\sum_{k=0}^{\infty} x^k = 1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x} \qquad (|x| < 1).$$

This series is of a special kind, a power series.

Definition 2.14. A *power series* is a series of the form

$$\sum_{k=0}^{\infty} a_k (x-a)^k.$$

The numbers a_n are called the coefficients. The number a is called the center of the power series.

Consider the power series

$$\sum_{n=1}^{\infty} \frac{x^n}{n} = x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots.$$

For what values of x, if any, does the series converge? To find all values of x for which the series converges, we use the ratio test, Theorem 1.18. We compute the limit

$$\lim_{n\to\infty} \left| \frac{\frac{x^{n+1}}{n+1}}{\frac{x^{n+1}}{n+1}} \right| = \lim_{n\to\infty} |x| \frac{n+1}{n} = |x|.$$

According to the ratio test, if the limit is less than 1, then the series converges absolutely. Therefore, $\sum_{n=1}^{\infty} \frac{x^n}{n}$ converges for $|x| < 1$. Also, if the limit is greater than 1, then the series diverges, in this case for $|x| > 1$. The test gives no information when the limit is 1, in our case $|x| = 1$. So our next task is to investigate the convergence (or divergence) of $\sum_{n=1}^{\infty} \frac{x^n}{n}$ when $x = 1$ and when $x = -1$. At $x = 1$, we get $\sum_{n=1}^{\infty} \frac{1}{n}$, the well-known harmonic series. We saw in Example 1.21 that it diverges. At $x = -1$ we get the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$. It converges by the alternating series theorem, Theorem 1.17.

Therefore, $\sum_{n=1}^{\infty} \frac{x^n}{n}$ converges pointwise for all x in $[-1,1)$. We have not shown that the convergence is uniform, so we do not know whether the function $f(x) = \sum_{n=1}^{\infty} \frac{x^n}{n}$ is continuous.

Sometimes a sequence of functions converges to a function that we know by another rule. If so, we know a great deal about that limit function. But this is not always the case. Some sequences of functions, including power series, converge to functions that we know only through sequential approximation. The next two theorems give us important information about the limit function of a power series. The first tells us about its domain. The second tells us about its continuity.

Theorem 2.13. *For a power series* $\sum_{n=0}^{\infty} c_n(x-a)^n$*, one of the following must hold:*

(a) The series converges absolutely for every x*.*
(b) The series converges only at $x = a$*.*
(c) There is a positive number R*, called the radius of convergence, such that the series converges absolutely for* $|x-a| < R$ *and diverges for* $|x-a| > R$*.*

In case (c), the series might or might not converge at $x = a-R$ *and at* $x = a+R$*.*

Proof. Let us first point out that if the series converges at some $x_0 \neq a$, then it converges absolutely for every x that is closer to a, that is, $|x-a| < |x_0-a|$. Here is why: The convergence of $\sum_{n=0}^{\infty} c_n(x_0-a)^n$ implies that the terms $c_n(x_0-a)^n$ tend to 0. In particular, there is an N such that $|c_n(x_0-a)^n| < 1$ for all $n > N$. If $0 < |x-a| < |x_0-a|$, set $r = \frac{|x_0-a|}{|x-a|}$. Then $r < 1$, and we get

$$\sum_{n=N+1}^{\infty} |c_n(x-a)^n| = \sum_{n=N+1}^{\infty} |c_n(x-a)^n| \left| \frac{(x_0-a)^n}{(x_0-a)^n} \right| \tag{2.11}$$

$$= \sum_{n=N+1}^{\infty} |c_n(x_0-a)^n| \left| \frac{(x-a)^n}{(x_0-a)^n} \right| \leq \sum_{n=N+1}^{\infty} r^n.$$

Therefore, $\sum_{n=0}^{\infty} c_n(x-a)^n$ converges absolutely by comparison with a geometric series.

