1 Functions

OVERVIEW Functions are fundamental to the study of calculus. In this chapter we review what functions are and how they are pictured as graphs, how they are combined and transformed, and ways they can be classified. We review the trigonometric functions, and we discuss misrepresentations that can occur when using calculators and computers to obtain a function's graph. We also discuss inverse, exponential, and logarithmic functions. The real number system, Cartesian coordinates, straight lines, circles, parabolas, and ellipses are reviewed in the Appendices.

1.1 Functions and Their Graphs

Functions are a tool for describing the real world in mathematical terms. A function can be represented by an equation, a graph, a numerical table, or a verbal description; we will use all four representations throughout this book. This section reviews these function ideas.

Functions; Domain and Range

The temperature at which water boils depends on the elevation above sea level (the boiling point drops as you ascend). The interest paid on a cash investment depends on the length of time the investment is held. The area of a circle depends on the radius of the circle. The distance an object travels at constant speed along a straight-line path depends on the elapsed time.

In each case, the value of one variable quantity, say y, depends on the value of another variable quantity, which we might call x. We say that "y is a function of x" and write this symbolically as

$$y = f(x) \qquad (\text{“}y \text{ equals } f \text{ of } x\text{”}).$$

In this notation, the symbol f represents the function, the letter x is the **independent variable** representing the input value of f, and y is the **dependent variable** or output value of f at x.

> **DEFINITION** A **function** f from a set D to a set Y is a rule that assigns a *unique* (single) element $f(x) \in Y$ to each element $x \in D$.

The set D of all possible input values is called the **domain** of the function. The set of all output values of $f(x)$ as x varies throughout D is called the **range** of the function. The range may not include every element in the set Y. The domain and range of a function can be any sets of objects, but often in calculus they are sets of real numbers interpreted as points of a coordinate line. (In Chapters 13–16, we will encounter functions for which the elements of the sets are points in the coordinate plane or in space.)

Often a function is given by a formula that describes how to calculate the output value from the input variable. For instance, the equation $A = \pi r^2$ is a rule that calculates the area A of a circle from its radius r (so r, interpreted as a length, can only be positive in this formula). When we define a function $y = f(x)$ with a formula and the domain is not stated explicitly or restricted by context, the domain is assumed to be the largest set of real x-values for which the formula gives real y-values, which is called the **natural domain**. If we want to restrict the domain in some way, we must say so. The domain of $y = x^2$ is the entire set of real numbers. To restrict the domain of the function to, say, positive values of x, we would write "$y = x^2, x > 0$."

Changing the domain to which we apply a formula usually changes the range as well. The range of $y = x^2$ is $[0, \infty)$. The range of $y = x^2, x \geq 2$, is the set of all numbers obtained by squaring numbers greater than or equal to 2. In set notation (see Appendix 1), the range is $\{x^2 | x \geq 2\}$ or $\{y | y \geq 4\}$ or $[4, \infty)$.

FIGURE 1.1 A diagram showing a function as a kind of machine.

D = domain set $\quad$ Y = set containing the range

FIGURE 1.2 A function from a set D to a set Y assigns a unique element of Y to each element in D.

When the range of a function is a set of real numbers, the function is said to be **real-valued**. The domains and ranges of most real-valued functions of a real variable we consider are intervals or combinations of intervals. The intervals may be open, closed, or half open, and may be finite or infinite. Sometimes the range of a function is not easy to find.

A function f is like a machine that produces an output value $f(x)$ in its range whenever we feed it an input value x from its domain (Figure 1.1). The function keys on a calculator give an example of a function as a machine. For instance, the $\sqrt{x}$ key on a calculator gives an output value (the square root) whenever you enter a nonnegative number x and press the $\sqrt{x}$ key.

A function can also be pictured as an **arrow diagram** (Figure 1.2). Each arrow associates an element of the domain D with a unique or single element in the set Y. In Figure 1.2, the arrows indicate that $f(a)$ is associated with a, $f(x)$ is associated with x, and so on. Notice that a function can have the same *value* at two different input elements in the domain (as occurs with $f(a)$ in Figure 1.2), but each input element x is assigned a *single* output value $f(x)$.

EXAMPLE 1 Let's verify the natural domains and associated ranges of some simple functions. The domains in each case are the values of x for which the formula makes sense.

Function	Domain (x)	Range (y)
$y = x^2$	$(-\infty, \infty)$	$[0, \infty)$
$y = 1/x$	$(-\infty, 0) \cup (0, \infty)$	$(-\infty, 0) \cup (0, \infty)$
$y = \sqrt{x}$	$[0, \infty)$	$[0, \infty)$
$y = \sqrt{4 - x}$	$(-\infty, 4]$	$[0, \infty)$
$y = \sqrt{1 - x^2}$	$[-1, 1]$	$[0, 1]$

Solution The formula $y = x^2$ gives a real y-value for any real number x, so the domain is $(-\infty, \infty)$. The range of $y = x^2$ is $[0, \infty)$ because the square of any real number is nonnegative and every nonnegative number y is the square of its own square root, $y = \left(\sqrt{y}\right)^2$ for $y \geq 0$.

The formula $y = 1/x$ gives a real y-value for every x except $x = 0$. For consistency in the rules of arithmetic, *we cannot divide any number by zero*. The range of $y = 1/x$, the set of reciprocals of all nonzero real numbers, is the set of all nonzero real numbers, since $y = 1/(1/y)$. That is, for $y \neq 0$ the number $x = 1/y$ is the input assigned to the output value y.

The formula $y = \sqrt{x}$ gives a real y-value only if $x \geq 0$. The range of $y = \sqrt{x}$ is $[0, \infty)$ because every nonnegative number is some number's square root (namely, it is the square root of its own square).

In $y = \sqrt{4 - x}$, the quantity $4 - x$ cannot be negative. That is, $4 - x \geq 0$, or $x \leq 4$. The formula gives real y-values for all $x \leq 4$. The range of $\sqrt{4 - x}$ is $[0, \infty)$, the set of all nonnegative numbers.

To find out, we could plot more points. But how would we then connect *them*? The basic question still remains: How do we know for sure what the graph looks like between the points we plot? Calculus answers this question, as we will see in Chapter 4. Meanwhile, we will have to settle for plotting points and connecting them as best we can.

Representing a Function Numerically

We have seen how a function may be represented algebraically by a formula (the area function) and visually by a graph (Example 2). Another way to represent a function is **numerically**, through a table of values. Numerical representations are often used by engineers and experimental scientists. From an appropriate table of values, a graph of the function can be obtained using the method illustrated in Example 2, possibly with the aid of a computer. The graph consisting of only the points in the table is called a **scatterplot**.

EXAMPLE 3 Musical notes are pressure waves in the air. The data associated with Figure 1.6 give recorded pressure displacement versus time in seconds of a musical note produced by a tuning fork. The table provides a representation of the pressure function over time. If we first make a scatterplot and then connect approximately the data points (t, p) from the table, we obtain the graph shown in the figure.

Time	Pressure	Time	Pressure
0.00091	−0.080	0.00362	0.217
0.00108	0.200	0.00379	0.480
0.00125	0.480	0.00398	0.681
0.00144	0.693	0.00416	0.810
0.00162	0.816	0.00435	0.827
0.00180	0.844	0.00453	0.749
0.00198	0.771	0.00471	0.581
0.00216	0.603	0.00489	0.346
0.00234	0.368	0.00507	0.077
0.00253	0.099	0.00525	−0.164
0.00271	−0.141	0.00543	−0.320
0.00289	−0.309	0.00562	−0.354
0.00307	−0.348	0.00579	−0.248
0.00325	−0.248	0.00598	−0.035
0.00344	−0.041		

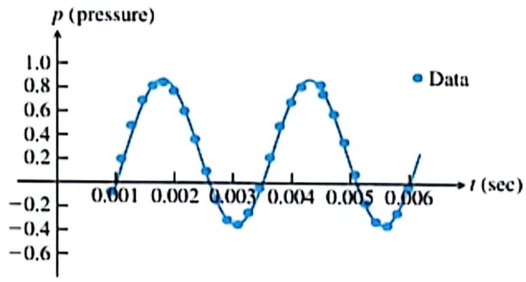


FIGURE 1.6 A smooth curve through the plotted points gives a graph of the pressure function represented by the accompanying tabled data (Example 3).

The Vertical Line Test for a Function

Not every curve in the coordinate plane can be the graph of a function. A function f can have only one value $f(x)$ for each x in its domain, so *no vertical* line can intersect the graph of a function more than once. If a is in the domain of the function f, then the vertical line $x = a$ will intersect the graph of f at the single point $(a, f(a))$.

A circle cannot be the graph of a function, since some vertical lines intersect the circle twice. The circle graphed in Figure 1.7a, however, does contain the graphs of functions of x, such as the upper semicircle defined by the function $f(x) = \sqrt{1 - x^2}$ and the lower semicircle defined by the function $g(x) = -\sqrt{1 - x^2}$ (Figures 1.7b and 1.7c).

The formula $y = \sqrt{1 - x^2}$ gives a real y-value for every x in the closed interval from -1 to 1. Outside this domain, $1 - x^2$ is negative and its square root is not a real number. The values of $1 - x^2$ vary from 0 to 1 on the given domain, and the square roots of these values do the same. The range of $\sqrt{1 - x^2}$ is $[0, 1]$. ■

Graphs of Functions

If f is a function with domain D, its **graph** consists of the points in the Cartesian plane whose coordinates are the input-output pairs for f. In set notation, the graph is

$$\{(x, f(x)) \mid x \in D\}.$$

The graph of the function $f(x) = x + 2$ is the set of points with coordinates (x, y) for which $y = x + 2$. Its graph is the straight line sketched in Figure 1.3.

The graph of a function f is a useful picture of its behavior. If (x, y) is a point on the graph, then $y = f(x)$ is the height of the graph above (or below) the point x. The height may be positive or negative, depending on the sign of $f(x)$ (Figure 1.4).

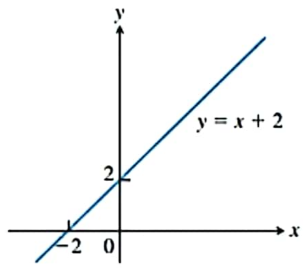


FIGURE 1.3 The graph of $f(x) = x + 2$ is the set of points (x, y) for which y has the value $x + 2$.

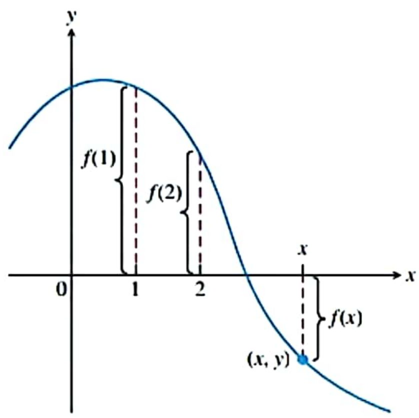


FIGURE 1.4 If (x, y) lies on the graph of f, then the value $y = f(x)$ is the height of the graph above the point x (or below x if $f(x)$ is negative).

EXAMPLE 2 Graph the function $y = x^2$ over the interval $[-2, 2]$.

Solution Make a table of xy-pairs that satisfy the equation $y = x^2$. Plot the points (x, y) whose coordinates appear in the table, and draw a *smooth* curve (labeled with its equation) through the plotted points (see Figure 1.5). ■

x	$y = x^2$
-2	4
-1	1
0	0
1	1
$\frac{3}{2}$	$\frac{9}{4}$
2	4

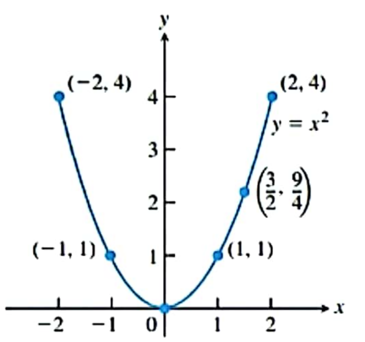
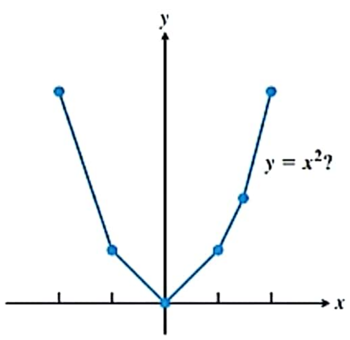
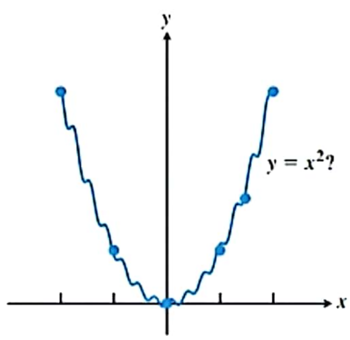
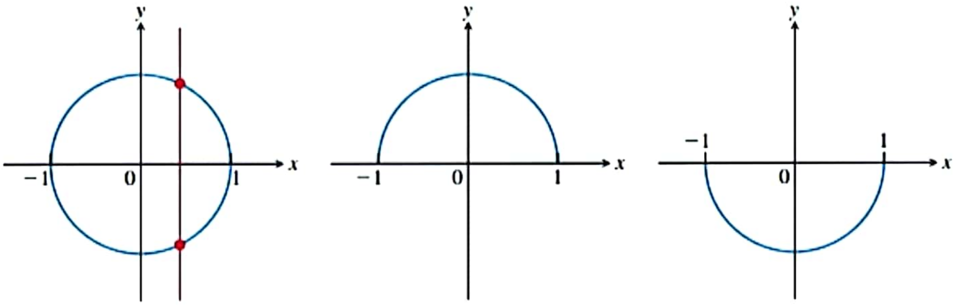


FIGURE 1.5 Graph of the function in Example 2.

How do we know that the graph of $y = x^2$ doesn't look like one of these curves?







(a) $x^2 + y^2 = 1$ (b) $y = \sqrt{1 - x^2}$ (c) $y = -\sqrt{1 - x^2}$

FIGURE 1.7 (a) The circle is not the graph of a function; it fails the vertical line test. (b) The upper semicircle is the graph of a function $f(x) = \sqrt{1 - x^2}$. (c) The lower semicircle is the graph of a function $g(x) = -\sqrt{1 - x^2}$.

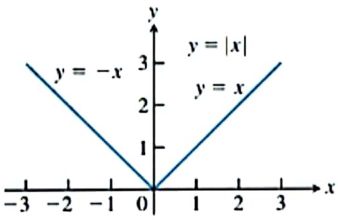


FIGURE 1.8 The absolute value function has domain $(-\infty, \infty)$ and range $[0, \infty)$.

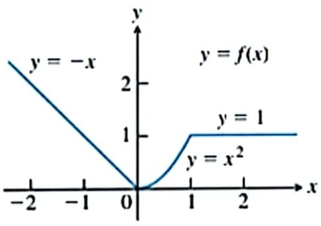


FIGURE 1.9 To graph the function $y = f(x)$ shown here, we apply different formulas to different parts of its domain (Example 4).

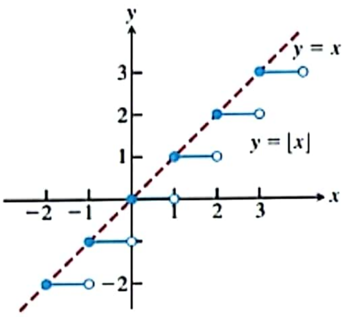


FIGURE 1.10 The graph of the greatest integer function $y = \lfloor x \rfloor$ lies on or below the line $y = x$, so it provides an integer floor for x (Example 5).

Piecewise-Defined Functions

Sometimes a function is described in pieces by using different formulas on different parts of its domain. One example is the **absolute value function**

$$|x| = \begin{cases} x, & x \geq 0 \qquad \text{First formula} \\ -x, & x < 0, \qquad \text{Second formula} \end{cases}$$

whose graph is given in Figure 1.8. The right-hand side of the equation means that the function equals x if $x \geq 0$, and equals $-x$ if $x < 0$. Piecewise-defined functions often arise when real-world data are modeled. Here are some other examples.

EXAMPLE 4 The function

$$f(x) = \begin{cases} -x, & x < 0 \qquad \text{First formula} \\ x^2, & 0 \leq x \leq 1 \qquad \text{Second formula} \\ 1, & x > 1 \qquad \text{Third formula} \end{cases}$$

is defined on the entire real line but has values given by different formulas, depending on the position of x. The values of f are given by $y = -x$ when $x < 0$, $y = x^2$ when $0 \leq x \leq 1$, and $y = 1$ when $x > 1$. The function, however, is *just one function* whose domain is the entire set of real numbers (Figure 1.9). ■

EXAMPLE 5 The function whose value at any number x is the *greatest integer less than or equal to* x is called the **greatest integer function** or the **integer floor function**. It is denoted $\lfloor x \rfloor$. Figure 1.10 shows the graph. Observe that

$$\lfloor 2.4 \rfloor = 2, \quad \lfloor 1.9 \rfloor = 1, \quad \lfloor 0 \rfloor = 0, \quad \lfloor -1.2 \rfloor = -2,$$
$$\lfloor 2 \rfloor = 2, \quad \lfloor 0.2 \rfloor = 0, \quad \lfloor -0.3 \rfloor = -1, \quad \lfloor -2 \rfloor = -2.$$ ■

EXAMPLE 6 The function whose value at any number x is the *smallest integer greater than or equal to* x is called the **least integer function** or the **integer ceiling function**. It is denoted $\lceil x \rceil$. Figure 1.11 shows the graph. For positive values of x, this function might represent, for example, the cost of parking x hours in a parking lot that charges \$1 for each hour or part of an hour. ■

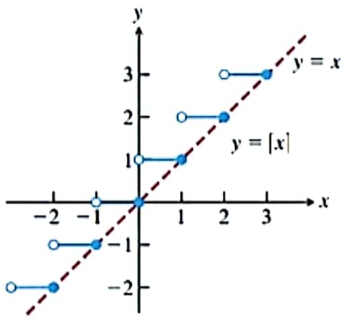


FIGURE 1.11 The graph of the least integer function $y = \lceil x \rceil$ lies on or above the line $y = x$, so it provides an integer ceiling for x (Example 6).

Increasing and Decreasing Functions

If the graph of a function *climbs* or *rises* as you move from left to right, we say that the function is *increasing*. If the graph *descends* or *falls* as you move from left to right, the function is *decreasing*.

> **DEFINITIONS** Let f be a function defined on an interval I and let x_1 and x_2 be any two points in I.
>
> **1.** If $f(x_2) > f(x_1)$ whenever $x_1 < x_2$, then f is said to be **increasing** on I.
>
> **2.** If $f(x_2) < f(x_1)$ whenever $x_1 < x_2$, then f is said to be **decreasing** on I.

It is important to realize that the definitions of increasing and decreasing functions must be satisfied for *every* pair of points x_1 and x_2 in I with $x_1 < x_2$. Because we use the inequality $<$ to compare the function values, instead of $\leq$, it is sometimes said that f is *strictly* increasing or decreasing on I. The interval I may be finite (also called bounded) or infinite (unbounded) and by definition never consists of a single point (Appendix 1).

EXAMPLE 7 The function graphed in Figure 1.9 is decreasing on $(-\infty, 0]$ and increasing on $[0, 1]$. The function is neither increasing nor decreasing on the interval $[1, \infty)$ because of the strict inequalities used to compare the function values in the definitions. ■

Even Functions and Odd Functions: Symmetry

The graphs of *even* and *odd* functions have characteristic symmetry properties.

> **DEFINITIONS** A function $y = f(x)$ is an
>
> **even function of x** if $f(-x) = f(x)$,
>
> **odd function of x** if $f(-x) = -f(x)$,
>
> for every x in the function's domain.

The names *even* and *odd* come from powers of x. If y is an even power of x, as in $y = x^2$ or $y = x^4$, it is an even function of x because $(-x)^2 = x^2$ and $(-x)^4 = x^4$. If y is an odd power of x, as in $y = x$ or $y = x^3$, it is an odd function of x because $(-x)^1 = -x$ and $(-x)^3 = -x^3$.

The graph of an even function is **symmetric about the y-axis**. Since $f(-x) = f(x)$, a point (x, y) lies on the graph if and only if the point $(-x, y)$ lies on the graph (Figure 1.12a). A reflection across the y-axis leaves the graph unchanged.

The graph of an odd function is **symmetric about the origin**. Since $f(-x) = -f(x)$, a point (x, y) lies on the graph if and only if the point $(-x, -y)$ lies on the graph (Figure 1.12b). Equivalently, a graph is symmetric about the origin if a rotation of 180° about the origin leaves the graph unchanged. Notice that the definitions imply that both x and $-x$ must be in the domain of f.

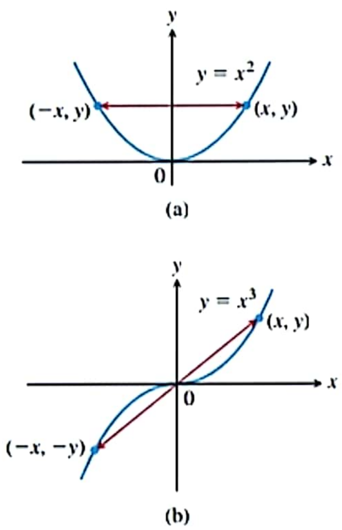


FIGURE 1.12 (a) The graph of $y = x^2$ (an even function) is symmetric about the y-axis. (b) The graph of $y = x^3$ (an odd function) is symmetric about the origin.

EXAMPLE 8 Here are several functions illustrating the definition.

$f(x) = x^2$	Even function: $(-x)^2 = x^2$ for all x; symmetry about y-axis.
$f(x) = x^2 + 1$	Even function: $(-x)^2 + 1 = x^2 + 1$ for all x; symmetry about y-axis (Figure 1.13a).
$f(x) = x$	Odd function: $(-x) = -x$ for all x; symmetry about the origin.
$f(x) = x + 1$	Not odd: $f(-x) = -x + 1$, but $-f(x) = -x - 1$. The two are not equal. Not even: $(-x) + 1 \neq x + 1$ for all $x \neq 0$ (Figure 1.13b). ■

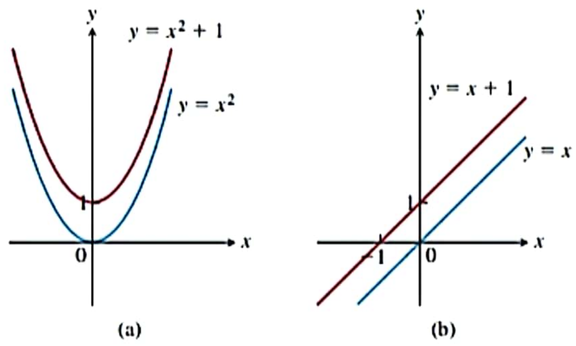


FIGURE 1.13 (a) When we add the constant term 1 to the function $y = x^2$, the resulting function $y = x^2 + 1$ is still even and its graph is still symmetric about the y-axis. (b) When we add the constant term 1 to the function $y = x$, the resulting function $y = x + 1$ is no longer odd, since the symmetry about the origin is lost. The function $y = x + 1$ is also not even (Example 8).

Common Functions

A variety of important types of functions are frequently encountered in calculus. We identify and briefly describe them here.

Linear Functions A function of the form $f(x) = mx + b$, for constants m and b, is called a **linear function**. Figure 1.14a shows an array of lines $f(x) = mx$ where $b = 0$, so these lines pass through the origin. The function $f(x) = x$ where $m = 1$ and $b = 0$ is called the **identity function**. Constant functions result when the slope $m = 0$ (Figure 1.14b). A linear function with positive slope whose graph passes through the origin is called a *proportionality* relationship.

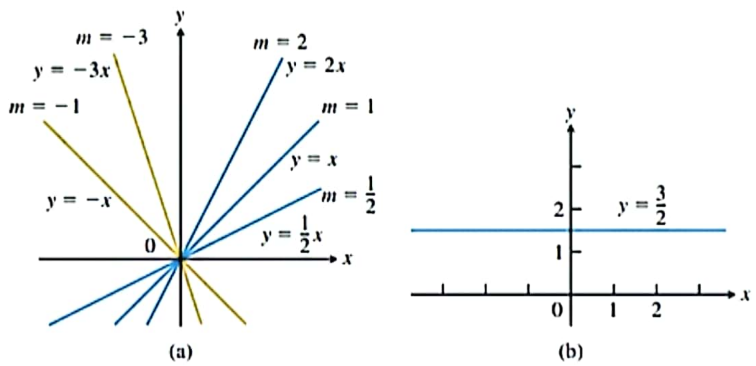


FIGURE 1.14 (a) Lines through the origin with slope m. (b) A constant function with slope $m = 0$.

> **DEFINITION** Two variables y and x are **proportional** (to one another) if one is always a constant multiple of the other; that is, if $y = kx$ for some nonzero constant k.

If the variable y is proportional to the reciprocal $1/x$, then sometimes it is said that y is **inversely proportional** to x (because $1/x$ is the multiplicative inverse of x).

Power Functions A function $f(x) = x^a$, where a is a constant, is called a **power function**. There are several important cases to consider.

(a) $a = n$, a positive integer.

The graphs of $f(x) = x^n$, for $n = 1, 2, 3, 4, 5$, are displayed in Figure 1.15. These functions are defined for all real values of x. Notice that as the power n gets larger, the curves tend to flatten toward the x-axis on the interval $(-1, 1)$, and to rise more steeply for $|x| > 1$. Each curve passes through the point $(1, 1)$ and through the origin. The graphs of functions with even powers are symmetric about the y-axis; those with odd powers are symmetric about the origin. The even-powered functions are decreasing on the interval $(-\infty, 0]$ and increasing on $[0, \infty)$; the odd-powered functions are increasing over the entire real line $(-\infty, \infty)$.

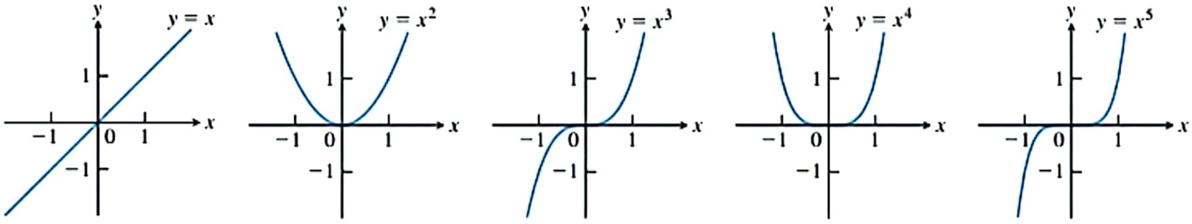


FIGURE 1.15 Graphs of $f(x) = x^n$, $n = 1, 2, 3, 4, 5$, defined for $-\infty < x < \infty$.

(b) $a = -1$ or $a = -2$.

The graphs of the functions $f(x) = x^{-1} = 1/x$ and $g(x) = x^{-2} = 1/x^2$ are shown in Figure 1.16. Both functions are defined for all $x \neq 0$ (you can never divide by zero). The graph of $y = 1/x$ is the hyperbola $xy = 1$, which approaches the coordinate axes far from the origin. The graph of $y = 1/x^2$ also approaches the coordinate axes. The graph of the function f is symmetric about the origin; f is decreasing on the intervals $(-\infty, 0)$ and $(0, \infty)$. The graph of the function g is symmetric about the y-axis; g is increasing on $(-\infty, 0)$ and decreasing on $(0, \infty)$.

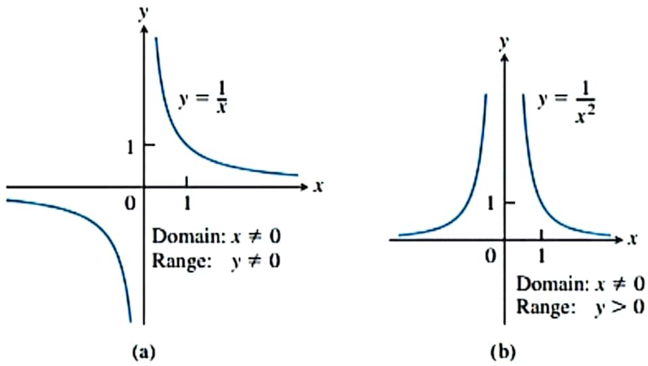


FIGURE 1.16 Graphs of the power functions $f(x) = x^a$ for part (a) $a = -1$ and for part (b) $a = -2$.

(c) $a = \frac{1}{2}, \frac{1}{3}, \frac{3}{2}$, and $\frac{2}{3}$.

The functions $f(x) = x^{1/2} = \sqrt{x}$ and $g(x) = x^{1/3} = \sqrt[3]{x}$ are the **square root** and **cube root** functions, respectively. The domain of the square root function is $[0, \infty)$, but the cube root function is defined for all real x. Their graphs are displayed in Figure 1.17, along with the graphs of $y = x^{3/2}$ and $y = x^{2/3}$. (Recall that $x^{3/2} = (x^{1/2})^3$ and $x^{2/3} = (x^{1/3})^2$.)

Polynomials A function p is a **polynomial** if

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

where n is a nonnegative integer and the numbers $a_0, a_1, a_2, \ldots, a_n$ are real constants (called the **coefficients** of the polynomial). All polynomials have domain $(-\infty, \infty)$. If the

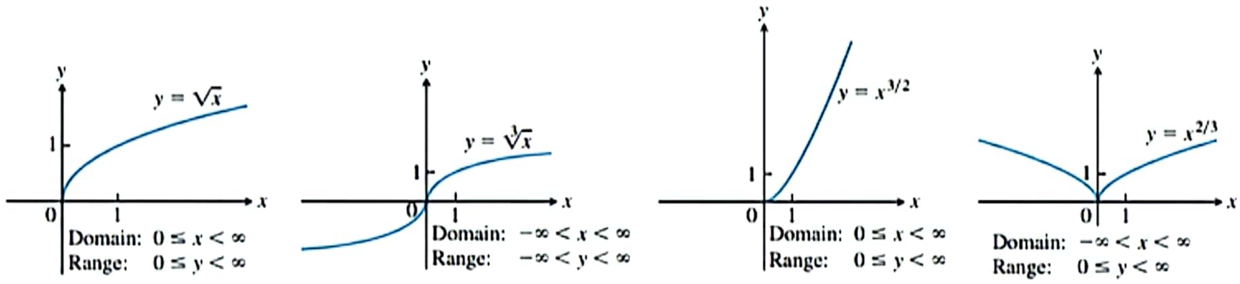


FIGURE 1.17 Graphs of the power functions $f(x) = x^a$ for $a = \frac{1}{2}, \frac{1}{3}, \frac{3}{2}$, and $\frac{2}{3}$.

leading coefficient $a_n \neq 0$ and $n > 0$, then n is called the **degree** of the polynomial. Linear functions with $m \neq 0$ are polynomials of degree 1. Polynomials of degree 2, usually written as $p(x) = ax^2 + bx + c$, are called **quadratic functions.** Likewise, **cubic functions** are polynomials $p(x) = ax^3 + bx^2 + cx + d$ of degree 3. Figure 1.18 shows the graphs of three polynomials. Techniques to graph polynomials are studied in Chapter 4.

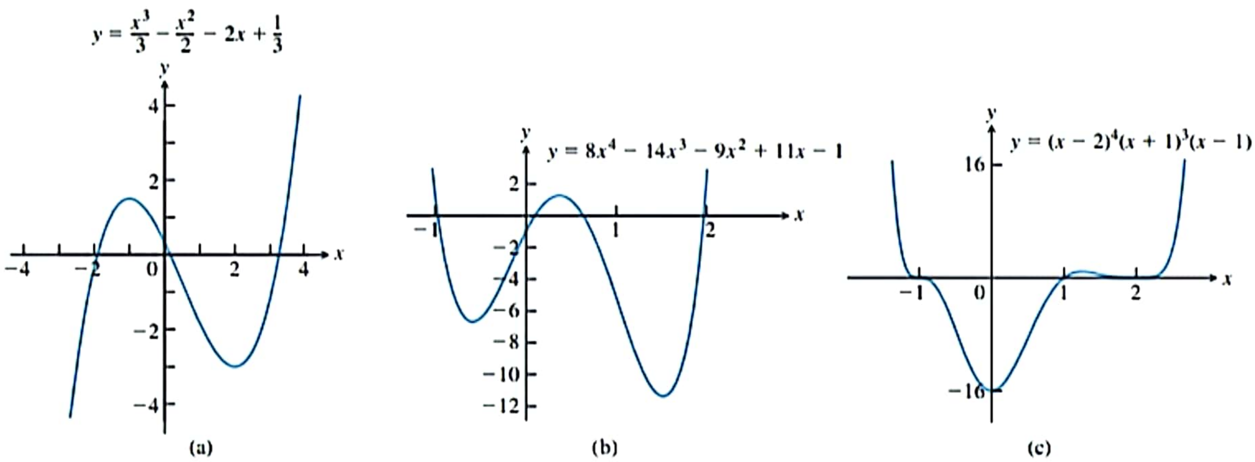


FIGURE 1.18 Graphs of three polynomial functions.

Rational Functions A **rational function** is a quotient or ratio $f(x) = p(x)/q(x)$, where p and q are polynomials. The domain of a rational function is the set of all real x for which $q(x) \neq 0$. The graphs of several rational functions are shown in Figure 1.19.

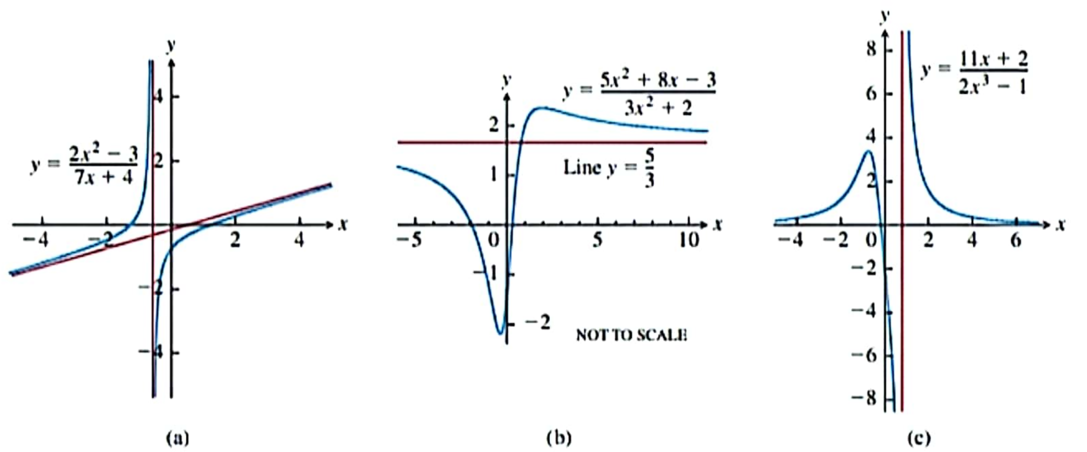
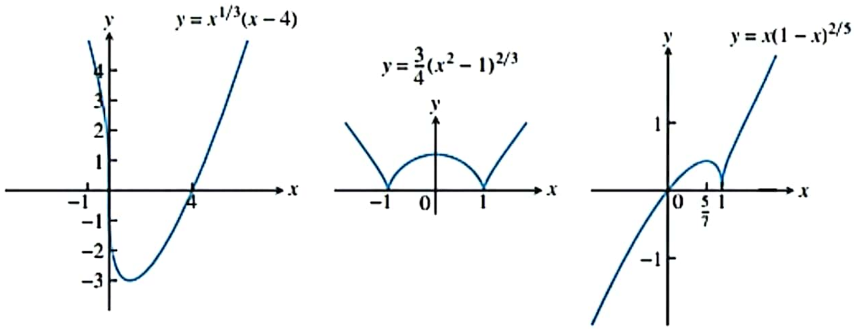


FIGURE 1.19 Graphs of three rational functions. The straight red lines approached by the graphs are called *asymptotes* and are not part of the graphs. We discuss asymptotes in Section 2.6.

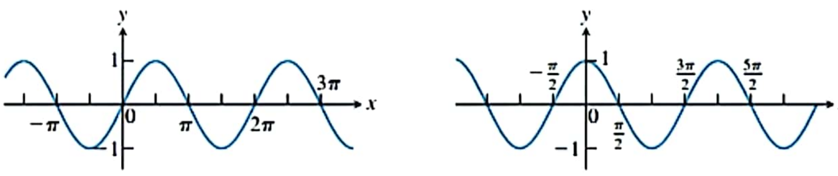
Algebraic Functions Any function constructed from polynomials using algebraic operations (addition, subtraction, multiplication, division, and taking roots) lies within the class of **algebraic functions**. All rational functions are algebraic, but also included are more complicated functions (such as those satisfying an equation like $y^3 - 9xy + x^3 = 0$, studied in Section 3.7). Figure 1.20 displays the graphs of three algebraic functions.



(a) (b) (c)

FIGURE 1.20 Graphs of three algebraic functions.

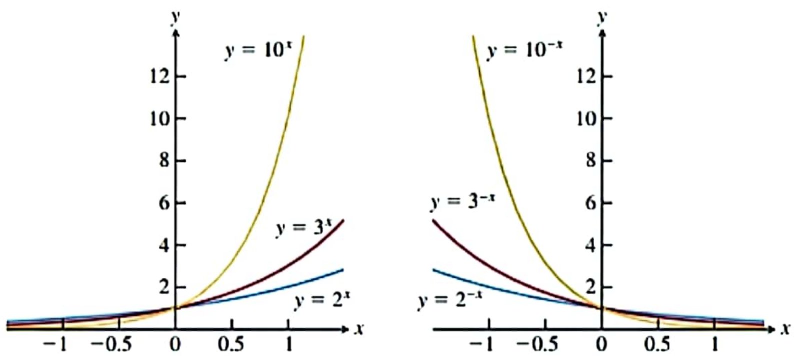
Trigonometric Functions The six basic trigonometric functions are reviewed in Section 1.3. The graphs of the sine and cosine functions are shown in Figure 1.21.



(a) $f(x) = \sin x$ (b) $f(x) = \cos x$

FIGURE 1.21 Graphs of the sine and cosine functions.

Exponential Functions Functions of the form $f(x) = a^x$, where the base $a > 0$ is a positive constant and $a \neq 1$, are called **exponential functions**. All exponential functions have domain $(-\infty, \infty)$ and range $(0, \infty)$, so an exponential function never assumes the value 0. We discuss exponential functions in Section 1.5. The graphs of some exponential functions are shown in Figure 1.22.



(a) (b)

FIGURE 1.22 Graphs of exponential functions.

1.2 Combining Functions; Shifting and Scaling Graphs

In this section we look at the main ways functions are combined or transformed to form new functions.

Sums, Differences, Products, and Quotients

Like numbers, functions can be added, subtracted, multiplied, and divided (except where the denominator is zero) to produce new functions. If f and g are functions, then for every x that belongs to the domains of both f and g (that is, for $x \in D(f) \cap D(g)$), we define functions $f + g$, $f - g$, and fg by the formulas

$$(f + g)(x) = f(x) + g(x)$$
$$(f - g)(x) = f(x) - g(x)$$
$$(fg)(x) = f(x)g(x).$$

Notice that the + sign on the left-hand side of the first equation represents the operation of addition of *functions*, whereas the + on the right-hand side of the equation means addition of the real numbers $f(x)$ and $g(x)$.

At any point of $D(f) \cap D(g)$ at which $g(x) \neq 0$, we can also define the function f/g by the formula

$$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)} \qquad (\text{where } g(x) \neq 0).$$

Functions can also be multiplied by constants: If c is a real number, then the function cf is defined for all x in the domain of f by

$$(cf)(x) = cf(x).$$

EXAMPLE 1 The functions defined by the formulas

$$f(x) = \sqrt{x} \qquad \text{and} \qquad g(x) = \sqrt{1 - x}$$

have domains $D(f) = [0, \infty)$ and $D(g) = (-\infty, 1]$. The points common to these domains are the points

$$[0, \infty) \cap (-\infty, 1] = [0, 1].$$

The following table summarizes the formulas and domains for the various algebraic combinations of the two functions. We also write $f \cdot g$ for the product function fg.