Now consider the three possibilities we have listed in the theorem. It might happen that the series converges for every x. If so, it converges absolutely for every x by what we have just shown. This covers the first case.

The other possibility is that the series converges for some number x_0, but not for every number. If there is only one such x_0, then it must be a, since the series

$$c_0 + c_1(a-a) + c_2(a-a)^2 + \cdots = c_0$$

certainly converges. This covers the second case.

Finally, there may be an $x_0 \neq a$ for which the series converges, though the series does not converge for every number. We will use the least upper bound principle, Theorems 1.2 and 1.3, to describe R. Let S be the set of numbers x for which the series converges. Then S is not empty, because a and x_0 are in S, as well as every number closer to a than x_0. Also, S is bounded, because if there were arbitrarily large (positive or negative) numbers in S, then all numbers closer to a would be in S, i.e., S would be all the numbers. Therefore, S has a least upper bound M and a greatest lower bound m, which means that if

$$m < x < M,$$

then the series converges at x. We ask you in Problem 2.65 to show that m and M are the same distance from a:

$$m < a < M \qquad \text{and} \quad a - m = M - a$$

and that the convergence is absolute in (m, M). Set $R = M - a$. This concludes the proof. □

Theorem 2.14. *A power series* $\sum_{n=0}^{\infty} a_n(x-a)^n$ *converges uniformly to its limit function on every closed interval* $|x-a| \le r$*, where* r *is less than the radius of convergence* R*.*

In particular, the limit function is continuous in $(a-R, a+R)$*.*

Proof. If the radius of convergence of $\sum_{n=0}^{\infty} a_n(x-a)^n = f(x)$ is $R = 0$, the series converges at only one point, $x = a$. The series is then just $f(a) = a_0 + 0 + \cdots$, which converges uniformly on that domain.

Suppose $R > 0$ or R is infinite, and take any positive $r < R$. Then the number $a + r$ is in the interval of convergence, so according to Theorem 2.13, $\sum_{n=0}^{\infty} a_n r^n$ converges absolutely. Then for every x with $|x-a| \le r$,

$$\left| f(x) - \sum_{n=0}^{k} |a_n(x-a)^n| \right| \le \sum_{n=k+1}^{\infty} |a_n(x-a)^n| \le \sum_{n=k+1}^{\infty} |a_n r^n|.$$

The last expression is independent of x and tends to 0 as k tends to infinity. Therefore, f is the uniform limit of its partial sums, which are continuous, on $|x-a| \le r$. According to Theorem 2.11, f is continuous on $[a-r, a+r]$.

Since every point of $(a-R, a+R)$ is contained in such a closed interval, f is continuous on $(a-R, a+R)$. □

The radius of convergence, R, of a power series can often be found by the ratio test. If that fails, there is another test, called the root test, which we describe in Problem 2.67.

Example 2.36. To find the interval of convergence of $\sum_{n=0}^{\infty} 2^n (x-3)^n$, we use the ratio test:

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

When $2|x-3| < 1$, the series converges absolutely. When $2|x-3| > 1$, the series diverges. What happens when $2|x-3| = 1$?

(a) At $x = 2.5$, $2(x-3) = -1$, and $\sum_{n=0}^{\infty} 2^n (x-3)^n = \sum_{n=0}^{\infty} (-1)^n$ diverges.

(b) At $x = 3.5$, $2(x-3) = 1$, and $\sum_{n=0}^{\infty} 2^n (x-3)^n = \sum_{n=0}^{\infty} 1^n$ diverges.

Conclusion: $f(x) = \sum_{n=0}^{\infty} 2^n (x-3)^n$ converges for all x with $2|x-3| < 1$, i.e., in $(2.5, 3.5)$. Also, according to Theorem 2.14, the series converges uniformly to f on every closed interval $|x-3| \le r < \frac{1}{2}$, and f is continuous on $(2.5, 3.5)$.

Example 2.37. To find the interval of convergence of $\sum_{n=0}^{\infty} \frac{x^n}{n!}$, we use the ratio test:

$$\lim_{n\to\infty} \left| \frac{\frac{x^{n+1}}{(n+1)!}}{\frac{x^n}{n!}} \right| = \lim_{n\to\infty} \frac{|x|}{n+1} = 0 < 1.$$

Since $0 < 1$ for all x, the series converges for all x. It converges uniformly on every closed interval $|x-0| \le r$. So $f(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ is continuous on $(-\infty, \infty)$. In Chap. 4, we will see that this power series converges to a function that we know by another rule.

2.6c Approximating the Functions $\sqrt{x}$ *and* e^x

We close this section by looking at three examples of sequences of functions $\{f_n\}$ that are not power series that converge uniformly to the important functions $\sqrt{x}$, $|x|$, and e^x. In the case of e^x, we use the sequence of continuous functions $e_n(x) = \left(1 + \frac{x}{n}\right)^n$, and thus we prove that e^x is a continuous function.

Approximating $\sqrt{x}$. In Sect. 1.3a, we constructed a sequence of approximations $s_1, s_2, s_3, \ldots$ that converged to the square root of 2. There is nothing special about the number 2. The same construction can be used to generate a sequence of numbers that tends to the square root of any positive number x. Here is how:

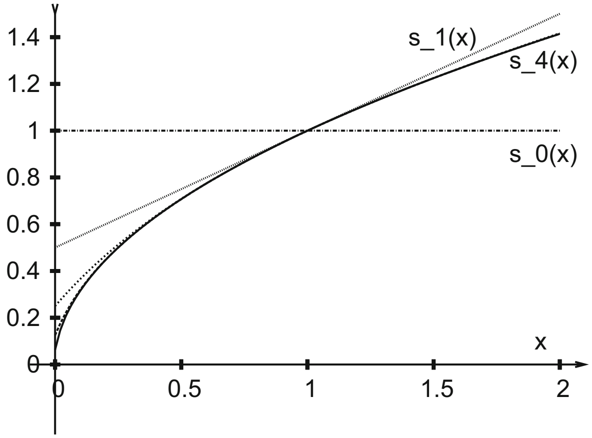

Fig. 2.40 The functions $s_n(x)$ converge to $\sqrt{x}$. The cases $0 \le n \le 4$ are shown. Note that $\sqrt{x}$ is not plotted

Suppose s is an approximation to the square root of x. To find a better approximation, we note that the product of s and $\frac{x}{s}$ is x. If s happens to be larger than $\frac{x}{s}$, then $s^2 > s\frac{x}{s} = x > \left(\frac{x}{s}\right)^2$, so $s > \sqrt{x} > \frac{x}{s}$, that is, the square root of x lies between these two. A similar argument shows that $\frac{x}{s} > \sqrt{x} > s$ if s happens to be less than $\frac{x}{s}$. So we take as the next approximation the arithmetic mean of the two:

$$\text{new approximation } = \frac{1}{2}\left(s + \frac{x}{s}\right).$$

Rather than start with an arbitrary first approximation, we start with $s_0 = 1$ and construct a sequence of approximations $s_1, s_2, \ldots$ as follows:

$$s_{n+1} = \frac{1}{2}\left(s_n + \frac{x}{s_n}\right).$$

The approximations s_n depend on the number x whose square root we seek; in other words, s_n is a function of x. How much does s_{n+1} differ from $\sqrt{x}$?

$$s_{n+1} - \sqrt{x} = \frac{1}{2}\left(s_n + \frac{x}{s_n}\right) - \sqrt{x}.$$

We bring the fractions on the right to a common denominator:

$$s_{n+1} - \sqrt{x} = \frac{1}{2s_n}(s_n^2 + x - 2s_n\sqrt{x}). \tag{2.12}$$

The expression in parentheses on the right is a perfect square, $(s_n - \sqrt{x})^2$. So we can rewrite Eq. (2.12) as

$$s_{n+1} - \sqrt{x} = \frac{1}{2s_n}(s_n - \sqrt{x})^2, \qquad (n \geq 0). \tag{2.13}$$

This formula implies that s_{n+1} is greater than $\sqrt{x}$ except when $s_n = \sqrt{x}$.