Function	Formula	Domain
$f + g$	$(f + g)(x) = \sqrt{x} + \sqrt{1 - x}$	$[0, 1] = D(f) \cap D(g)$
$f - g$	$(f - g)(x) = \sqrt{x} - \sqrt{1 - x}$	$[0, 1]$
$g - f$	$(g - f)(x) = \sqrt{1 - x} - \sqrt{x}$	$[0, 1]$
$f \cdot g$	$(f \cdot g)(x) = f(x)g(x) = \sqrt{x(1 - x)}$	$[0, 1]$
f/g	$\frac{f}{g}(x) = \frac{f(x)}{g(x)} = \sqrt{\frac{x}{1 - x}}$	$[0, 1)$ ($x = 1$ excluded)
g/f	$\frac{g}{f}(x) = \frac{g(x)}{f(x)} = \sqrt{\frac{1 - x}{x}}$	$(0, 1]$ ($x = 0$ excluded)

The graph of the function $f + g$ is obtained from the graphs of f and g by adding the corresponding y-coordinates $f(x)$ and $g(x)$ at each point $x \in D(f) \cap D(g)$, as in Figure 1.25. The graphs of $f + g$ and $f \cdot g$ from Example 1 are shown in Figure 1.26.

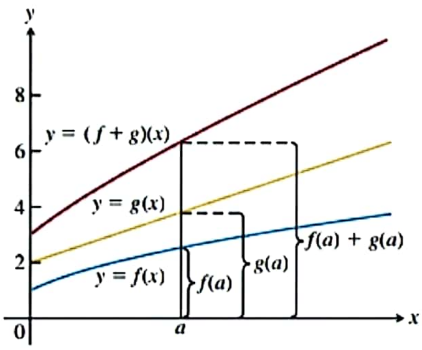


FIGURE 1.25 Graphical addition of two functions.

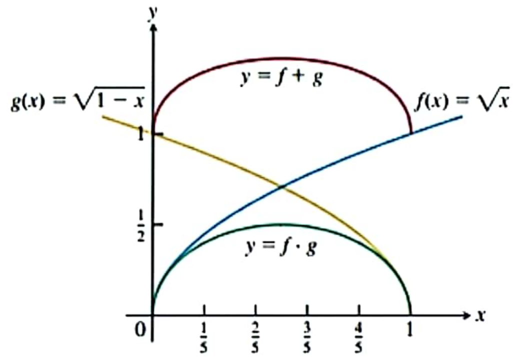


FIGURE 1.26 The domain of the function $f + g$ is the intersection of the domains of f and g, the interval $[0, 1]$ on the x-axis where these domains overlap. This interval is also the domain of the function $f \cdot g$ (Example 1).

Composite Functions

Composition is another method for combining functions.

DEFINITION If f and g are functions, the **composite** function $f \circ g$ ("f composed with g") is defined by

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

The domain of $f \circ g$ consists of the numbers x in the domain of g for which $g(x)$ lies in the domain of f.

The definition implies that $f \circ g$ can be formed when the range of g lies in the domain of f. To find $(f \circ g)(x)$, *first* find $g(x)$ and *second* find $f(g(x))$. Figure 1.27 pictures $f \circ g$ as a machine diagram, and Figure 1.28 shows the composite as an arrow diagram.

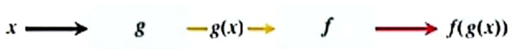


FIGURE 1.27 A composite function $f \circ g$ uses the output $g(x)$ of the first function g as the input for the second function f.

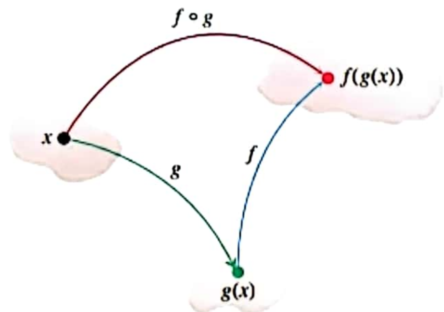


FIGURE 1.28 Arrow diagram for $f \circ g$. If x lies in the domain of g and $g(x)$ lies in the domain of f, then the functions f and g can be composed to form $(f \circ g)(x)$.

To evaluate the composite function $g \circ f$ (when defined), we find $f(x)$ first and then $g(f(x))$. The domain of $g \circ f$ is the set of numbers x in the domain of f such that $f(x)$ lies in the domain of g.

The functions $f \circ g$ and $g \circ f$ are usually quite different.

EXAMPLE 2 If $f(x) = \sqrt{x}$ and $g(x) = x + 1$, find

(a) $(f \circ g)(x)$ **(b)** $(g \circ f)(x)$ **(c)** $(f \circ f)(x)$ **(d)** $(g \circ g)(x)$.

Solution

	Composite	Domain
(a)	$(f \circ g)(x) = f(g(x)) = \sqrt{g(x)} = \sqrt{x+1}$	$[-1, \infty)$
(b)	$(g \circ f)(x) = g(f(x)) = f(x) + 1 = \sqrt{x} + 1$	$[0, \infty)$
(c)	$(f \circ f)(x) = f(f(x)) = \sqrt{f(x)} = \sqrt{\sqrt{x}} = x^{1/4}$	$[0, \infty)$
(d)	$(g \circ g)(x) = g(g(x)) = g(x) + 1 = (x + 1) + 1 = x + 2$	$(-\infty, \infty)$

To see why the domain of $f \circ g$ is $[-1, \infty)$, notice that $g(x) = x + 1$ is defined for all real x but belongs to the domain of f only if $x + 1 \geq 0$, that is to say, when $x \geq -1$. ■

Notice that if $f(x) = x^2$ and $g(x) = \sqrt{x}$, then $(f \circ g)(x) = \left(\sqrt{x}\right)^2 = x$. However, the domain of $f \circ g$ is $[0, \infty)$, not $(-\infty, \infty)$, since $\sqrt{x}$ requires $x \geq 0$.

Shifting a Graph of a Function

A common way to obtain a new function from an existing one is by adding a constant to each output of the existing function, or to its input variable. The graph of the new function is the graph of the original function shifted vertically or horizontally, as follows.

Shift Formulas

Vertical Shifts

$y = f(x) + k$ Shifts the graph of f *up* k units if $k > 0$
Shifts it *down* $|k|$ units if $k < 0$

Horizontal Shifts

$y = f(x + h)$ Shifts the graph of f *left* h units if $h > 0$
Shifts it *right* $|h|$ units if $h < 0$

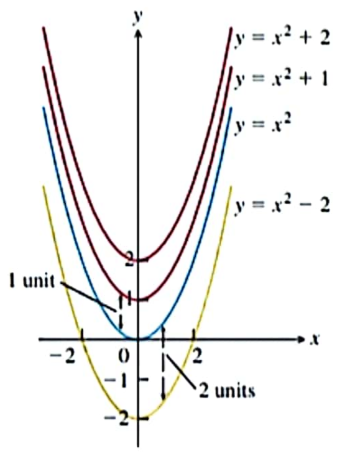


FIGURE 1.29 To shift the graph of $f(x) = x^2$ up (or down), we add positive (or negative) constants to the formula for f (Examples 3a and b).

EXAMPLE 3

(a) Adding 1 to the right-hand side of the formula $y = x^2$ to get $y = x^2 + 1$ shifts the graph up 1 unit (Figure 1.29).

(b) Adding -2 to the right-hand side of the formula $y = x^2$ to get $y = x^2 - 2$ shifts the graph down 2 units (Figure 1.29).

(c) Adding 3 to x in $y = x^2$ to get $y = (x + 3)^2$ shifts the graph 3 units to the left, while adding -2 shifts the graph 2 units to the right (Figure 1.30).

(d) Adding -2 to x in $y = |x|$, and then adding -1 to the result, gives $y = |x - 2| - 1$ and shifts the graph 2 units to the right and 1 unit down (Figure 1.31). ■

Scaling and Reflecting a Graph of a Function

To scale the graph of a function $y = f(x)$ is to stretch or compress it, vertically or horizontally. This is accomplished by multiplying the function f, or the independent variable x, by an appropriate constant c. Reflections across the coordinate axes are special cases where $c = -1$.

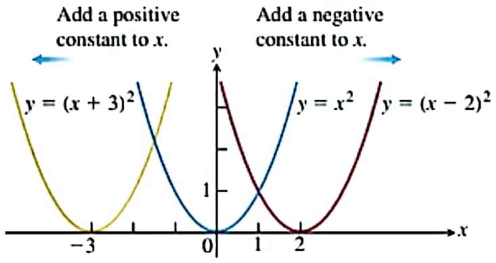


FIGURE 1.30 To shift the graph of $y = x^2$ to the left, we add a positive constant to x (Example 3c). To shift the graph to the right, we add a negative constant to x.

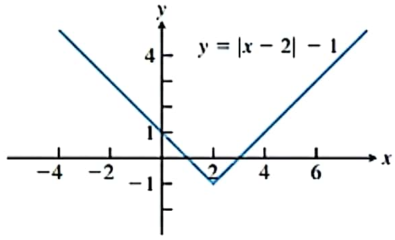


FIGURE 1.31 The graph of $y = |x|$ shifted 2 units to the right and 1 unit down (Example 3d).

Vertical and Horizontal Scaling and Reflecting Formulas

For $c > 1$, the graph is scaled:

$y = cf(x)$	Stretches the graph of f vertically by a factor of c.
$y = \frac{1}{c}f(x)$	Compresses the graph of f vertically by a factor of c.
$y = f(cx)$	Compresses the graph of f horizontally by a factor of c.
$y = f(x/c)$	Stretches the graph of f horizontally by a factor of c.

For $c = -1$, the graph is reflected:

$y = -f(x)$	Reflects the graph of f across the x-axis.
$y = f(-x)$	Reflects the graph of f across the y-axis.

EXAMPLE 4 Here we scale and reflect the graph of $y = \sqrt{x}$.

(a) **Vertical:** Multiplying the right-hand side of $y = \sqrt{x}$ by 3 to get $y = 3\sqrt{x}$ stretches the graph vertically by a factor of 3, whereas multiplying by $1/3$ compresses the graph by a factor of 3 (Figure 1.32).

(b) **Horizontal:** The graph of $y = \sqrt{3x}$ is a horizontal compression of the graph of $y = \sqrt{x}$ by a factor of 3, and $y = \sqrt{x/3}$ is a horizontal stretching by a factor of 3 (Figure 1.33). Note that $y = \sqrt{3x} = \sqrt{3}\sqrt{x}$ so a horizontal compression *may* correspond to a vertical stretching by a different scaling factor. Likewise, a horizontal stretching may correspond to a vertical compression by a different scaling factor.

(c) **Reflection:** The graph of $y = -\sqrt{x}$ is a reflection of $y = \sqrt{x}$ across the x-axis, and $y = \sqrt{-x}$ is a reflection across the y-axis (Figure 1.34). ■

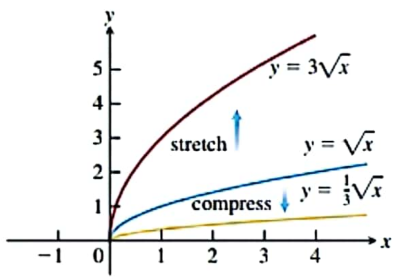


FIGURE 1.32 Vertically stretching and compressing the graph $y = \sqrt{x}$ by a factor of 3 (Example 4a).

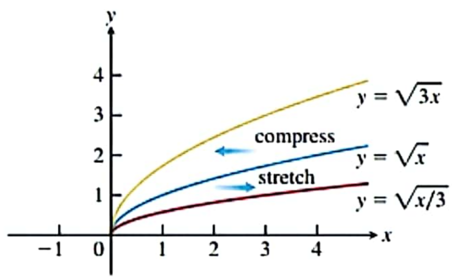


FIGURE 1.33 Horizontally stretching and compressing the graph $y = \sqrt{x}$ by a factor of 3 (Example 4b).

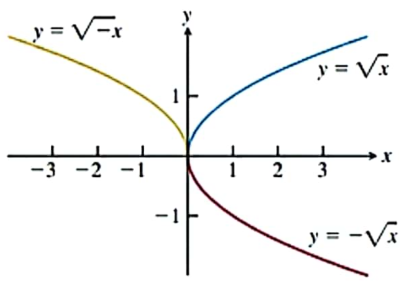


FIGURE 1.34 Reflections of the graph $y = \sqrt{x}$ across the coordinate axes (Example 4c).

Graphing

In Exercises 67–74, graph each function, not by plotting points, but by starting with the graph of one of the standard functions presented in Figures 1.14–1.17 and applying an appropriate transformation.

67. $y = -\sqrt{2x + 1}$

68. $y = \sqrt{1 - \frac{x}{2}}$

69. $y = (x - 1)^3 + 2$

70. $y = (1 - x)^3 + 2$

71. $y = \frac{1}{2x} - 1$

72. $y = \frac{2}{x^2} + 1$

73. $y = -\sqrt[3]{x}$

74. $y = (-2x)^{2/3}$

75. Graph the function $y = |x^2 - 1|$.

76. Graph the function $y = \sqrt{|x|}$.

Combining Functions

77. Assume that f is an even function, g is an odd function, and both f and g are defined on the entire real line $(-\infty, \infty)$. Which of the following (where defined) are even? odd?

a. fg **b.** f/g **c.** g/f
d. $f^2 = ff$ **e.** $g^2 = gg$ **f.** $f \circ g$
g. $g \circ f$ **h.** $f \circ f$ **i.** $g \circ g$

78. Can a function be both even and odd? Give reasons for your answer.

T **79.** (*Continuation of Example 1.*) Graph the functions $f(x) = \sqrt{x}$ and $g(x) = \sqrt{1 - x}$ together with their (a) sum, (b) product, (c) two differences, (d) two quotients.

T **80.** Let $f(x) = x - 7$ and $g(x) = x^2$. Graph f and g together with $f \circ g$ and $g \circ f$.

1.3 Trigonometric Functions

This section reviews radian measure and the basic trigonometric functions.

Angles

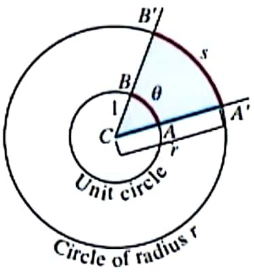


FIGURE 1.36 The radian measure of the central angle $A'CB'$ is the number $\theta = s/r$. For a unit circle of radius $r = 1$, θ is the length of arc AB that central angle ACB cuts from the unit circle.

Angles are measured in degrees or radians. The number of **radians** in the central angle $A'CB'$ within a circle of radius r is defined as the number of "radius units" contained in the arc s subtended by that central angle. If we denote this central angle by θ when measured in radians, this means that $\theta = s/r$ (Figure 1.36), or

$$s = r\theta \qquad (\theta \text{ in radians}). \tag{1}$$

If the circle is a unit circle having radius $r = 1$, then from Figure 1.36 and Equation (1), we see that the central angle θ measured in radians is just the length of the arc that the angle cuts from the unit circle. Since one complete revolution of the unit circle is 360° or 2π radians, we have

$$\pi \text{ radians} = 180° \tag{2}$$

and

$$1 \text{ radian} = \frac{180}{\pi} (\approx 57.3) \text{ degrees} \qquad \text{or} \qquad 1 \text{ degree} = \frac{\pi}{180} (\approx 0.017) \text{ radians}.$$

Table 1.1 shows the equivalence between degree and radian measures for some basic angles.

TABLE 1.1 Angles measured in degrees and radians

Degrees	−180	−135	−90	−45	0	30	45	60	90	120	135	150	180	270	360
θ (radians)	$-\pi$	$\frac{-3\pi}{4}$	$\frac{-\pi}{2}$	$\frac{-\pi}{4}$	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{6}$	π	$\frac{3\pi}{2}$	2π

An angle in the xy-plane is said to be in **standard position** if its vertex lies at the origin and its initial ray lies along the positive x-axis (Figure 1.37). Angles measured counterclockwise from the positive x-axis are assigned positive measures; angles measured clockwise are assigned negative measures.

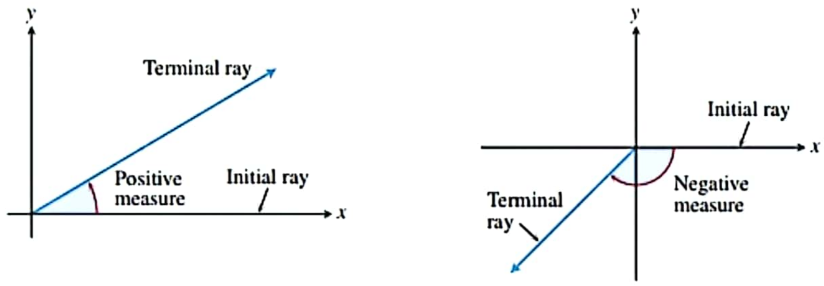


FIGURE 1.37 Angles in standard position in the xy-plane.

Angles describing counterclockwise rotations can go arbitrarily far beyond 2π radians or 360°. Similarly, angles describing clockwise rotations can have negative measures of all sizes (Figure 1.38).

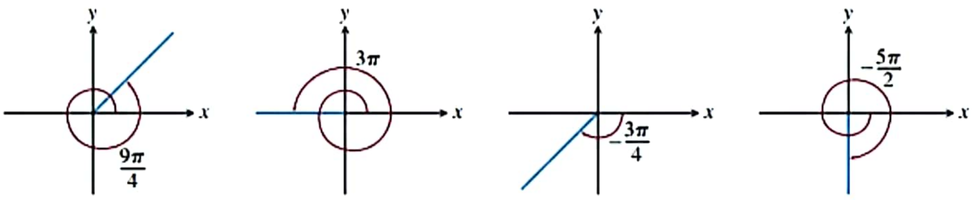
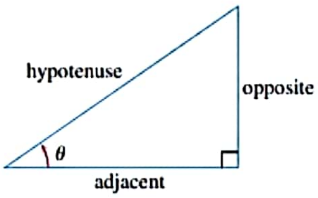


FIGURE 1.38 Nonzero radian measures can be positive or negative and can go beyond 2π.

Angle Convention: Use Radians From now on, in this book it is assumed that all angles are measured in radians unless degrees or some other unit is stated explicitly. When we talk about the angle $\pi/3$, we mean $\pi/3$ radians (which is 60°), not $\pi/3$ degrees. We use radians because it simplifies many of the operations in calculus, and some results we will obtain involving the trigonometric functions are not true when angles are measured in degrees.



$$\sin\theta = \frac{\text{opp}}{\text{hyp}} \qquad \csc\theta = \frac{\text{hyp}}{\text{opp}}$$
$$\cos\theta = \frac{\text{adj}}{\text{hyp}} \qquad \sec\theta = \frac{\text{hyp}}{\text{adj}}$$
$$\tan\theta = \frac{\text{opp}}{\text{adj}} \qquad \cot\theta = \frac{\text{adj}}{\text{opp}}$$

FIGURE 1.39 Trigonometric ratios of an acute angle.

The Six Basic Trigonometric Functions

You are probably familiar with defining the trigonometric functions of an acute angle in terms of the sides of a right triangle (Figure 1.39). We extend this definition to obtuse and negative angles by first placing the angle in standard position in a circle of radius r. We then define the trigonometric functions in terms of the coordinates of the point $P(x, y)$ where the angle's terminal ray intersects the circle (Figure 1.40).

$$\textbf{sine:} \quad \sin\theta = \frac{y}{r} \qquad \textbf{cosecant:} \quad \csc\theta = \frac{r}{y}$$
$$\textbf{cosine:} \quad \cos\theta = \frac{x}{r} \qquad \textbf{secant:} \quad \sec\theta = \frac{r}{x}$$
$$\textbf{tangent:} \quad \tan\theta = \frac{y}{x} \qquad \textbf{cotangent:} \quad \cot\theta = \frac{x}{y}$$

These extended definitions agree with the right-triangle definitions when the angle is acute.

Notice also that whenever the quotients are defined,

$$\tan\theta = \frac{\sin\theta}{\cos\theta} \qquad \cot\theta = \frac{1}{\tan\theta}$$
$$\sec\theta = \frac{1}{\cos\theta} \qquad \csc\theta = \frac{1}{\sin\theta}$$

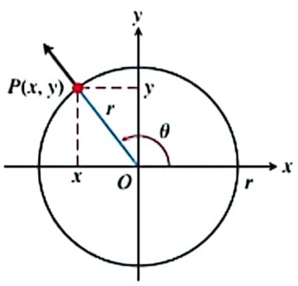


FIGURE 1.40 The trigonometric functions of a general angle θ are defined in terms of x, y, and r.

As you can see, $\tan\theta$ and $\sec\theta$ are not defined if $x = \cos\theta = 0$. This means they are not defined if θ is $\pm\pi/2, \pm 3\pi/2, \ldots$. Similarly, $\cot\theta$ and $\csc\theta$ are not defined for values of θ for which $y = 0$, namely $\theta = 0, \pm\pi, \pm 2\pi, \ldots$.

The exact values of these trigonometric ratios for some angles can be read from the triangles in Figure 1.41. For instance,

$$\sin\frac{\pi}{4} = \frac{1}{\sqrt{2}} \qquad \sin\frac{\pi}{6} = \frac{1}{2} \qquad \sin\frac{\pi}{3} = \frac{\sqrt{3}}{2}$$

$$\cos\frac{\pi}{4} = \frac{1}{\sqrt{2}} \qquad \cos\frac{\pi}{6} = \frac{\sqrt{3}}{2} \qquad \cos\frac{\pi}{3} = \frac{1}{2}$$

$$\tan\frac{\pi}{4} = 1 \qquad \tan\frac{\pi}{6} = \frac{1}{\sqrt{3}} \qquad \tan\frac{\pi}{3} = \sqrt{3}$$

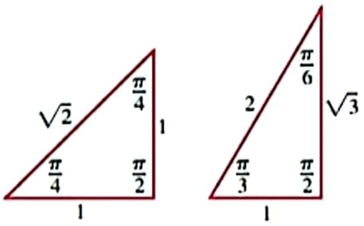


FIGURE 1.41 Radian angles and side lengths of two common triangles.

The CAST rule (Figure 1.42) is useful for remembering when the basic trigonometric functions are positive or negative. For instance, from the triangle in Figure 1.43, we see that

$$\sin\frac{2\pi}{3} = \frac{\sqrt{3}}{2}, \qquad \cos\frac{2\pi}{3} = -\frac{1}{2}, \qquad \tan\frac{2\pi}{3} = -\sqrt{3}.$$

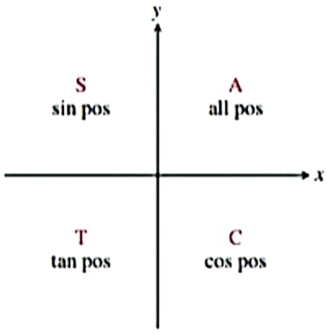


FIGURE 1.42 The CAST rule, remembered by the statement "Calculus Activates Student Thinking," tells which trigonometric functions are positive in each quadrant.

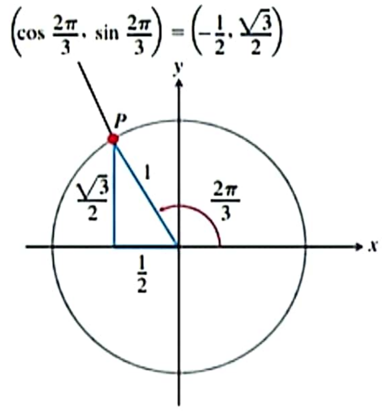


FIGURE 1.43 The triangle for calculating the sine and cosine of $2\pi/3$ radians. The side lengths come from the geometry of right triangles.

Using a similar method we determined the values of $\sin\theta$, $\cos\theta$, and $\tan\theta$ shown in Table 1.2.

TABLE 1.2 Values of $\sin\theta$, $\cos\theta$, and $\tan\theta$ for selected values of θ

Degrees	−180	−135	−90	−45	0	30	45	60	90	120	135	150	180	270	360
θ (radians)	$-\pi$	$\frac{-3\pi}{4}$	$\frac{-\pi}{2}$	$\frac{-\pi}{4}$	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{6}$	π	$\frac{3\pi}{2}$	2π
$\sin\theta$	0	$\frac{-\sqrt{2}}{2}$	−1	$\frac{-\sqrt{2}}{2}$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0	−1	0
$\cos\theta$	−1	$\frac{-\sqrt{2}}{2}$	0	$\frac{\sqrt{2}}{2}$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$\frac{-\sqrt{2}}{2}$	$\frac{-\sqrt{3}}{2}$	−1	0	1
$\tan\theta$	0	1		−1	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$		$-\sqrt{3}$	−1	$\frac{-\sqrt{3}}{3}$	0		0

Periodicity and Graphs of the Trigonometric Functions

When an angle of measure θ and an angle of measure $\theta + 2\pi$ are in standard position, their terminal rays coincide. The two angles therefore have the same trigonometric function values: $\sin(\theta + 2\pi) = \sin\theta$, $\tan(\theta + 2\pi) = \tan\theta$, and so on. Similarly, $\cos(\theta - 2\pi) = \cos\theta$, $\sin(\theta - 2\pi) = \sin\theta$, and so on. We describe this repeating behavior by saying that the six basic trigonometric functions are *periodic*.

Periods of Trigonometric Functions

Period π: $\tan(x + \pi) = \tan x$
$\cot(x + \pi) = \cot x$

Period 2π: $\sin(x + 2\pi) = \sin x$
$\cos(x + 2\pi) = \cos x$
$\sec(x + 2\pi) = \sec x$
$\csc(x + 2\pi) = \csc x$

DEFINITION A function $f(x)$ is **periodic** if there is a positive number p such that $f(x + p) = f(x)$ for every value of x. The smallest such value of p is the **period** of f.

When we graph trigonometric functions in the coordinate plane, we usually denote the independent variable by x instead of θ. Figure 1.44 shows that the tangent and cotangent functions have period $p = \pi$, and the other four functions have period 2π. Also, the symmetries in these graphs reveal that the cosine and secant functions are even and the other four functions are odd (although this does not prove those results).

Even

$\cos(-x) = \cos x$
$\sec(-x) = \sec x$

Odd

$\sin(-x) = -\sin x$
$\tan(-x) = -\tan x$
$\csc(-x) = -\csc x$
$\cot(-x) = -\cot x$

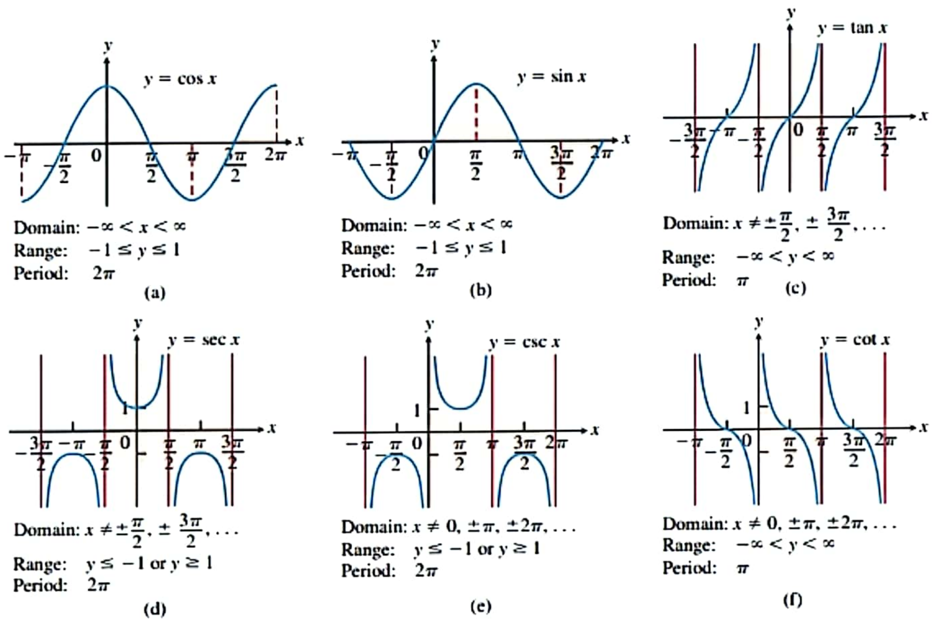


FIGURE 1.44 Graphs of the six basic trigonometric functions using radian measure. The shading for each trigonometric function indicates its periodicity.

Trigonometric Identities

The coordinates of any point $P(x, y)$ in the plane can be expressed in terms of the point's distance r from the origin and the angle θ that ray OP makes with the positive x-axis (Figure 1.40). Since $x/r = \cos\theta$ and $y/r = \sin\theta$, we have

$$x = r\cos\theta, \quad y = r\sin\theta.$$

When $r = 1$ we can apply the Pythagorean theorem to the reference right triangle in Figure 1.45 and obtain the equation

$$\cos^2\theta + \sin^2\theta = 1. \tag{3}$$

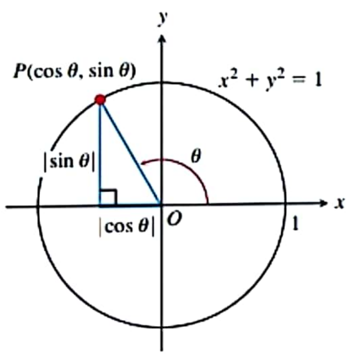


FIGURE 1.45 The reference triangle for a general angle θ.

This equation, true for all values of θ, is the most frequently used identity in trigonometry. Dividing this identity in turn by $\cos^2\theta$ and $\sin^2\theta$ gives

$$1 + \tan^2\theta = \sec^2\theta$$
$$1 + \cot^2\theta = \csc^2\theta$$

The following formulas hold for all angles A and B (Exercise 58).

Addition Formulas

$$\begin{aligned} \cos(A + B) &= \cos A\cos B - \sin A\sin B \\ \sin(A + B) &= \sin A\cos B + \cos A\sin B \end{aligned} \tag{4}$$

There are similar formulas for $\cos(A - B)$ and $\sin(A - B)$ (Exercises 35 and 36). All the trigonometric identities needed in this book derive from Equations (3) and (4). For example, substituting θ for both A and B in the addition formulas gives

Double-Angle Formulas

$$\begin{aligned} \cos 2\theta &= \cos^2\theta - \sin^2\theta \\ \sin 2\theta &= 2\sin\theta\cos\theta \end{aligned} \tag{5}$$

Additional formulas come from combining the equations

$$\cos^2\theta + \sin^2\theta = 1, \qquad \cos^2\theta - \sin^2\theta = \cos 2\theta.$$

We add the two equations to get $2\cos^2\theta = 1 + \cos 2\theta$ and subtract the second from the first to get $2\sin^2\theta = 1 - \cos 2\theta$. This results in the following identities, which are useful in integral calculus.

Half-Angle Formulas

$$\cos^2\theta = \frac{1 + \cos 2\theta}{2} \tag{6}$$

$$\sin^2\theta = \frac{1 - \cos 2\theta}{2} \tag{7}$$

The Law of Cosines

If a, b, and c are sides of a triangle ABC and if θ is the angle opposite c, then

$$c^2 = a^2 + b^2 - 2ab\cos\theta. \tag{8}$$

This equation is called the **law of cosines**.

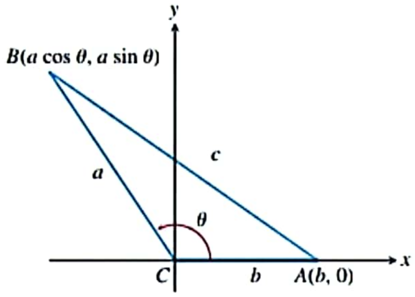


FIGURE 1.46 The square of the distance between A and B gives the law of cosines.

We can see why the law holds if we introduce coordinate axes with the origin at C and the positive x-axis along one side of the triangle, as in Figure 1.46. The coordinates of A are $(b, 0)$; the coordinates of B are $(a\cos\theta, a\sin\theta)$. The square of the distance between A and B is therefore

$$\begin{aligned} c^2 &= (a\cos\theta - b)^2 + (a\sin\theta)^2 \\ &= a^2\underbrace{(\cos^2\theta + \sin^2\theta)}_{1} + b^2 - 2ab\cos\theta \\ &= a^2 + b^2 - 2ab\cos\theta. \end{aligned}$$

The law of cosines generalizes the Pythagorean theorem. If $\theta = \pi/2$, then $\cos\theta = 0$ and $c^2 = a^2 + b^2$.

Two Special Inequalities

For any angle θ measured in radians, the sine and cosine functions satisfy

$$\boxed{-|\theta| \le \sin\theta \le |\theta| \quad \text{and} \quad -|\theta| \le 1 - \cos\theta \le |\theta|.}$$

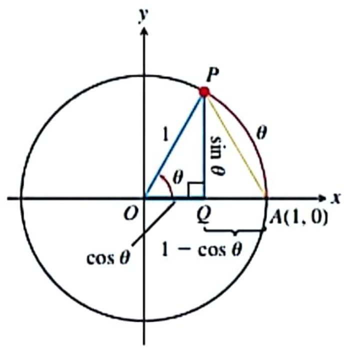


FIGURE 1.47 From the geometry of this figure, drawn for $\theta > 0$, we get the inequality $\sin^2\theta + (1 - \cos\theta)^2 \le \theta^2$.

To establish these inequalities, we picture θ as a nonzero angle in standard position (Figure 1.47). The circle in the figure is a unit circle, so $|\theta|$ equals the length of the circular arc AP. The length of line segment AP is therefore less than $|\theta|$.

Triangle APQ is a right triangle with sides of length

$$QP = |\sin\theta|, \qquad AQ = 1 - \cos\theta.$$

From the Pythagorean theorem and the fact that $AP < |\theta|$, we get

$$\sin^2\theta + (1 - \cos\theta)^2 = (AP)^2 \le \theta^2. \tag{9}$$

The terms on the left-hand side of Equation (9) are both positive, so each is smaller than their sum and hence is less than or equal to θ^2:

$$\sin^2\theta \le \theta^2 \quad \text{and} \quad (1 - \cos\theta)^2 \le \theta^2.$$

By taking square roots, this is equivalent to saying that

$$|\sin\theta| \le |\theta| \quad \text{and} \quad |1 - \cos\theta| \le |\theta|,$$

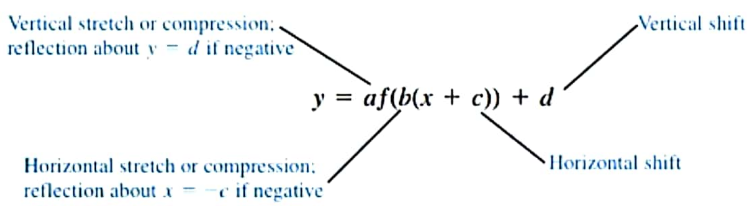
so

$$-|\theta| \le \sin\theta \le |\theta| \quad \text{and} \quad -|\theta| \le 1 - \cos\theta \le |\theta|.$$

These inequalities will be useful in the next chapter.

Transformations of Trigonometric Graphs

The rules for shifting, stretching, compressing, and reflecting the graph of a function summarized in the following diagram apply to the trigonometric functions we have discussed in this section.



63. A triangle has side $c = 2$ and angles $A = \pi/4$ and $B = \pi/3$. Find the length a of the side opposite A.

T **64. The approximation $\sin x \approx x$** It is often useful to know that, when x is measured in radians, $\sin x \approx x$ for numerically small values of x. In Section 3.11, we will see why the approximation holds. The approximation error is less than 1 in 5000 if $|x| < 0.1$.

a. With your grapher in radian mode, graph $y = \sin x$ and $y = x$ together in a viewing window about the origin. What do you see happening as x nears the origin?

b. With your grapher in degree mode, graph $y = \sin x$ and $y = x$ together about the origin again. How is the picture different from the one obtained with radian mode?

General Sine Curves

For

$$f(x) = A\sin\left(\frac{2\pi}{B}(x - C)\right) + D,$$

identify A, B, C, and D for the sine functions in Exercises 65–68 and sketch their graphs.

65. $y = 2\sin(x + \pi) - 1$

66. $y = \frac{1}{2}\sin(\pi x - \pi) + \frac{1}{2}$

67. $y = -\frac{2}{\pi}\sin\left(\frac{\pi}{2}t\right) + \frac{1}{\pi}$

68. $y = \frac{L}{2\pi}\sin\frac{2\pi t}{L},\quad L > 0$

COMPUTER EXPLORATIONS

In Exercises 69–72, you will explore graphically the general sine function

$$f(x) = A\sin\left(\frac{2\pi}{B}(x - C)\right) + D$$

as you change the values of the constants A, B, C, and D. Use a CAS or computer grapher to perform the steps in the exercises.

69. The period B Set the constants $A = 3$, $C = D = 0$.

a. Plot $f(x)$ for the values $B = 1, 3, 2\pi, 5\pi$ over the interval $-4\pi \le x \le 4\pi$. Describe what happens to the graph of the general sine function as the period increases.

b. What happens to the graph for negative values of B? Try it with $B = -3$ and $B = -2\pi$.

70. The horizontal shift C Set the constants $A = 3$, $B = 6$, $D = 0$.

a. Plot $f(x)$ for the values $C = 0$, 1, and 2 over the interval $-4\pi \le x \le 4\pi$. Describe what happens to the graph of the general sine function as C increases through positive values.

b. What happens to the graph for negative values of C?

c. What smallest positive value should be assigned to C so the graph exhibits no horizontal shift? Confirm your answer with a plot.

71. The vertical shift D Set the constants $A = 3$, $B = 6$, $C = 0$.

a. Plot $f(x)$ for the values $D = 0$, 1, and 3 over the interval $-4\pi \le x \le 4\pi$. Describe what happens to the graph of the general sine function as D increases through positive values.

b. What happens to the graph for negative values of D?

72. The amplitude A Set the constants $B = 6$, $C = D = 0$.

a. Describe what happens to the graph of the general sine function as A increases through positive values. Confirm your answer by plotting $f(x)$ for the values $A = 1$, 5, and 9.

b. What happens to the graph for negative values of A?

1.4 Graphing with Software

Today a number of hardware devices, including computers, calculators, and smartphones, have graphing applications based on software that enables us to graph very complicated functions with high precision. Many of these functions could not otherwise be easily graphed. However, some care must be taken when using such graphing software, and in this section we address some of the issues that may be involved. In Chapter 4 we will see how calculus helps us determine that we are accurately viewing all the important features of a function's graph.

Graphing Windows

When using software for graphing, a portion of the graph is displayed in a **display** or **viewing window**. Depending on the software, the default window may give an incomplete or misleading picture of the graph. We use the term *square window* when the units or scales used on both axes are the same. This term does not mean that the display window itself is square (usually it is rectangular), but instead it means that the x-unit is the same length as the y-unit.

When a graph is displayed in the default mode, the x-unit may differ from the y-unit of scaling in order to capture essential features of the graph. This difference in scaling can cause visual distortions that may lead to erroneous interpretations of the function's behavior.

Some graphing software allows us to set the viewing window by specifying one or both of the intervals, $a \le x \le b$ and $c \le y \le d$, and it may allow for equalizing the scales used for the axes as well. The software selects equally spaced x-values in $[a, b]$ and then plots the points $(x, f(x))$. A point is plotted if and only if x lies in the domain of the function and $f(x)$ lies within the interval $[c, d]$. A short line segment is then drawn between each plotted point and its next neighboring point. We now give illustrative examples of some common problems that may occur with this procedure.

EXAMPLE 1 Graph the function $f(x) = x^3 - 7x^2 + 28$ in each of the following display or viewing windows:

(a) $[-10, 10]$ by $[-10, 10]$ **(b)** $[-4, 4]$ by $[-50, 10]$ **(c)** $[-4, 10]$ by $[-60, 60]$

Solution

(a) We select $a = -10$, $b = 10$, $c = -10$, and $d = 10$ to specify the interval of x-values and the range of y-values for the window. The resulting graph is shown in Figure 1.48a. It appears that the window is cutting off the bottom part of the graph and that the interval of x-values is too large. Let's try the next window.

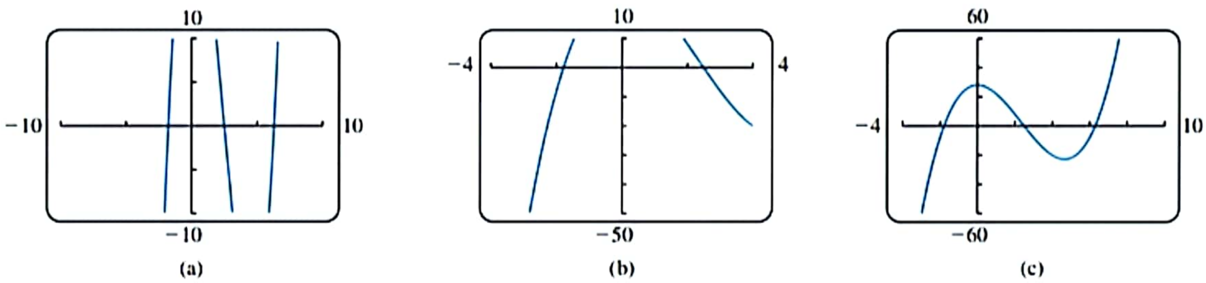


FIGURE 1.48 The graph of $f(x) = x^3 - 7x^2 + 28$ in different viewing windows. Selecting a window that gives a clear picture of a graph is often a trial-and-error process (Example 1). The default window used by the software may automatically display the graph in (c).