Since the denominator s_n on the right in Eq. (2.13) is greater than $s_n - \sqrt{x}$, we deduce that

$$s_{n+1} - \sqrt{x} < \frac{1}{2}(s_n - \sqrt{x}).$$

Applying this inequality n times, we get

$$s_{n+1} - \sqrt{x} < \frac{1}{2^n}(s_1 - \sqrt{x}) = \left(\frac{1}{2}\right)^n \left(\frac{1+x}{2} - \sqrt{x}\right). \tag{2.14}$$

Note that in Eq. (2.14), the factor $\frac{1+x}{2} - \sqrt{x}$ is less than $\frac{1+c}{2}$ whenever $x \leq c$. Therefore, inequality (2.14) implies

$$s_{n+1}(x) - \sqrt{x} \leq \frac{1+c}{2^n}.$$

It follows that the sequence of functions $s_n(x)$ converges uniformly to the function $\sqrt{x}$ over every finite interval $[0, c]$ of the positive axis (Fig. 2.40). The rate of convergence is even faster than what we have proved here, as we discuss in Sect. 5.3c.

Example 2.38. We show how to approximate $f(x) = |x|$ by a sequence of rational functions. Let $f_n(x) = s_n(x^2)$, where s_n is the sequence of functions derived in the preceding example that converge to $\sqrt{x}$. The $s_n(x)$ converge uniformly to $\sqrt{x}$, and x^2 is continuous on every closed interval. By Theorem 2.12, $s_n(x^2)$ converges uniformly to $\sqrt{x^2} = |x|$.

We indicate in Fig. 2.41 the graphs of $s_2(x^2)$, $s_3(x^2)$, and $s_5(x^2)$, which are rational approximations to $|x|$.

Approximating e^x. Take the functions $e_n(x) = \left(1 + \frac{x}{n}\right)^n$. We shall show, with your help, that they converge uniformly to the function e^x over every finite interval $[-c, c]$ (Fig. 2.42).

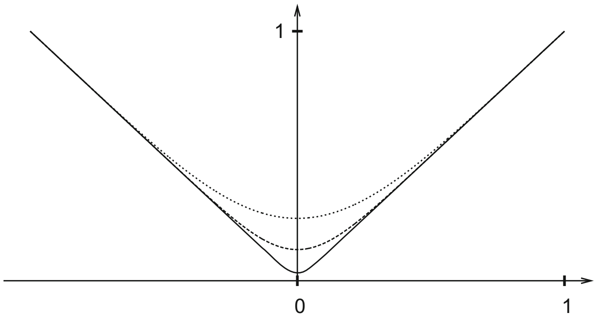

Fig. 2.41 Rational approximations of $|x|$ in Example 2.38

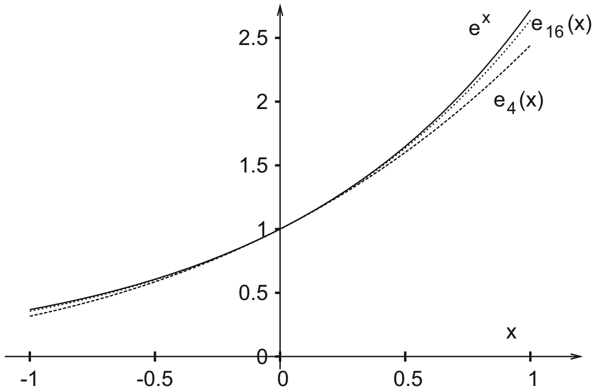

Fig. 2.42 The exponential function e^x and the functions $e_n(x) = \left(1+\frac{x}{n}\right)^n$ for $n = 4$ and $n = 16$ are graphed on the interval $[-1,1]$

Let us return to Sect. 1.4. There, we showed that the sequence of numbers $e_n = \left(1+\dfrac{1}{n}\right)^n$ is increasing and bounded, and therefore, by the monotone convergence theorem, it has a limit, a number that we have denoted by e.