(b) We see some new features of the graph (Figure 1.48b), but the top is missing and we need to view more to the right of $x = 4$ as well. The next window should help.

(c) Figure 1.48c shows the graph in this new viewing window. Observe that we get a more complete picture of the graph in this window, and it is a reasonable graph of a third-degree polynomial. ■

EXAMPLE 2 When a graph is displayed, the x-unit may differ from the y-unit, as in the graphs shown in Figures 1.48b and 1.48c. The result is distortion in the picture, which may be misleading. The display window can be made square by compressing or stretching the units on one axis to match the scale on the other, giving the true graph. Many software systems have built-in options to make the window "square." If yours does not, you may have to bring to your viewing some foreknowledge of the true picture.

Figure 1.49a shows the graphs of the perpendicular lines $y = x$ and $y = -x + 3\sqrt{2}$, together with the semicircle $y = \sqrt{9 - x^2}$, in a nonsquare $[-4, 4]$ by $[-6, 8]$ display window. Notice the distortion. The lines do not appear to be perpendicular, and the semicircle appears to be elliptical in shape.

Figure 1.49b shows the graphs of the same functions in a square window in which the x-units are scaled to be the same as the y-units. Notice that the scaling on the x-axis for Figure 1.49a has been compressed in Figure 1.49b to make the window square. Figure 1.49c gives an enlarged view of Figure 1.49b with a square $[-3, 3]$ by $[0, 4]$ window. ■

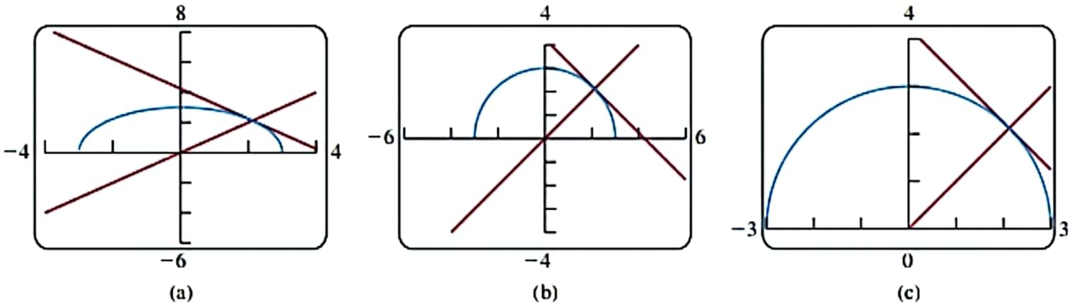


FIGURE 1.49 Graphs of the perpendicular lines $y = x$ and $y = -x + 3\sqrt{2}$ and of the semicircle $y = \sqrt{9 - x^2}$ appear distorted (a) in a nonsquare window, but clear (b) and (c) in square windows (Example 2). Some software may not provide options for the views in (b) or (c).

If the denominator of a rational function is zero at some x-value within the viewing window, graphing software may produce a steep near-vertical line segment from the top to the bottom of the window. Example 3 illustrates steep line segments.

Sometimes the graph of a trigonometric function oscillates very rapidly. When graphing software plots the points of the graph and connects them, many of the maximum and minimum points are actually missed. The resulting graph is then very misleading.

EXAMPLE 3 Graph the function $f(x) = \sin 100x$.

Solution Figure 1.50a shows the graph of f in the viewing window $[-12, 12]$ by $[-1, 1]$. We see that the graph looks very strange because the sine curve should oscillate periodically between -1 and 1. This behavior is not exhibited in Figure 1.50a. We might experiment with a smaller viewing window, say $[-6, 6]$ by $[-1, 1]$, but the graph is not better (Figure 1.50b). The difficulty is that the period of the trigonometric function $y = \sin 100x$ is very small ($2\pi/100 \approx 0.063$). If we choose the much smaller viewing window $[-0.1, 0.1]$ by $[-1, 1]$ we get the graph shown in Figure 1.50c. This graph reveals the expected oscillations of a sine curve.

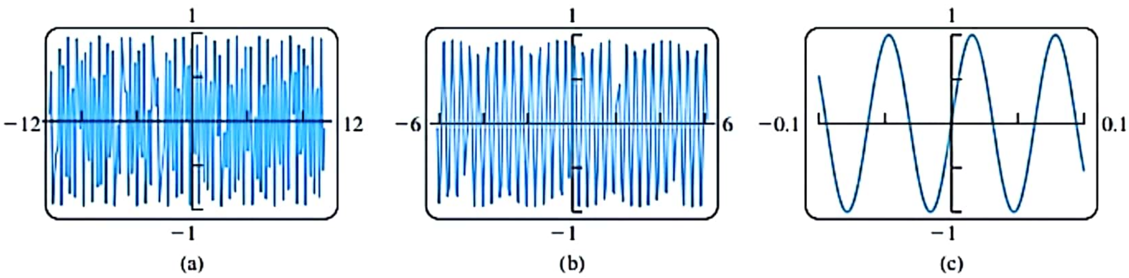


FIGURE 1.50 Graphs of the function $y = \sin 100x$ in three viewing windows. Because the period is $2\pi/100 \approx 0.063$, the smaller window in (c) best displays the true aspects of this rapidly oscillating function (Example 3).

EXAMPLE 4 Graph the function $y = \cos x + \frac{1}{200}\sin 200x$.

Solution In the viewing window $[-6, 6]$ by $[-1, 1]$ the graph appears much like the cosine function with some very small sharp wiggles on it (Figure 1.51a). We get a better look when we significantly reduce the window to $[-0.2, 0.2]$ by $[0.97, 1.01]$, obtaining the graph in Figure 1.51b. We now see the small but rapid oscillations of the second term, $(1/200)\sin 200x$, added to the comparatively larger values of the cosine curve.

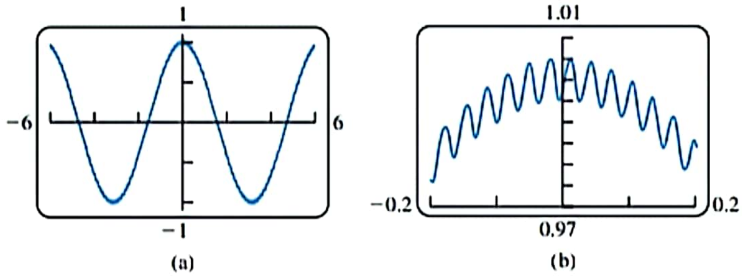


FIGURE 1.51 In (b) we see a close-up view of the function $y = \cos x + \frac{1}{200}\sin 200x$ graphed in (a). The term $\cos x$ clearly dominates the second term, $\frac{1}{200}\sin 200x$, which produces the rapid oscillations along the cosine curve. Both views are needed for a clear idea of the graph (Example 4).

Obtaining a Complete Graph

Some graphing software will not display the portion of a graph for $f(x)$ when $x < 0$. Usually that happens because of the algorithm the software is using to calculate the function values. Sometimes we can obtain the complete graph by defining the formula for the function in a different way, as illustrated in the next example.

EXAMPLE 5 Graph the function $y = x^{1/3}$.

Solution Some graphing software displays the graph shown in Figure 1.52a. When we compare it with the graph of $y = x^{1/3} = \sqrt[3]{x}$ in Figure 1.17, we see that the left branch for $x < 0$ is missing. The reason the graphs differ is that the software algorithm calculates $x^{1/3}$ as $e^{(1/3)\ln x}$. Since the logarithmic function is not defined for negative values of x, the software can produce only the right branch, where $x > 0$. (Logarithmic and exponential functions are introduced in the next two sections.)

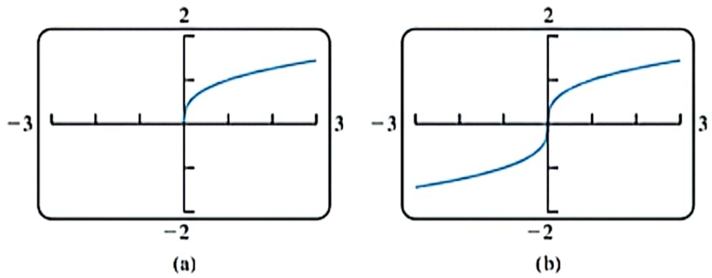


FIGURE 1.52 The graph of $y = x^{1/3}$ is missing the left branch in (a). In (b) we graph the function $f(x) = \frac{x}{|x|} \cdot |x|^{1/3}$, obtaining both branches. (See Example 5.)

To obtain the full picture showing both branches, we can graph the function

$$f(x) = \frac{x}{|x|} \cdot |x|^{1/3}.$$

This function equals $x^{1/3}$ except at $x = 0$ (where f is undefined, although $0^{1/3} = 0$). A graph of f is displayed in Figure 1.52b. ■

Capturing the Trend of Collected Data

We have pointed out that applied scientists and analysts often collect data to study a particular issue or phenomenon of interest. If there is no known principle or physical law

2

Limits and Continuity

OVERVIEW Mathematicians of the seventeenth century were keenly interested in the study of motion for objects on or near the earth and the motion of planets and stars. This study involved both the speed of the object and its direction of motion at any instant, and they knew the direction at a given instant was along a line tangent to the path of motion. The concept of a limit is fundamental to finding the velocity of a moving object and the tangent to a curve. In this chapter we develop the limit, first intuitively and then formally. We use limits to describe the way a function varies. Some functions vary *continuously*; small changes in x produce only small changes in $f(x)$. Other functions can have values that jump, vary erratically, or tend to increase or decrease without bound. The notion of limit gives a precise way to distinguish between these behaviors.

2.1 Rates of Change and Tangents to Curves

Calculus is a tool that helps us understand how a change in one quantity is related to a change in another. How does the speed of a falling object change as a function of time? How does the level of water in a barrel change as a function of the amount of liquid poured into it? We see change occurring in nearly everything we observe in the world and universe, and powerful modern instruments help us see more and more. In this section we introduce the ideas of average and instantaneous rates of change, and show that they are closely related to the slope of a curve at a point P on the curve. We give precise developments of these important concepts in the next chapter, but for now we use an informal approach so you will see how they lead naturally to the main idea of this chapter, the *limit*. The idea of a limit plays a foundational role throughout calculus.

Average and Instantaneous Speed

HISTORICAL BIOGRAPHY*
Galileo Galilei
(1564–1642)

In the late sixteenth century, Galileo discovered that a solid object dropped from rest (not moving) near the surface of the earth and allowed to fall freely will fall a distance proportional to the square of the time it has been falling. This type of motion is called **free fall**. It assumes negligible air resistance to slow the object down, and that gravity is the only force acting on the falling object. If y denotes the distance fallen in feet after t seconds, then Galileo's law is

$$y = 16t^2,$$

where 16 is the (approximate) constant of proportionality. (If y is measured in meters, the constant is 4.9.)

A moving object's **average speed** during an interval of time is found by dividing the distance covered by the time elapsed. The unit of measure is length per unit time: kilometers per hour, feet (or meters) per second, or whatever is appropriate to the problem at hand.

*To learn more about the historical figures mentioned in the text and the development of many major elements and topics of calculus, visit **www.aw.com/thomas**.

EXAMPLE 1 A rock breaks loose from the top of a tall cliff. What is its average speed

(a) during the first 2 sec of fall?

(b) during the 1-sec interval between second 1 and second 2?

Solution The average speed of the rock during a given time interval is the change in distance, Δy, divided by the length of the time interval, Δt. (Increments like Δy and Δt are reviewed in Appendix 3, and pronounced "delta y" and "delta t.") Measuring distance in feet and time in seconds, we have the following calculations:

(a) For the first 2 sec: $\dfrac{\Delta y}{\Delta t} = \dfrac{16(2)^2 - 16(0)^2}{2 - 0} = 32\dfrac{\text{ft}}{\text{sec}}$

(b) From sec 1 to sec 2: $\dfrac{\Delta y}{\Delta t} = \dfrac{16(2)^2 - 16(1)^2}{2 - 1} = 48\dfrac{\text{ft}}{\text{sec}}$ ■

We want a way to determine the speed of a falling object at a single instant t_0, instead of using its average speed over an interval of time. To do this, we examine what happens when we calculate the average speed over shorter and shorter time intervals starting at t_0. The next example illustrates this process. Our discussion is informal here, but it will be made precise in Chapter 3.

EXAMPLE 2 Find the speed of the falling rock in Example 1 at $t = 1$ and $t = 2$ sec.

Solution We can calculate the average speed of the rock over a time interval $[t_0, t_0 + h]$, having length $\Delta t = h$, as

$$\frac{\Delta y}{\Delta t} = \frac{16(t_0 + h)^2 - 16t_0^2}{h}. \tag{1}$$

We cannot use this formula to calculate the "instantaneous" speed at the exact moment t_0 by simply substituting $h = 0$, because we cannot divide by zero. But we *can* use it to calculate average speeds over increasingly short time intervals starting at $t_0 = 1$ and $t_0 = 2$. When we do so, by taking smaller and smaller values of h, we see a pattern (Table 2.1).

TABLE 2.1 Average speeds over short time intervals $[t_0, t_0 + h]$

Average speed: $\dfrac{\Delta y}{\Delta t} = \dfrac{16(t_0 + h)^2 - 16t_0^2}{h}$

Length of time interval h	Average speed over interval of length h starting at $t_0 = 1$	Average speed over interval of length h starting at $t_0 = 2$
1	48	80
0.1	33.6	65.6
0.01	32.16	64.16
0.001	32.016	64.016
0.0001	32.0016	64.0016

The average speed on intervals starting at $t_0 = 1$ seems to approach a limiting value of 32 as the length of the interval decreases. This suggests that the rock is falling at a speed of 32 ft/sec at $t_0 = 1$ sec. Let's confirm this algebraically.

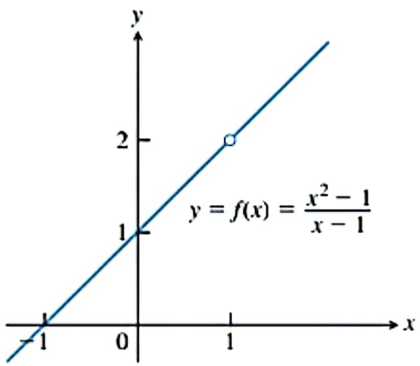
2.2 Limit of a Function and Limit Laws

In Section 2.1 we saw that limits arise when finding the instantaneous rate of change of a function or the tangent to a curve. Here we begin with an informal definition of *limit* and show how we can calculate the values of limits. A precise definition is presented in the next section.

HISTORICAL ESSAY
Limits

Limits of Function Values

Frequently when studying a function $y = f(x)$, we find ourselves interested in the function's behavior *near* a particular point c, but not *at* c. This might be the case, for instance, if c is an irrational number, like π or $\sqrt{2}$, whose values can only be approximated by "close" rational numbers at which we actually evaluate the function instead. Another situation occurs when trying to evaluate a function at c leads to division by zero, which is undefined. We encountered this last circumstance when seeking the instantaneous rate of change in y by considering the quotient function $\Delta y/h$ for h closer and closer to zero. Here's a specific example in which we explore numerically how a function behaves near a particular point at which we cannot directly evaluate the function.



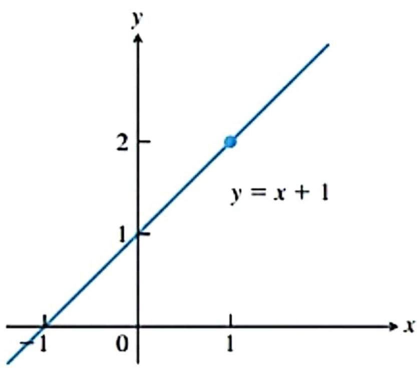


FIGURE 2.7 The graph of f is identical with the line $y = x + 1$ except at $x = 1$, where f is not defined (Example 1).

EXAMPLE 1 How does the function

$$f(x) = \frac{x^2 - 1}{x - 1}$$

behave near $x = 1$?

Solution The given formula defines f for all real numbers x except $x = 1$ (we cannot divide by zero). For any $x \neq 1$, we can simplify the formula by factoring the numerator and canceling common factors:

$$f(x) = \frac{(x - 1)(x + 1)}{x - 1} = x + 1 \qquad \text{for} \qquad x \neq 1.$$

The graph of f is the line $y = x + 1$ with the point $(1, 2)$ *removed*. This removed point is shown as a "hole" in Figure 2.7. Even though $f(1)$ is not defined, it is clear that we can make the value of $f(x)$ *as close as we want* to 2 by choosing x *close enough* to 1 (Table 2.2). ■

TABLE 2.2 As x gets closer to 1, $f(x)$ gets closer to 2.

x	$f(x) = \dfrac{x^2-1}{x-1}$
0.9	1.9
1.1	2.1
0.99	1.99
1.01	2.01
0.999	1.999
1.001	2.001
0.999999	1.999999
1.000001	2.000001

Generalizing the idea illustrated in Example 1, suppose $f(x)$ is defined on an open interval about c, *except possibly at c itself.* If $f(x)$ is arbitrarily close to the number L (as close to L as we like) for all x sufficiently close to c, we say that f approaches the **limit** L as x approaches c, and write

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

which is read "the limit of $f(x)$ as x approaches c is L." For instance, in Example 1 we would say that $f(x)$ approaches the *limit* 2 as x approaches 1, and write

$$\lim_{x \to 1} f(x) = 2, \qquad \text{or} \qquad \lim_{x \to 1} \frac{x^2 - 1}{x - 1} = 2.$$

Essentially, the definition says that the values of $f(x)$ are close to the number L whenever x is close to c (on either side of c).

Our definition here is "informal" because phrases like *arbitrarily close* and *sufficiently close* are imprecise; their meaning depends on the context. (To a machinist manufacturing a piston, *close* may mean *within a few thousandths of an inch*. To an astronomer studying distant galaxies, *close* may mean *within a few thousand light-years*.) Nevertheless, the definition is clear enough to enable us to recognize and evaluate limits of many specific functions. We will need the precise definition given in Section 2.3, however, when we set out to prove theorems about limits or study complicated functions. Here are several more examples exploring the idea of limits.

EXAMPLE 2 The limit value of a function does not depend on how the function is defined at the point being approached. Consider the three functions in Figure 2.8. The function f has limit 2 as $x \to 1$ even though f is not defined at $x = 1$. The function g has limit 2 as $x \to 1$ even though $2 \neq g(1)$. The function h is the only one of the three functions in Figure 2.8 whose limit as $x \to 1$ equals its value at $x = 1$. For h, we have $\lim_{x \to 1} h(x) = h(1)$. This equality of limit and function value is of special importance, and we return to it in Section 2.5. ■

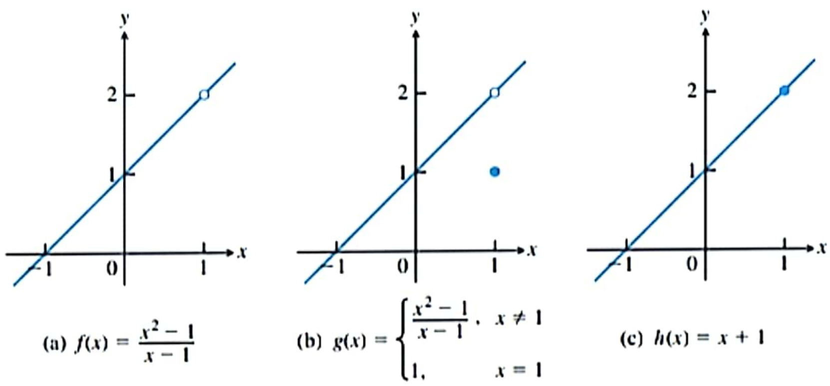


FIGURE 2.8 The limits of $f(x)$, $g(x)$, and $h(x)$ all equal 2 as x approaches 1. However, only $h(x)$ has the same function value as its limit at $x = 1$ (Example 2).