We can show by similar arguments (see Problems 2.73 and 2.74) that for every positive x, the sequence of numbers $e_n(x)$ is increasing and bounded, whence by the monotone convergence theorem, it converges pointwise to a number $e(x)$ that depends on x. Note that $e_n(1) = e_n$, so $e(1) = \mathrm{e}$.

It remains to show that the limit function $e(x)$ is the exponential function e^x, and that convergence is uniform over every finite interval. To do this, we first show that $e(x) = \mathrm{e}^x$ when x is rational. We do this by showing that

$$e(r+s) = e(r)e(s)$$

for every pair of positive rational numbers r and s. We know from Sect. 2.5c that this relation implies that $e(x)$ is an exponential function for rational numbers.

Let r and s be any positive rational numbers. We can find a common denominator d such that

$$r = \frac{p}{d}, \qquad s = \frac{q}{d}$$

and p, q, and d are positive whole numbers. By manipulating $r+s$ algebraically, we obtain

$$e(r+s) = e\left(\frac{p}{d} + \frac{q}{d}\right) = e\left(\frac{1}{d}(p+q)\right).$$

We claim that

$$e(kx) = \big(e(x)\big)^k \tag{2.15}$$

for positive integers k. Here is a proof of the claim. Since $\left(1+\frac{x}{n}\right)^n$ converges to $e(x)$, for every positive integer k, $\left(1+\frac{kx}{n}\right)^n$ converges to $e(kx)$. Set $n = km$; we get that $\left(1+\frac{kx}{km}\right)^{km} = \left(1+\frac{x}{m}\right)^{mk}$ tends to $e(x)^k$. This proves Eq. (2.15).

Set $x = 1/d$ and $k = p+q$ in Eq. (2.15). We get

$$e\left(\frac{1}{d}(p+q)\right) = \left(e\left(\frac{1}{d}\right)\right)^{p+q} = \left(e\left(\frac{1}{d}\right)\right)^p \left(e\left(\frac{1}{d}\right)\right)^q = e\left(\frac{p}{d}\right)e\left(\frac{q}{d}\right) = e(r)e(s).$$

This concludes the proof that $e(x)$ is an exponential function a^x for x rational. Since $e(1) = \mathrm{e}$, it follows that $e(x) = \mathrm{e}^x$.

We turn now to showing that $e_n(x)$ converges *uniformly* to $e(x)$ on every finite interval $[-c, c]$. Our proof that the sequence $e_n(x)$ converges for every x as n tends to infinity used the monotone convergence theorem. Unfortunately, this gives no information as to how fast these sequences converge, and therefore it is useless in proving the uniformity of convergence. We will show that

$$\text{if} \quad -c \le x \le c, \ \text{ then } \ e(x) - e_n(x) < \frac{k}{n},$$

for some constant k that depends on c. This is sufficient to prove the uniform convergence.

We make use of the following inequality:

$$a^n - b^n < (a-b)na^n \qquad \text{if} \quad 1 < b < a. \tag{2.16}$$

First we prove the inequality: We start from the observation that for all a and b,

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \cdots + b^{n-1}),$$

which we see by carrying out the multiplication on the right-hand side. Then in the case $0 < b < a$, we have for each power that $b^k < a^k$, so in the factor $(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \cdots + b^{n-1})$, there are n terms each less than a^{n-1}. This proves that $a^n - b^n < (a-b)na^{n-1}$. In the case $1 < a$, we may append one more factor of a, and this proves the inequality. We will use this inequality twice in two different ways to show uniform convergence.

Since $e_n(x)$ is an increasing sequence,

$$e(x) \geq \left(1+\frac{x}{n}\right)^n.$$

Take the nth root of this inequality and use Eq. (2.15) with $k=n$ to express the nth root of $e(x)$. We get

$$e\left(\frac{x}{n}\right) = \big(e(x)\big)^{\frac{1}{n}} \geq 1+\frac{x}{n} \geq 1.$$

The first use of inequality (2.16) will be to show that for $n > x$,

$$e\left(\frac{x}{n}\right) < \frac{1}{1-\frac{x}{n}}. \tag{2.17}$$

Set $a = 1+\dfrac{x}{n}$ and $b=1$ in Eq. (2.16). We get

$$a^n - b^n = e_n(x) - 1 < (a-b)na^n = \frac{x}{n}n\left(1+\frac{x}{n}\right)^n = xe_n(x).$$

Letting n tend to infinity, we get in the limit $e(x)-1 < xe(x)$, or $(1-x)e(x) < 1$. Thus if $x < 1$, then $1-x$ is positive, and we get $e(x) < \dfrac{1}{1-x}$. But if $n > x$, then $\dfrac{x}{n} < 1$, whence $e\big(\frac{x}{n}\big) < \dfrac{1}{1-\frac{x}{n}}$. This proves Eq. (2.17).

For the second use of inequality (2.16), set $a = e\big(\frac{x}{n}\big)$ and $b = 1+\dfrac{x}{n}$. We get

$$e(x)-e_n(x) = \left(e(\tfrac{x}{n})\right)^n - \left(1+\frac{x}{n}\right)^n = a^n - b^n \leq (a-b)na^n$$

$$= \left(e\left(\frac{x}{n}\right) - \left(1+\frac{x}{n}\right)\right) n \left(e(\tfrac{x}{n})\right)^n = \left(e\left(\frac{x}{n}\right) - \left(1+\frac{x}{n}\right)\right) ne(x). \tag{2.18}$$

Combining the two results, set Eq. (2.17) into the right side of Eq. (2.18) to get

$$e(x)-e_n(x) < \left(\frac{1}{1-\frac{x}{n}} - \left(1+\frac{x}{n}\right)\right) ne(x) = \left(\frac{\frac{x^2}{n^2}}{1-\frac{x}{n}}\right) ne(x). \tag{2.19}$$

So for n greater than x,

$$e(x)-e_n(x) \leq \frac{1}{n}\frac{x^2 e(x)}{1-\frac{x}{n}}. \tag{2.20}$$

For $n > 2x$, the denominator on the right in Eq. (2.20) is greater than $\frac{1}{2}$, so

$$e(x)-e_n(x) < \frac{1}{n}2e(x)x^2 < \frac{2}{n}e(c)c^2$$

for every x in $[-c, c]$. This shows that as n tends to infinity, $e_n(x)$ tends to $e(x)$ uniformly on every finite x-interval. This concludes the proof. □

Example 2.39. We know that $g_n(x) = \left(1 + \frac{x}{n}\right)^n$ converges uniformly to e^x for x in any interval $[a, b]$. By Theorem 2.12, then,

(a) $\left(1 + \frac{x^2}{n}\right)^n = g_n(x^2)$ converges uniformly to e^{x^2};
(b) $\left(1 - \frac{x}{n}\right)^n = g_n(-x)$ converges uniformly to e^{-x};
(c) $\log(g_n(x)) = n \log\left(1 + \frac{x}{n}\right)$ converges uniformly to $\log(e^x) = x$.

Problems

2.59. Use the identity $1 + x + x^2 + x^3 + x^4 = \frac{1 - x^5}{1 - x}$ to estimate the accuracy of the approximation

$$1 + x + x^2 + x^3 + x^4 \approx \frac{1}{1 - x}$$

on $-\frac{1}{2} \le x \le \frac{1}{2}$.