EXAMPLE 3

(a) If f is the **identity function** $f(x) = x$, then for any value of c (Figure 2.9a),

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

(b) If f is the **constant function** $f(x) = k$ (function with the constant value k), then for any value of c (Figure 2.9b),

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

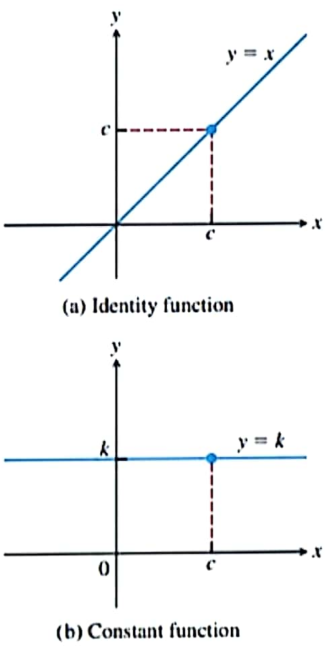


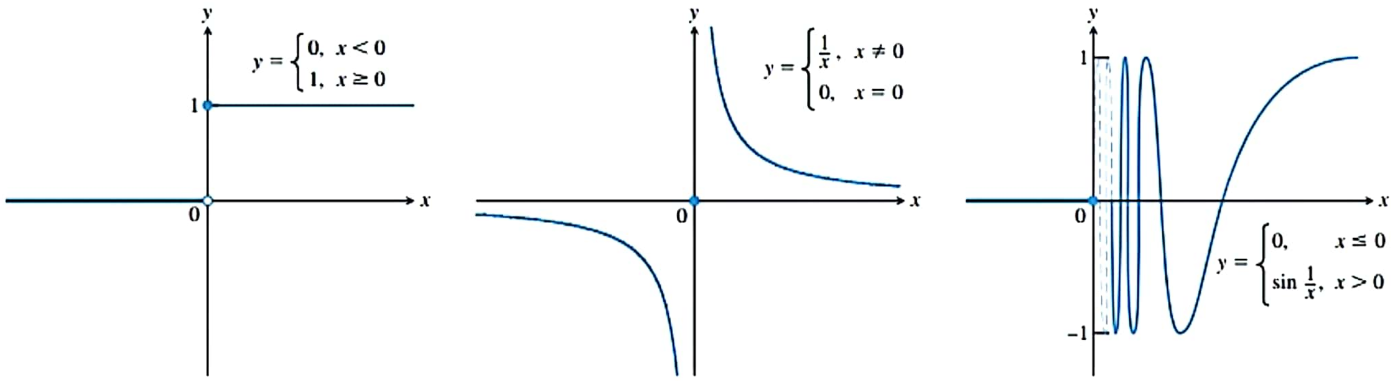
FIGURE 2.9 The functions in Example 3 have limits at all points c.

For instances of each of these rules we have

$$\lim_{x \to 3} x = 3 \qquad \text{and} \qquad \lim_{x \to -7} (4) = \lim_{x \to 2} (4) = 4.$$

We prove these rules in Example 3 in Section 2.3. ■

A function may not have a limit at a particular point. Some ways that limits can fail to exist are illustrated in Figure 2.10 and described in the next example.



(a) Unit step function $U(x)$ (b) $g(x)$ (c) $f(x)$

FIGURE 2.10 None of these functions has a limit as x approaches 0 (Example 4).

EXAMPLE 4 Discuss the behavior of the following functions, explaining why they have no limit as $x \to 0$.

(a) $U(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$

(b) $g(x) = \begin{cases} \dfrac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$

(c) $f(x) = \begin{cases} 0, & x \leq 0 \\ \sin\dfrac{1}{x}, & x > 0 \end{cases}$

Solution

(a) It *jumps*: The **unit step function** $U(x)$ has no limit as $x \to 0$ because its values jump at $x = 0$. For negative values of x arbitrarily close to zero, $U(x) = 0$. For positive values of x arbitrarily close to zero, $U(x) = 1$. There is no *single* value L approached by $U(x)$ as $x \to 0$ (Figure 2.10a).

(b) It *grows too "large" to have a limit*: $g(x)$ has no limit as $x \to 0$ because the values of g grow arbitrarily large in absolute value as $x \to 0$ and do not stay close to *any* fixed real number (Figure 2.10b). We say the function is *not bounded*.

(c) It *oscillates too much to have a limit*: $f(x)$ has no limit as $x \to 0$ because the function's values oscillate between $+1$ and -1 in every open interval containing 0. The values do not stay close to any one number as $x \to 0$ (Figure 2.10c). ■

The Limit Laws

To calculate limits of functions that are arithmetic combinations of functions having known limits, we can use several fundamental rules.

THEOREM 1—Limit Laws If L, M, c, and k are real numbers and

$$\lim_{x\to c} f(x) = L \quad \text{and} \quad \lim_{x\to c} g(x) = M, \quad \text{then}$$

1. *Sum Rule:* $\lim_{x\to c}(f(x) + g(x)) = L + M$
2. *Difference Rule:* $\lim_{x\to c}(f(x) - g(x)) = L - M$
3. *Constant Multiple Rule:* $\lim_{x\to c}(k \cdot f(x)) = k \cdot L$
4. *Product Rule:* $\lim_{x\to c}(f(x) \cdot g(x)) = L \cdot M$
5. *Quotient Rule:* $\lim_{x\to c} \dfrac{f(x)}{g(x)} = \dfrac{L}{M}, \quad M \neq 0$
6. *Power Rule:* $\lim_{x\to c}[f(x)]^n = L^n$, n a positive integer
7. *Root Rule:* $\lim_{x\to c} \sqrt[n]{f(x)} = \sqrt[n]{L} = L^{1/n}$, n a positive integer

(If n is even, we assume that $\lim_{x\to c} f(x) = L > 0$.)

In words, the Sum Rule says that the limit of a sum is the sum of the limits. Similarly, the next rules say that the limit of a difference is the difference of the limits; the limit of a constant times a function is the constant times the limit of the function; the limit of a product is the product of the limits; the limit of a quotient is the quotient of the limits (provided that the limit of the denominator is not 0); the limit of a positive integer power (or root) of a function is the integer power (or root) of the limit (provided that the root of the limit is a real number).

It is reasonable that the properties in Theorem 1 are true (although these intuitive arguments do not constitute proofs). If x is sufficiently close to c, then $f(x)$ is close to L and $g(x)$ is close to M, from our informal definition of a limit. It is then reasonable that $f(x) + g(x)$ is close to $L + M$; $f(x) - g(x)$ is close to $L - M$; $kf(x)$ is close to kL; $f(x)g(x)$ is close to LM; and $f(x)/g(x)$ is close to L/M if M is not zero. We prove the Sum Rule in Section 2.3, based on a precise definition of limit. Rules 2–5 are proved in Appendix 4. Rule 6 is obtained by applying Rule 4 repeatedly. Rule 7 is proved in more advanced texts. The Sum, Difference, and Product Rules can be extended to any number of functions, not just two.

EXAMPLE 5 Use the observations $\lim_{x\to c} k = k$ and $\lim_{x\to c} x = c$ (Example 3) and the fundamental rules of limits to find the following limits.

(a) $\lim_{x\to c}(x^3 + 4x^2 - 3)$

(b) $\lim_{x\to c} \dfrac{x^4 + x^2 - 1}{x^2 + 5}$

(c) $\lim_{x\to -2} \sqrt{4x^2 - 3}$

Solution

(a) $\lim_{x\to c}(x^3 + 4x^2 - 3) = \lim_{x\to c} x^3 + \lim_{x\to c} 4x^2 - \lim_{x\to c} 3$ — Sum and Difference Rules

$= c^3 + 4c^2 - 3$ — Power and Multiple Rules

(b) $\lim_{x\to c}\dfrac{x^4 + x^2 - 1}{x^2 + 5} = \dfrac{\lim_{x\to c}(x^4 + x^2 - 1)}{\lim_{x\to c}(x^2 + 5)}$ — Quotient Rule

$= \dfrac{\lim_{x\to c} x^4 + \lim_{x\to c} x^2 - \lim_{x\to c} 1}{\lim_{x\to c} x^2 + \lim_{x\to c} 5}$ — Sum and Difference Rules

$= \dfrac{c^4 + c^2 - 1}{c^2 + 5}$ — Power or Product Rule

(c) $\lim_{x\to -2}\sqrt{4x^2 - 3} = \sqrt{\lim_{x\to -2}(4x^2 - 3)}$ — Root Rule with $n = 2$

$= \sqrt{\lim_{x\to -2} 4x^2 - \lim_{x\to -2} 3}$ — Difference Rule

$= \sqrt{4(-2)^2 - 3}$ — Product and Multiple Rules

$= \sqrt{16 - 3}$

$= \sqrt{13}$ ■

Theorem 1 simplifies the task of calculating limits of polynomials and rational functions. To evaluate the limit of a polynomial function as x approaches c, merely substitute c for x in the formula for the function. To evaluate the limit of a rational function as x approaches a point c *at which the denominator is not zero*, substitute c for x in the formula for the function. (See Examples 5a and 5b.) We state these results formally as theorems.

THEOREM 2—Limits of Polynomials

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

$$\lim_{x\to c} P(x) = P(c) = a_n c^n + a_{n-1}c^{n-1} + \cdots + a_0.$$

THEOREM 3—Limits of Rational Functions

If $P(x)$ and $Q(x)$ are polynomials and $Q(c) \neq 0$, then

$$\lim_{x\to c}\frac{P(x)}{Q(x)} = \frac{P(c)}{Q(c)}.$$

EXAMPLE 6 The following calculation illustrates Theorems 2 and 3:

$$\lim_{x\to -1}\frac{x^3 + 4x^2 - 3}{x^2 + 5} = \frac{(-1)^3 + 4(-1)^2 - 3}{(-1)^2 + 5} = \frac{0}{6} = 0$$ ■

Identifying Common Factors

It can be shown that if $Q(x)$ is a polynomial and $Q(c) = 0$, then $(x - c)$ is a factor of $Q(x)$. Thus, if the numerator and denominator of a rational function of x are both zero at $x = c$, they have $(x - c)$ as a common factor.

Eliminating Common Factors from Zero Denominators

Theorem 3 applies only if the denominator of the rational function is not zero at the limit point c. If the denominator is zero, canceling common factors in the numerator and denominator may reduce the fraction to one whose denominator is no longer zero at c. If this happens, we can find the limit by substitution in the simplified fraction.

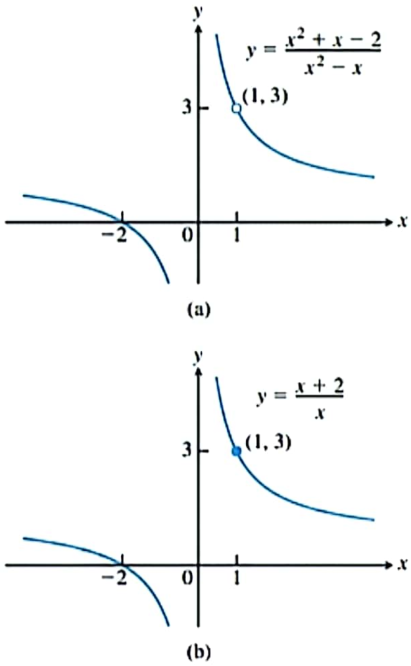


FIGURE 2.11 The graph of $f(x) = (x^2 + x - 2)/(x^2 - x)$ in part (a) is the same as the graph of $g(x) = (x + 2)/x$ in part (b) except at $x = 1$, where f is undefined. The functions have the same limit as $x \to 1$ (Example 7).

EXAMPLE 7 Evaluate

$$\lim_{x \to 1} \frac{x^2 + x - 2}{x^2 - x}.$$

Solution We cannot substitute $x = 1$ because it makes the denominator zero. We test the numerator to see if it, too, is zero at $x = 1$. It is, so it has a factor of $(x - 1)$ in common with the denominator. Canceling this common factor gives a simpler fraction with the same values as the original for $x \neq 1$:

$$\frac{x^2 + x - 2}{x^2 - x} = \frac{(x - 1)(x + 2)}{x(x - 1)} = \frac{x + 2}{x}, \qquad \text{if } x \neq 1.$$

Using the simpler fraction, we find the limit of these values as $x \to 1$ by Theorem 3:

$$\lim_{x \to 1} \frac{x^2 + x - 2}{x^2 - x} = \lim_{x \to 1} \frac{x + 2}{x} = \frac{1 + 2}{1} = 3.$$

See Figure 2.11.

Using Calculators and Computers to Estimate Limits

When we cannot use the Quotient Rule in Theorem 1 because the limit of the denominator is zero, we can try using a calculator or computer to guess the limit numerically as x gets closer and closer to c. We used this approach in Example 1, but calculators and computers can sometimes give false values and misleading impressions for functions that are undefined at a point or fail to have a limit there. Usually the problem is associated with rounding errors, as we now illustrate.

EXAMPLE 8 Estimate the value of $\displaystyle\lim_{x \to 0} \frac{\sqrt{x^2 + 100} - 10}{x^2}$.

Solution Table 2.3 lists values of the function obtained on a calculator for several points approaching $x = 0$. As x approaches 0 through the points ± 1, ± 0.5, ± 0.10, and ± 0.01, the function seems to approach the number 0.05.

As we take even smaller values of x, ± 0.0005, ± 0.0001, ± 0.00001, and ± 0.000001, the function appears to approach the number 0.

Is the answer 0.05 or 0, or some other value? We resolve this question in the next example.

TABLE 2.3 Computed values of $f(x) = \dfrac{\sqrt{x^2 + 100} - 10}{x^2}$ near $x = 0$

x	$f(x)$	
± 1	0.049876	approaches 0.05?
± 0.5	0.049969	
± 0.1	0.049999	
± 0.01	0.050000	
± 0.0005	0.050000	approaches 0?
± 0.0001	0.000000	
± 0.00001	0.000000	
± 0.000001	0.000000	

Using a computer or calculator may give ambiguous results, as in the last example. The calculator could not keep track of enough digits to avoid rounding errors in computing the values of $f(x)$ when x is very small. We cannot substitute $x = 0$ in the problem, and the numerator and denominator have no obvious common factors (as they did in Example 7). Sometimes, however, we can create a common factor algebraically.

EXAMPLE 9 Evaluate

$$\lim_{x \to 0} \frac{\sqrt{x^2 + 100} - 10}{x^2}.$$

Solution This is the limit we considered in Example 8. We can create a common factor by multiplying both numerator and denominator by the conjugate radical expression $\sqrt{x^2 + 100} + 10$ (obtained by changing the sign after the square root). The preliminary algebra rationalizes the numerator:

$$\begin{aligned}
\frac{\sqrt{x^2 + 100} - 10}{x^2} &= \frac{\sqrt{x^2 + 100} - 10}{x^2} \cdot \frac{\sqrt{x^2 + 100} + 10}{\sqrt{x^2 + 100} + 10} \\
&= \frac{x^2 + 100 - 100}{x^2\left(\sqrt{x^2 + 100} + 10\right)} \\
&= \frac{x^2}{x^2\left(\sqrt{x^2 + 100} + 10\right)} && \text{Common factor } x^2 \\
&= \frac{1}{\sqrt{x^2 + 100} + 10}. && \text{Cancel } x^2 \text{ for } x \neq 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\lim_{x \to 0} \frac{\sqrt{x^2 + 100} - 10}{x^2} &= \lim_{x \to 0} \frac{1}{\sqrt{x^2 + 100} + 10} \\
&= \frac{1}{\sqrt{0^2 + 100} + 10} && \text{Denominator not 0 at } x = 0\text{; substitute.} \\
&= \frac{1}{20} = 0.05.
\end{aligned}$$

This calculation provides the correct answer, in contrast to the ambiguous computer results in Example 8. ■

We cannot always algebraically resolve the problem of finding the limit of a quotient where the denominator becomes zero. In some cases the limit might then be found with the aid of some geometry applied to the problem (see the proof of Theorem 7 in Section 2.4), or through methods of calculus (illustrated in Section 4.5). The next theorems give helpful tools by using function comparisons.

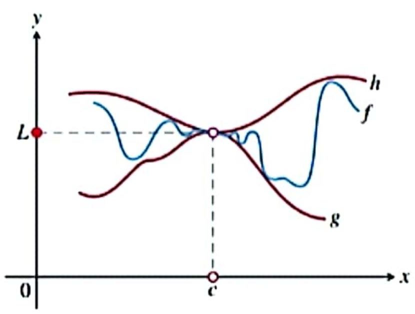


FIGURE 2.12 The graph of f is sandwiched between the graphs of g and h.

The Sandwich Theorem

The following theorem enables us to calculate a variety of limits. It is called the Sandwich Theorem because it refers to a function f whose values are sandwiched between the values of two other functions g and h that have the same limit L at a point c. Being trapped between the values of two functions that approach L, the values of f must also approach L (Figure 2.12). You will find a proof in Appendix 4.

THEOREM 4—The Sandwich Theorem Suppose that $g(x) \le f(x) \le h(x)$ for all x in some open interval containing c, except possibly at $x = c$ itself. Suppose also that

$$\lim_{x\to c} g(x) = \lim_{x\to c} h(x) = L.$$

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

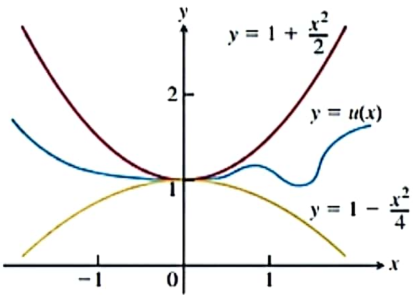


FIGURE 2.13 Any function $u(x)$ whose graph lies in the region between $y = 1 + (x^2/2)$ and $y = 1 - (x^2/4)$ has limit 1 as $x \to 0$ (Example 10).

The Sandwich Theorem is also called the Squeeze Theorem or the Pinching Theorem.

EXAMPLE 10 Given that

$$1 - \frac{x^2}{4} \le u(x) \le 1 + \frac{x^2}{2} \qquad \text{for all } x \ne 0,$$

find $\lim_{x\to 0} u(x)$, no matter how complicated u is.

Solution Since

$$\lim_{x\to 0}(1 - (x^2/4)) = 1 \qquad \text{and} \qquad \lim_{x\to 0}(1 + (x^2/2)) = 1,$$

the Sandwich Theorem implies that $\lim_{x\to 0} u(x) = 1$ (Figure 2.13). ■

EXAMPLE 11 The Sandwich Theorem helps us establish several important limit rules:

(a) $\lim_{\theta\to 0} \sin\theta = 0$ **(b)** $\lim_{\theta\to 0} \cos\theta = 1$

(c) For any function f, $\lim_{x\to c} |f(x)| = 0$ implies $\lim_{x\to c} f(x) = 0$.

Solution

(a) In Section 1.3 we established that $-|\theta| \le \sin\theta \le |\theta|$ for all θ (see Figure 2.14a). Since $\lim_{\theta\to 0}(-|\theta|) = \lim_{\theta\to 0} |\theta| = 0$, we have

$$\lim_{\theta\to 0} \sin\theta = 0.$$

(b) From Section 1.3, $0 \le 1 - \cos\theta \le |\theta|$ for all θ (see Figure 2.14b), and we have $\lim_{\theta\to 0}(1 - \cos\theta) = 0$ or

$$\lim_{\theta\to 0} \cos\theta = 1.$$

(c) Since $-|f(x)| \le f(x) \le |f(x)|$ and $-|f(x)|$ and $|f(x)|$ have limit 0 as $x \to c$, it follows that $\lim_{x\to c} f(x) = 0$. ■

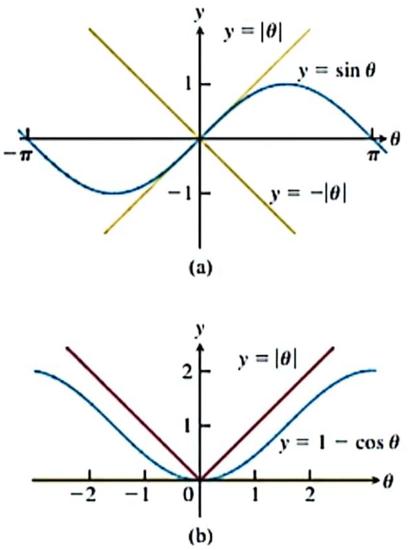


FIGURE 2.14 The Sandwich Theorem confirms the limits in Example 11.