2.60. In this problem, we explore another geometric meaning for geometric series. Refer to Fig. 2.43, where a line is drawn from the top point of the unit circle through the point (x, y) in the first quadrant of the circle. The point z where the line hits the axis is called the *stereographic projection* of the point (x, y). The shaded triangles are all similar. Justify the following statements.

(a) $z = \frac{x}{1 - y}$.
(b) The height of the nth triangle is y times the height of the $(n-1)$st triangle.
(c) z is the sum of the series $z = x + xy + xy^2 + xy^3 + \cdots = \frac{x}{1 - y}$.

2.61. We gave an outline of the proof of part (a) of Theorem 2.12. Let us fill in the details.

(a) Explain why

$$\left|f(x) + g(x) - \left(f_n(x) + g_n(x)\right)\right| \le |f(x) - f_n(x)| + |g(x) - g_n(x)|$$

for all x.
(b) Explain why given any tolerance $\varepsilon > 0$, there is an N_1 such that $|f(x) - f_n(x)| < \frac{\varepsilon}{2}$ for all x when $n > N_1$, and why there is an N_2 such that $|g(x) - g_n(x)| < \frac{\varepsilon}{2}$ for all x when $n > N_2$.

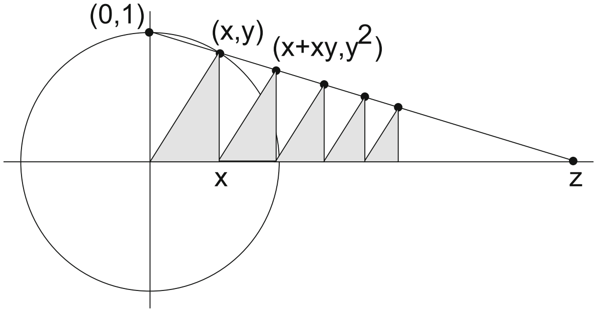

Fig. 2.43 Stereographic projection is the sum of a geometric series

(c) Explain why given any tolerance $\varepsilon > 0$, there is a number N such that

$$|f(x) - f_n(x)| + |g(x) - g_n(x)| < \varepsilon$$

for all x whenever $n > N$.

(d) Explain why given any tolerance $\varepsilon > 0$, there is a number N such that

$$\left| f(x) + g(x) - \big(f_n(x) + g_n(x)\big) \right| < \varepsilon$$

for all x whenever $n > N$.

(e) Explain why $f_n + g_n$ converges uniformly to $f + g$.

2.62. Use Theorem 2.12 to find an interval $a \le t \le b$ on which the convergence

$$1 + e^{-t} + e^{-2t} + e^{-3t} + \cdots = \frac{1}{1 - e^{-t}}$$

is uniform.

2.63. A power series $f(x) = \sum_{n=0}^{\infty} a_n(x-2)^n$ is known to converge at $x = 4$. At what other values of x must it converge? Find the largest open interval on which we can be sure that f is continuous.

2.64. For each pair of series, which one has the larger radius of convergence? In two cases, they have the same radius.

(a) $\sum_{n=0}^{\infty} x^n$ or $\sum_{n=0}^{\infty} 3^n x^n$

(b) $\sum_{n=0}^{\infty} x^n$ or $\sum_{n=0}^{\infty} \frac{x^n}{n!}$

(c) $\sum_{n=0}^{\infty} n(x-2)^n$ or $\sum_{n=0}^{\infty} (x-3)^n$

(d) $1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$ or $\frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots$

2.65. Fill in the missing step that we have indicated in the proof of Theorem 2.13.

2.66. Which of these series represent a continuous function on (at least) $[-1,1]$?

(a) $\sum_{n=0}^{\infty} x^n$
(b) $\sum_{n=0}^{\infty} \left(\frac{1}{10}\right)^n x^n$
(c) $\sum_{n=0}^{\infty} \left(\frac{1}{10}\right)^n (x-2)^n$
(d) $1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\cdots$

2.67. Consider a power series $\sum_{n=0}^{\infty} a_n x^n$. Suppose the limit $L = \lim_{n\to\infty} |a_n|^{1/n}$ exists and is positive. Justify the following steps, which prove that $1/L$ is the radius of convergence of the series. This is the *root test*.