Another important property of limits is given by the next theorem. A proof is given in the next section.

THEOREM 5 If $f(x) \le g(x)$ for all x in some open interval containing c, except possibly at $x = c$ itself, and the limits of f and g both exist as x approaches c, then

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

Caution The assertion resulting from replacing the less than or equal to ($\le$) inequality by the strict less than ($<$) inequality in Theorem 5 is false. Figure 2.14a shows that for $\theta \ne 0$, $-|\theta| < \sin\theta < |\theta|$. So $\lim_{\theta\to 0} \sin\theta = 0 = \lim_{\theta\to 0} |\theta|$, not $\lim_{\theta\to 0} \sin\theta < \lim_{\theta\to 0} |\theta|$.

16. $\lim_{h\to 0^-} \dfrac{\sqrt{6} - \sqrt{5h^2 + 11h + 6}}{h}$

17. a. $\lim_{x\to -2^+} (x + 3)\dfrac{|x + 2|}{x + 2}$ b. $\lim_{x\to -2^-} (x + 3)\dfrac{|x + 2|}{x + 2}$

18. a. $\lim_{x\to 1^+} \dfrac{\sqrt{2x}(x - 1)}{|x - 1|}$ b. $\lim_{x\to 1^-} \dfrac{\sqrt{2x}(x - 1)}{|x - 1|}$

Use the graph of the greatest integer function $y = \lfloor x \rfloor$, Figure 1.10 in Section 1.1, to help you find the limits in Exercises 19 and 20.

19. a. $\lim_{\theta\to 3^+} \dfrac{\lfloor \theta \rfloor}{\theta}$ b. $\lim_{\theta\to 3^-} \dfrac{\lfloor \theta \rfloor}{\theta}$

20. a. $\lim_{t\to 4^+} (t - \lfloor t \rfloor)$ b. $\lim_{t\to 4^-} (t - \lfloor t \rfloor)$

Using $\lim_{\theta\to 0} \dfrac{\sin\theta}{\theta} = 1$

Find the limits in Exercises 21–42.

21. $\lim_{\theta\to 0} \dfrac{\sin\sqrt{2}\theta}{\sqrt{2}\theta}$

22. $\lim_{t\to 0} \dfrac{\sin kt}{t}$ (k constant)

23. $\lim_{y\to 0} \dfrac{\sin 3y}{4y}$

24. $\lim_{h\to 0^-} \dfrac{h}{\sin 3h}$

25. $\lim_{x\to 0} \dfrac{\tan 2x}{x}$

26. $\lim_{t\to 0} \dfrac{2t}{\tan t}$

27. $\lim_{x\to 0} \dfrac{x \csc 2x}{\cos 5x}$

28. $\lim_{x\to 0} 6x^2(\cot x)(\csc 2x)$

29. $\lim_{x\to 0} \dfrac{x + x\cos x}{\sin x \cos x}$

30. $\lim_{x\to 0} \dfrac{x^2 - x + \sin x}{2x}$

31. $\lim_{\theta\to 0} \dfrac{1 - \cos\theta}{\sin 2\theta}$

32. $\lim_{x\to 0} \dfrac{x - x\cos x}{\sin^2 3x}$

33. $\lim_{t\to 0} \dfrac{\sin(1 - \cos t)}{1 - \cos t}$

34. $\lim_{h\to 0} \dfrac{\sin(\sin h)}{\sin h}$

35. $\lim_{\theta\to 0} \dfrac{\sin\theta}{\sin 2\theta}$

36. $\lim_{x\to 0} \dfrac{\sin 5x}{\sin 4x}$

37. $\lim_{\theta\to 0} \theta\cos\theta$

38. $\lim_{\theta\to 0} \sin\theta\cot 2\theta$

39. $\lim_{x\to 0} \dfrac{\tan 3x}{\sin 8x}$

40. $\lim_{y\to 0} \dfrac{\sin 3y \cot 5y}{y \cot 4y}$

41. $\lim_{\theta\to 0} \dfrac{\tan\theta}{\theta^2 \cot 3\theta}$

42. $\lim_{\theta\to 0} \dfrac{\theta \cot 4\theta}{\sin^2\theta \cot^2 2\theta}$

Theory and Examples

43. Once you know $\lim_{x\to a^+} f(x)$ and $\lim_{x\to a^-} f(x)$ at an interior point of the domain of f, do you then know $\lim_{x\to a} f(x)$? Give reasons for your answer.

44. If you know that $\lim_{x\to c} f(x)$ exists, can you find its value by calculating $\lim_{x\to c^+} f(x)$? Give reasons for your answer.

45. Suppose that f is an odd function of x. Does knowing that $\lim_{x\to 0^+} f(x) = 3$ tell you anything about $\lim_{x\to 0^-} f(x)$? Give reasons for your answer.

46. Suppose that f is an even function of x. Does knowing that $\lim_{x\to 2^-} f(x) = 7$ tell you anything about either $\lim_{x\to -2^-} f(x)$ or $\lim_{x\to -2^+} f(x)$? Give reasons for your answer.

Formal Definitions of One-Sided Limits

47. Given $\epsilon > 0$, find an interval $I = (5, 5 + \delta)$, $\delta > 0$, such that if x lies in I, then $\sqrt{x - 5} < \epsilon$. What limit is being verified and what is its value?

48. Given $\epsilon > 0$, find an interval $I = (4 - \delta, 4)$, $\delta > 0$, such that if x lies in I, then $\sqrt{4 - x} < \epsilon$. What limit is being verified and what is its value?

Use the definitions of right-hand and left-hand limits to prove the limit statements in Exercises 49 and 50.

49. $\lim_{x\to 0^-} \dfrac{x}{|x|} = -1$

50. $\lim_{x\to 2^+} \dfrac{x - 2}{|x - 2|} = 1$

51. **Greatest integer function** Find (a) $\lim_{x\to 400^+} \lfloor x \rfloor$ and (b) $\lim_{x\to 400^-} \lfloor x \rfloor$; then use limit definitions to verify your findings. (c) Based on your conclusions in parts (a) and (b), can you say anything about $\lim_{x\to 400} \lfloor x \rfloor$? Give reasons for your answer.

52. **One-sided limits** Let $f(x) = \begin{cases} x^2 \sin(1/x), & x < 0 \\ \sqrt{x}, & x > 0. \end{cases}$

Find (a) $\lim_{x\to 0^+} f(x)$ and (b) $\lim_{x\to 0^-} f(x)$; then use limit definitions to verify your findings. (c) Based on your conclusions in parts (a) and (b), can you say anything about $\lim_{x\to 0} f(x)$? Give reasons for your answer.

2.5 Continuity

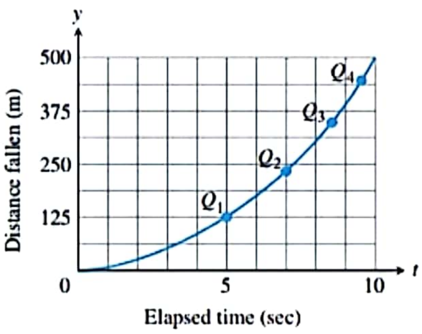


FIGURE 2.34 Connecting plotted points by an unbroken curve from experimental data $Q_1, Q_2, Q_3, \ldots$ for a falling object.

When we plot function values generated in a laboratory or collected in the field, we often connect the plotted points with an unbroken curve to show what the function's values are likely to have been at the points we did not measure (Figure 2.34). In doing so, we are assuming that we are working with a *continuous function*, so its outputs vary regularly and consistently with the inputs, and do not jump abruptly from one value to another without taking on the values in between. Intuitively, any function $y = f(x)$ whose graph can be sketched over its domain in one unbroken motion is an example of a continuous function. Such functions play an important role in the study of calculus and its applications.

Continuity at a Point

To understand continuity, it helps to consider a function like that in Figure 2.35, whose limits we investigated in Example 2 in the last section.

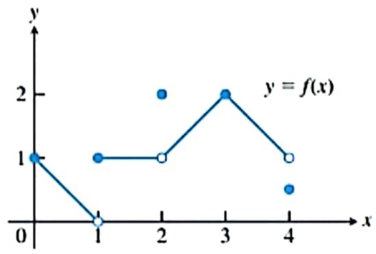


FIGURE 2.35 The function is not continuous at $x = 1, x = 2$, and $x = 4$ (Example 1).

EXAMPLE 1 At which numbers does the function f in Figure 2.35 appear to be not continuous? Explain why. What occurs at other numbers in the domain?

Solution First we observe that the domain of the function is the closed interval $[0, 4]$, so we will be considering the numbers x within that interval. From the figure, we notice right away that there are breaks in the graph at the numbers $x = 1$, $x = 2$, and $x = 4$. The breaks appear as jumps, which we identify later as "jump discontinuities." These are numbers for which the function is not continuous, and we discuss each in turn.

Numbers at which the graph of f has breaks:

At $x = 1$, the function fails to have a limit. It does have both a left-hand limit, $\lim_{x\to1^-} f(x) = 0$, as well as a right-hand limit, $\lim_{x\to1^+} f(x) = 1$, but the limit values are different, resulting in a jump in the graph. The function is not continuous at $x = 1$.

At $x = 2$, the function does have a limit, $\lim_{x\to2} f(x) = 1$, but the value of the function is $f(2) = 2$. The limit and function values are not the same, so there is a break in the graph and f is not continuous at $x = 2$.

At $x = 4$, the function does have a left-hand limit at this right endpoint, $\lim_{x\to4^-} f(x) = 1$, but again the value of the function $f(4) = \frac{1}{2}$ differs from the value of the limit. We see again a break in the graph of the function at this endpoint and the function is not continuous from the left.

Numbers at which the graph of f has no breaks:

At $x = 0$, the function has a right-hand limit at this left endpoint, $\lim_{x\to0^+} f(x) = 1$, and the value of the function is the same, $f(0) = 1$. So no break occurs in the graph of the function at this endpoint, and the function is continuous from the right at $x = 0$.

At $x = 3$, the function has a limit, $\lim_{x\to3} f(x) = 2$. Moreover, the limit is the same value as the function there, $f(3) = 2$. No break occurs in the graph and the function is continuous at $x = 3$.

At all other numbers $x = c$ in the domain, which we have not considered, the function has a limit equal to the value of the function at the point, so $\lim_{x\to c} f(x) = f(c)$. For example, $\lim_{x\to5/2} f(x) = f\left(\frac{5}{2}\right) = \frac{3}{2}$. No breaks appear in the graph of the function at any of these remaining numbers and the function is continuous at each of them. ■

The following definitions capture the continuity ideas we observed in Example 1.

DEFINITIONS Let c be a real number on the x-axis.

The function f is **continuous at c** if

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

The function f is **right-continuous at c (or continuous from the right)** if

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

The function f is **left-continuous at c (or continuous from the left)** if

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

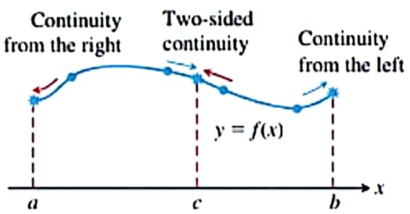


FIGURE 2.36 Continuity at points a, b, and c.

From Theorem 6, it follows immediately that a function f is continuous at an interior point c of its domain if and only if it is both right-continuous and left-continuous at c (Figure 2.36). We say that a function is **continuous over a closed interval** $[a, b]$ if it is right-continuous at a, left-continuous at b, and continuous at all interior points of the interval.

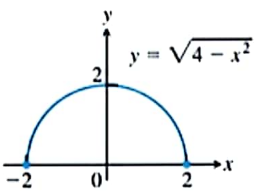


FIGURE 2.37 A function that is continuous over its domain (Example 2).

This definition applies to the infinite closed intervals $[a, \infty)$ and $(-\infty, b]$ as well, but only one endpoint is involved. If a function is not continuous at an interior point c of its domain, we say that f is **discontinuous at c**, and that c is a point of discontinuity of f. Note that a function f can be continuous, right-continuous, or left-continuous only at a point c for which $f(c)$ is defined.

EXAMPLE 2 The function $f(x) = \sqrt{4 - x^2}$ is continuous over its domain $[-2, 2]$ (Figure 2.37). It is right-continuous at $x = -2$, and left-continuous at $x = 2$. ■

EXAMPLE 3 The unit step function $U(x)$, graphed in Figure 2.38, is right-continuous at $x = 0$, but is neither left-continuous nor continuous there. It has a jump discontinuity at $x = 0$. ■

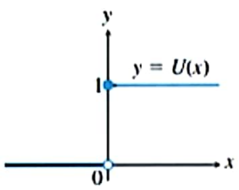


FIGURE 2.38 A function that has a jump discontinuity at the origin (Example 3).

We summarize continuity at an interior point in the form of a test.

Continuity Test

A function $f(x)$ is continuous at a point $x = c$ if and only if it meets the following three conditions.

1. $f(c)$ exists (c lies in the domain of f).
2. $\lim_{x \to c} f(x)$ exists (f has a limit as $x \to c$).
3. $\lim_{x \to c} f(x) = f(c)$ (the limit equals the function value).

For one-sided continuity and continuity at an endpoint of an interval, the limits in parts 2 and 3 of the test should be replaced by the appropriate one-sided limits.

EXAMPLE 4 The function $y = \lfloor x \rfloor$ introduced in Section 1.1 is graphed in Figure 2.39. It is discontinuous at every integer because the left-hand and right-hand limits are not equal as $x \to n$:

$$\lim_{x \to n^-} \lfloor x \rfloor = n - 1 \quad \text{and} \quad \lim_{x \to n^+} \lfloor x \rfloor = n.$$

Since $\lfloor n \rfloor = n$, the greatest integer function is right-continuous at every integer n (but not left-continuous).

The greatest integer function is continuous at every real number other than the integers. For example,

$$\lim_{x \to 1.5} \lfloor x \rfloor = 1 = \lfloor 1.5 \rfloor.$$

In general, if $n - 1 < c < n$, n an integer, then

$$\lim_{x \to c} \lfloor x \rfloor = n - 1 = \lfloor c \rfloor.$$ ■

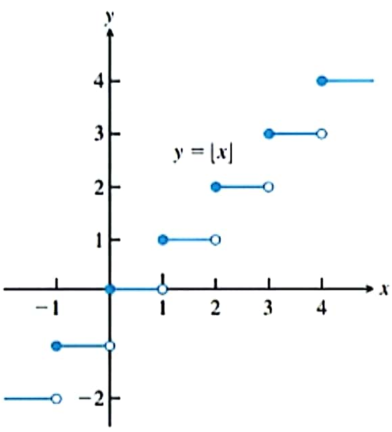


FIGURE 2.39 The greatest integer function is continuous at every noninteger point. It is right-continuous, but not left-continuous, at every integer point (Example 4).

Figure 2.40 displays several common types of discontinuities. The function in Figure 2.40a is continuous at $x = 0$. The function in Figure 2.40b would be continuous if it had $f(0) = 1$. The function in Figure 2.40c would be continuous if $f(0)$ were 1 instead of 2. The discontinuity in Figure 2.40c is **removable**. The function has a limit as $x \to 0$, and we can remove the discontinuity by setting $f(0)$ equal to this limit.

The discontinuities in Figure 2.40d through f are more serious: $\lim_{x \to 0} f(x)$ does not exist, and there is no way to improve the situation by changing f at 0. The step function in Figure 2.40d has a **jump discontinuity**: The one-sided limits exist but have different values. The function $f(x) = 1/x^2$ in Figure 2.40e has an **infinite discontinuity**. The function in Figure 2.40f has an **oscillating discontinuity**: It oscillates too much to have a limit as $x \to 0$.

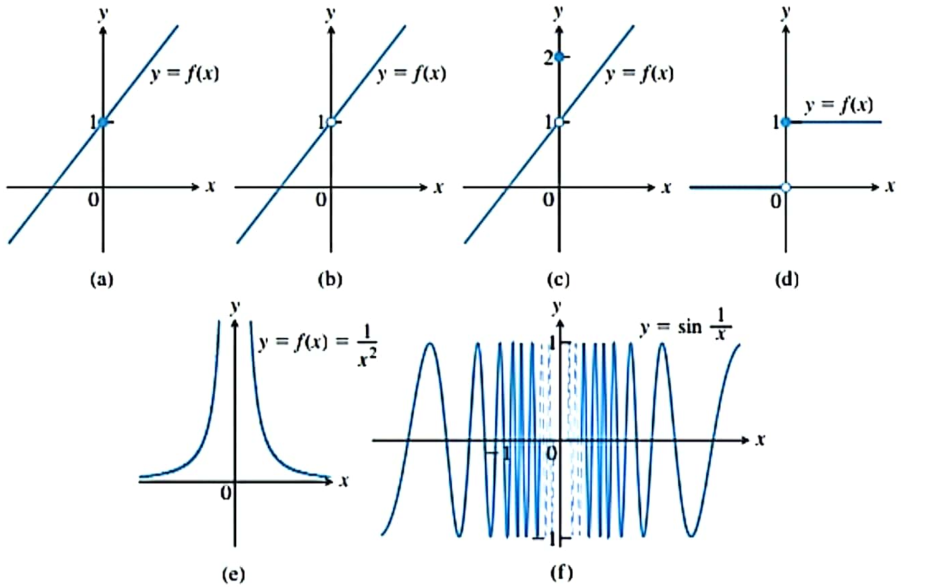


FIGURE 2.40 The function in (a) is continuous at $x = 0$; the functions in (b) through (f) are not.

Continuous Functions

Generally, we want to describe the continuity behavior of a function throughout its entire domain, not only at a single point. We know how to do that if the domain is a closed interval. In the same way, we define a **continuous function** as one that is continuous at every point in its domain. This is a property of the *function*. A function always has a specified domain, so if we change the domain, we change the function, and this may change its continuity property as well. If a function is discontinuous at one or more points of its domain, we say it is a **discontinuous** function.

EXAMPLE 5

(a) The function $y = 1/x$ (Figure 2.41) is a continuous function because it is continuous at every point of its domain. It has a point of discontinuity at $x = 0$, however, because it is not defined there; that is, it is discontinuous on any interval containing $x = 0$.

(b) The identity function $f(x) = x$ and constant functions are continuous everywhere by Example 3, Section 2.3. ■

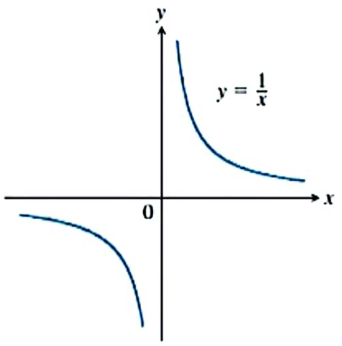


FIGURE 2.41 The function $y = 1/x$ is continuous over its natural domain. It has a point of discontinuity at the origin, so it is discontinuous on any interval containing $x = 0$ (Example 5).

Algebraic combinations of continuous functions are continuous wherever they are defined.

THEOREM 8—Properties of Continuous Functions If the functions f and g are continuous at $x = c$, then the following algebraic combinations are continuous at $x = c$.

1. *Sums:* $f + g$
2. *Differences:* $f - g$
3. *Constant multiples:* $k \cdot f$, for any number k
4. *Products:* $f \cdot g$
5. *Quotients:* f/g, provided $g(c) \neq 0$
6. *Powers:* f^n, n a positive integer
7. *Roots:* $\sqrt[n]{f}$, provided it is defined on an open interval containing c, where n is a positive integer

Continuous Extension to a Point

Sometimes the formula that describes a function f does not make sense at a point $x = c$. It might nevertheless be possible to extend the domain of f, to include $x = c$, creating a new function that is continuous at $x = c$. For example, the function $y = f(x) = (\sin x)/x$ is continuous at every point except $x = 0$, since the origin is not in its domain. Since $y = (\sin x)/x$ has a finite limit as $x \to 0$ (Theorem 7), we can extend the function's domain to include the point $x = 0$ in such a way that the extended function is continuous at $x = 0$. We define the new function

$$F(x) = \begin{cases} \dfrac{\sin x}{x}, & x \neq 0 \\ 1, & x = 0. \end{cases}$$

The function $F(x)$ is continuous at $x = 0$ because

$$\lim_{x \to 0} \frac{\sin x}{x} = F(0),$$

so it meets the requirements for continuity (Figure 2.47).