(a) Let $\sum_{n=0}^{\infty} p_n$ be a series of positive numbers for which $\lim_{n\to\infty} p_n^{1/n} = \ell$ exists and $\ell < 1$. Show that there is a number r, $0 < \ell < r < 1$, such that for N large enough, $p_n < r^n$, $n > N$. Conclude that $\sum_{n=0}^{\infty} p_n$ converges.
(b) Let $\sum_{n=0}^{\infty} p_n$ be a series of positive numbers for which $\lim_{n\to\infty} p_n^{1/n} = \ell$ exists and $\ell > 1$. Show that there is a number r, $1 < r < \ell$, such that for N large enough, $p_n > r^n$, $n > N$. Conclude that $\sum_{n=0}^{\infty} p_n$ diverges.
(c) Taking $p_n = |a_n x^n|$ for different choices of x, show that $1/L$ is the radius of convergence of $\sum_{n=0}^{\infty} a_n x^n$.

2.68. Suppose $\{p_n\}$ is a positive sequence whose partial sums $p_1 + \cdots + p_n$ are less than nL for some number L. Use the root test (Problem 2.67) to show that the series $\sum_{n=1}^{\infty} (p_1 p_2 p_3 \cdots p_n) x^n$ converges in $|x| < 1/L$.

2.69. Suppose the root test (Problem 2.67) indicates that a series $\sum_{n=0}^{\infty} a_n x^n$ has radius of convergence R. Show that according to the root test, $\sum_{n=0}^{\infty} n a_n x^n$ also has radius of convergence R. (See Problem 1.53.)

2.70. For each of the following series, determine (i) the values of x for which the series converges; (ii) the largest open interval on which the sum is continuous.

(a) $\sum_{n=0}^{\infty} \frac{x^n}{2^n}$
(b) $\sum_{n=0}^{\infty} \frac{(x-3)^{2n}}{(2n)!}$
(c) $\sum_{n=0}^{\infty} \sqrt{n}x^n$
(d) $\sum_{n=0}^{\infty} \left(\frac{x^n}{2^n} + \sqrt{n}x^n\right)$
(e) $\sum_{n=1}^{\infty} \frac{2^n+7^n}{3^n+5^n}x^n$

2.71. For some of the following series it is possible to give an algebraic formula for the function to which the series converges. In those cases, give such a formula, and state the domain of the function where possible.

(a) $1 - t^2 + t^4 - t^6 + \cdots$
(b) $\sum_{n=3}^{\infty} x^n$ Note the 3.
(c) $\sum_{n=0}^{\infty} \sqrt{n}x^n$
(d) $\sum_{n=0}^{\infty} \left(\frac{t^n}{2^n} + 3^n t^{2n}\right)$

2.72. Our sequence of functions $s_n(x)$ approximating $\sqrt{x}$ was defined recursively. Write explicit expressions for $s_2(x)$ and $s_3(x)$, and verify that they are rational functions.

2.73. Use the method explained in Sect. 1.4 to show that for each $x > 0$, the sequence $e_n(x) = \left(1 + \frac{x}{n}\right)^n$ is increasing.

2.74. Show that for each $x > 0$, the sequence $\{e_n(x)\}$ is bounded. *Hint*: For $x < 2$, $e_n(x) < \left(1 + \frac{2}{n}\right)^n$. Set $n = 2m$ to conclude that $e_m(x) < e^2$.

2.75. Find a sequence of functions that converges to e^{-x} on every interval $[a,b]$ by composing the sequence $e_n(x) = \left(1 + \frac{x}{n}\right)^n$ with a continuous function.