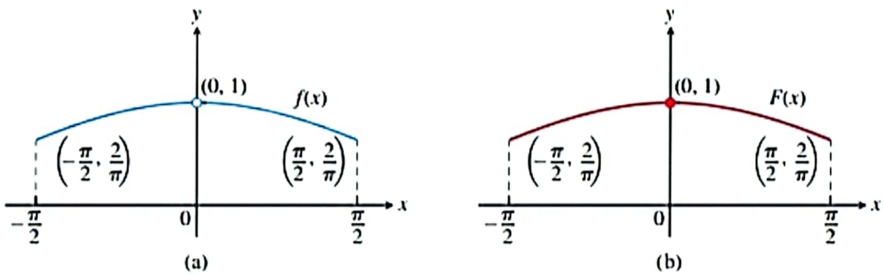


FIGURE 2.47 The graph (a) of $f(x) = (\sin x)/x$ for $-\pi/2 \le x \le \pi/2$ does not include the point (0, 1) because the function is not defined at $x = 0$. (b) We can remove the discontinuity from the graph by defining the new function $F(x)$ with $F(0) = 1$ and $F(x) = f(x)$ everywhere else. Note that $F(0) = \lim_{x \to 0} f(x)$.

More generally, a function (such as a rational function) may have a limit at a point where it is not defined. If $f(c)$ is not defined, but $\lim_{x \to c} f(x) = L$ exists, we can define a new function $F(x)$ by the rule

$$F(x) = \begin{cases} f(x), & \text{if } x \text{ is in the domain of } f \\ L, & \text{if } x = c. \end{cases}$$

The function F is continuous at $x = c$. It is called the **continuous extension of f** to $x = c$. For rational functions f, continuous extensions are often found by canceling common factors in the numerator and denominator.

EXAMPLE 12 Show that

$$f(x) = \frac{x^2 + x - 6}{x^2 - 4}, \quad x \neq 2$$

has a continuous extension to $x = 2$, and find that extension.

Solution Although $f(2)$ is not defined, if $x \neq 2$ we have

$$f(x) = \frac{x^2 + x - 6}{x^2 - 4} = \frac{(x - 2)(x + 3)}{(x - 2)(x + 2)} = \frac{x + 3}{x + 2}.$$

The new function

$$F(x) = \frac{x + 3}{x + 2}$$

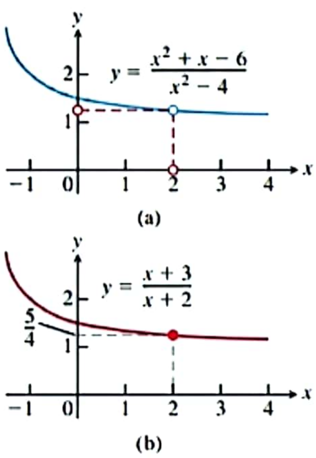


FIGURE 2.48 (a) The graph of $f(x)$ and (b) the graph of its continuous extension $F(x)$ (Example 12).

3

Derivatives

OVERVIEW In the beginning of Chapter 2, we discussed how to determine the slope of a curve at a point and how to measure the rate at which a function changes. Now that we have studied limits, we can define these ideas precisely and see that both are interpretations of the *derivative* of a function at a point. We then extend this concept from a single point to the *derivative function*, and we develop rules for finding this derivative function easily, without having to calculate any limits directly. These rules are used to find derivatives of most of the common functions reviewed in Chapter 1, as well as various combinations of them.

The derivative is one of the key ideas in calculus, and is used to study a wide range of problems in mathematics, science, economics, and medicine. These problems include finding points where a continuous function is zero, calculating the velocity and acceleration of a moving object, determining how the rate of flow of a liquid into a container changes the level of the liquid within it, describing the path followed by a light ray going from a point in air to a point in water, finding the number of items a manufacturing company should produce in order to maximize its profits, studying the spread of an infectious disease within a given population, or calculating the amount of blood the heart pumps in a minute based on how well the lungs are functioning.

3.1 Tangents and the Derivative at a Point

In this section we define the slope and tangent to a curve at a point, and the derivative of a function at a point. The derivative gives a way to find both the slope of a graph and the instantaneous rate of change of a function.

Finding a Tangent to the Graph of a Function

To find a tangent to an arbitrary curve $y = f(x)$ at a point $P(x_0, f(x_0))$, we use the procedure introduced in Section 2.1. We calculate the slope of the secant through P and a nearby point $Q(x_0 + h, f(x_0 + h))$. We then investigate the limit of the slope as $h \to 0$ (Figure 3.1). If the limit exists, we call it the slope of the curve at P and define the tangent at P to be the line through P having this slope.

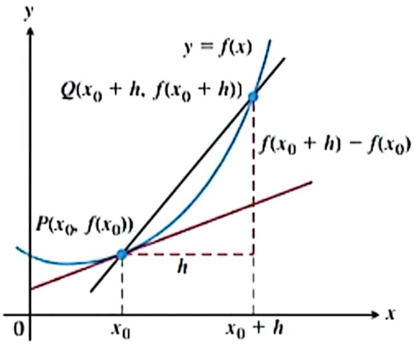


FIGURE 3.1 The slope of the tangent line at P is $\lim_{h\to 0} \dfrac{f(x_0 + h) - f(x_0)}{h}$.

DEFINITIONS The **slope of the curve** $y = f(x)$ at the point $P(x_0, f(x_0))$ is the number

$$m = \lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h} \quad \text{(provided the limit exists).}$$

The **tangent line** to the curve at P is the line through P with this slope.

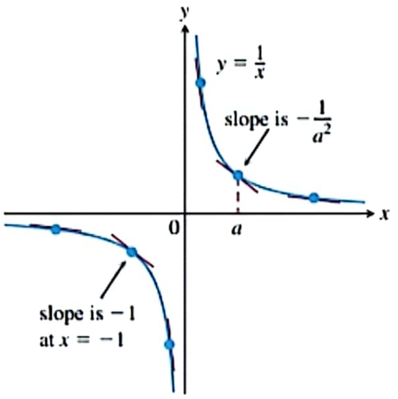


FIGURE 3.2 The tangent slopes, steep near the origin, become more gradual as the point of tangency moves away (Example 1).

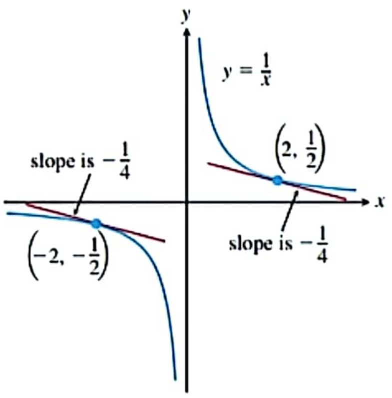


FIGURE 3.3 The two tangent lines to $y = 1/x$ having slope $-1/4$ (Example 1).

In Section 2.1, Example 3, we applied these definitions to find the slope of the parabola $f(x) = x^2$ at the point $P(2, 4)$ and the tangent line to the parabola at P. Let's look at another example.

EXAMPLE 1

(a) Find the slope of the curve $y = 1/x$ at any point $x = a \neq 0$. What is the slope at the point $x = -1$?

(b) Where does the slope equal $-1/4$?

(c) What happens to the tangent to the curve at the point $(a, 1/a)$ as a changes?

Solution

(a) Here $f(x) = 1/x$. The slope at $(a, 1/a)$ is

$$\lim_{h\to 0}\frac{f(a+h)-f(a)}{h} = \lim_{h\to 0}\frac{\dfrac{1}{a+h}-\dfrac{1}{a}}{h} = \lim_{h\to 0}\frac{1}{h}\frac{a-(a+h)}{a(a+h)}$$

$$= \lim_{h\to 0}\frac{-h}{ha(a+h)} = \lim_{h\to 0}\frac{-1}{a(a+h)} = -\frac{1}{a^2}.$$

Notice how we had to keep writing "$\lim_{h\to 0}$" before each fraction until the stage at which we could evaluate the limit by substituting $h = 0$. The number a may be positive or negative, but not 0. When $a = -1$, the slope is $-1/(-1)^2 = -1$ (Figure 3.2).

(b) The slope of $y = 1/x$ at the point where $x = a$ is $-1/a^2$. It will be $-1/4$ provided that

$$-\frac{1}{a^2} = -\frac{1}{4}.$$

This equation is equivalent to $a^2 = 4$, so $a = 2$ or $a = -2$. The curve has slope $-1/4$ at the two points $(2, 1/2)$ and $(-2, -1/2)$ (Figure 3.3).

(c) The slope $-1/a^2$ is always negative if $a \neq 0$. As $a \to 0^+$, the slope approaches $-\infty$ and the tangent becomes increasingly steep (Figure 3.2). We see this situation again as $a \to 0^-$. As a moves away from the origin in either direction, the slope approaches 0 and the tangent levels off becoming more and more horizontal. ■

Rates of Change: Derivative at a Point

The expression

$$\frac{f(x_0+h)-f(x_0)}{h}, \quad h \neq 0$$

is called the **difference quotient of f at x_0 with increment h**. If the difference quotient has a limit as h approaches zero, that limit is given a special name and notation.

The notation $f'(x_0)$ is read "f prime of x_0."

DEFINITION The **derivative of a function f at a point x_0**, denoted $f'(x_0)$, is

$$f'(x_0) = \lim_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h}$$

provided this limit exists.

If we interpret the difference quotient as the slope of a secant line, then the derivative gives the slope of the curve $y = f(x)$ at the point $P(x_0, f(x_0))$. Exercise 33 shows that the derivative of the linear function $f(x) = mx + b$ at any point x_0 is simply the slope of the line, so

$$f'(x_0) = m,$$

which is consistent with our definition of slope.

If we interpret the difference quotient as an average rate of change (Section 2.1), the derivative gives the function's instantaneous rate of change with respect to x at the point $x = x_0$. We study this interpretation in Section 3.4.

EXAMPLE 2 In Examples 1 and 2 in Section 2.1, we studied the speed of a rock falling freely from rest near the surface of the earth. We knew that the rock fell $y = 16t^2$ feet during the first t sec, and we used a sequence of average rates over increasingly short intervals to estimate the rock's speed at the instant $t = 1$. What was the rock's *exact* speed at this time?

Solution We let $f(t) = 16t^2$. The average speed of the rock over the interval between $t = 1$ and $t = 1 + h$ seconds, for $h > 0$, was found to be

$$\frac{f(1 + h) - f(1)}{h} = \frac{16(1 + h)^2 - 16(1)^2}{h} = \frac{16(h^2 + 2h)}{h} = 16(h + 2).$$

The rock's speed at the instant $t = 1$ is then

$$f'(1) = \lim_{h \to 0} 16(h + 2) = 16(0 + 2) = 32 \text{ ft/sec}.$$

Our original estimate of 32 ft/sec in Section 2.1 was right. ■

Summary

We have been discussing slopes of curves, lines tangent to a curve, the rate of change of a function, and the derivative of a function at a point. All of these ideas refer to the same limit.

> The following are all interpretations for the limit of the difference quotient,
>
> $$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$
>
> 1. The slope of the graph of $y = f(x)$ at $x = x_0$
> 2. The slope of the tangent to the curve $y = f(x)$ at $x = x_0$
> 3. The rate of change of $f(x)$ with respect to x at $x = x_0$
> 4. The derivative $f'(x_0)$ at a point

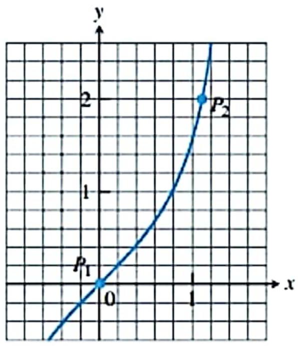
In the next sections, we allow the point x_0 to vary across the domain of the function f.

Exercises 3.1

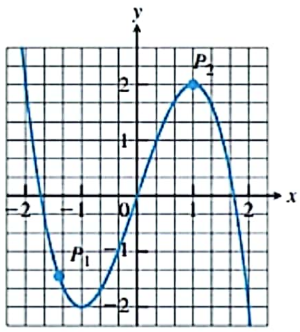
Slopes and Tangent Lines

In Exercises 1–4, use the grid and a straight edge to make a rough estimate of the slope of the curve (in y-units per x-unit) at the points P_1 and P_2.

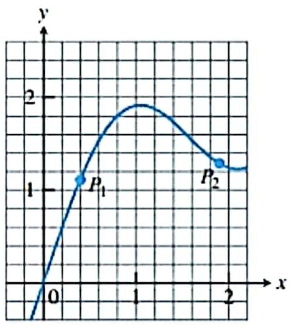
1.



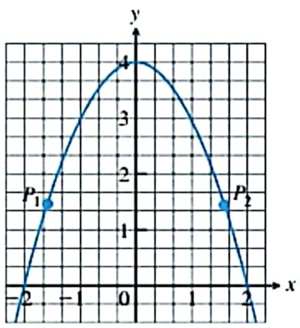
2.



3.



4.



In Exercises 5–10, find an equation for the tangent to the curve at the given point. Then sketch the curve and tangent together.

5. $y = 4 - x^2, \quad (-1, 3)$

6. $y = (x - 1)^2 + 1, \quad (1, 1)$

7. $y = 2\sqrt{x}, \quad (1, 2)$

8. $y = \dfrac{1}{x^2}, \quad (-1, 1)$

9. $y = x^3, \quad (-2, -8)$

10. $y = \dfrac{1}{x^3}, \quad \left(-2, -\dfrac{1}{8}\right)$

In Exercises 11–18, find the slope of the function's graph at the given point. Then find an equation for the line tangent to the graph there.

11. $f(x) = x^2 + 1, \quad (2, 5)$

12. $f(x) = x - 2x^2, \quad (1, -1)$

13. $g(x) = \dfrac{x}{x - 2}, \quad (3, 3)$

14. $g(x) = \dfrac{8}{x^2}, \quad (2, 2)$

15. $h(t) = t^3, \quad (2, 8)$

16. $h(t) = t^3 + 3t, \quad (1, 4)$

17. $f(x) = \sqrt{x}, \quad (4, 2)$

18. $f(x) = \sqrt{x + 1}, \quad (8, 3)$

In Exercises 19–22, find the slope of the curve at the point indicated.

19. $y = 5x - 3x^2, \quad x = 1$

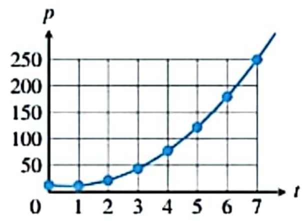
20. $y = x^3 - 2x + 7, \quad x = -2$

21. $y = \dfrac{1}{x - 1}, \quad x = 3$

22. $y = \dfrac{x - 1}{x + 1}, \quad x = 0$

Interpreting Derivative Values

23. Growth of yeast cells In a controlled laboratory experiment, yeast cells are grown in an automated cell culture system that counts the number P of cells present at hourly intervals. The number after t hours is shown in the accompanying figure.

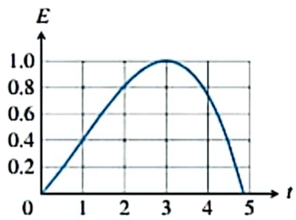


a. Explain what is meant by the derivative $P'(5)$. What are its units?

b. Which is larger, $P'(2)$ or $P'(3)$? Give a reason for your answer.

c. The quadratic curve capturing the trend of the data points (see Section 1.4) is given by $P(t) = 6.10t^2 - 9.28t + 16.43$. Find the instantaneous rate of growth when $t = 5$ hours.

24. Effectiveness of a drug On a scale from 0 to 1, the effectiveness E of a pain-killing drug t hours after entering the bloodstream is displayed in the accompanying figure.



a. At what times does the effectiveness appear to be increasing? What is true about the derivative at those times?

b. At what time would you estimate that the drug reaches its maximum effectiveness? What is true about the derivative at that time? What is true about the derivative as time increases in the 1 hour *before* your estimated time?

At what points do the graphs of the functions in Exercises 25 and 26 have horizontal tangents?

25. $f(x) = x^2 + 4x - 1$

26. $g(x) = x^3 - 3x$

27. Find equations of all lines having slope -1 that are tangent to the curve $y = 1/(x - 1)$.

28. Find an equation of the straight line having slope $1/4$ that is tangent to the curve $y = \sqrt{x}$.

Rates of Change

29. Object dropped from a tower An object is dropped from the top of a 100-m-high tower. Its height above ground after t sec is $100 - 4.9t^2$ m. How fast is it falling 2 sec after it is dropped?

3.2 The Derivative as a Function

HISTORICAL ESSAY
The Derivative

In the last section we defined the derivative of $y = f(x)$ at the point $x = x_0$ to be the limit

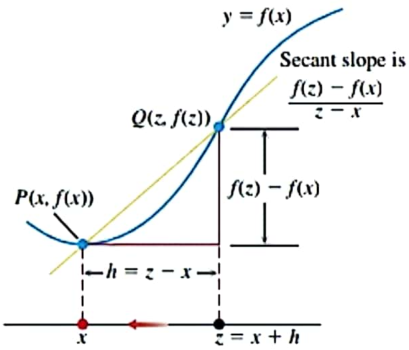
$$f'(x_0) = \lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

We now investigate the derivative as a *function* derived from f by considering the limit at each point x in the domain of f.

DEFINITION The **derivative** of the function $f(x)$ with respect to the variable x is the function f' whose value at x is

$$f'(x) = \lim_{h\to 0} \frac{f(x + h) - f(x)}{h},$$

provided the limit exists.



Derivative of f at x is

$$f'(x) = \lim_{h\to 0} \frac{f(x + h) - f(x)}{h}$$
$$= \lim_{z\to x} \frac{f(z) - f(x)}{z - x}$$

FIGURE 3.4 Two forms for the difference quotient.

We use the notation $f(x)$ in the definition to emphasize the independent variable x with respect to which the derivative function $f'(x)$ is being defined. The domain of f' is the set of points in the domain of f for which the limit exists, which means that the domain may be the same as or smaller than the domain of f. If f' exists at a particular x, we say that f is **differentiable (has a derivative) at x**. If f' exists at every point in the domain of f, we call f **differentiable**.

If we write $z = x + h$, then $h = z - x$ and h approaches 0 if and only if z approaches x. Therefore, an equivalent definition of the derivative is as follows (see Figure 3.4). This formula is sometimes more convenient to use when finding a derivative function, and focuses on the point z that approaches x.

Alternative Formula for the Derivative

$$f'(x) = \lim_{z\to x} \frac{f(z) - f(x)}{z - x}$$

Calculating Derivatives from the Definition

The process of calculating a derivative is called **differentiation**. To emphasize the idea that differentiation is an operation performed on a function $y = f(x)$, we use the notation

$$\frac{d}{dx} f(x)$$

as another way to denote the derivative $f'(x)$. Example 1 of Section 3.1 illustrated the differentiation process for the function $y = 1/x$ when $x = a$. For x representing any point in the domain, we get the formula

$$\frac{d}{dx}\left(\frac{1}{x}\right) = -\frac{1}{x^2}.$$

Derivative of the Reciprocal Function

$$\frac{d}{dx}\left(\frac{1}{x}\right) = -\frac{1}{x^2}, \quad x \neq 0$$

Here are two more examples in which we allow x to be any point in the domain of f.

EXAMPLE 1 Differentiate $f(x) = \dfrac{x}{x-1}$.

Solution We use the definition of derivative, which requires us to calculate $f(x+h)$ and then subtract $f(x)$ to obtain the numerator in the difference quotient. We have

$$f(x) = \frac{x}{x-1} \quad \text{and} \quad f(x+h) = \frac{(x+h)}{(x+h)-1}, \text{ so}$$

$$\begin{aligned}
f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} && \text{Definition} \\
&= \lim_{h\to 0} \frac{\dfrac{x+h}{x+h-1} - \dfrac{x}{x-1}}{h} \\
&= \lim_{h\to 0} \frac{1}{h} \cdot \frac{(x+h)(x-1) - x(x+h-1)}{(x+h-1)(x-1)} && \frac{a}{b} - \frac{c}{d} = \frac{ad - cb}{bd} \\
&= \lim_{h\to 0} \frac{1}{h} \cdot \frac{-h}{(x+h-1)(x-1)} && \text{Simplify.} \\
&= \lim_{h\to 0} \frac{-1}{(x+h-1)(x-1)} = \frac{-1}{(x-1)^2}. && \text{Cancel } h \neq 0.
\end{aligned}$$

EXAMPLE 2

(a) Find the derivative of $f(x) = \sqrt{x}$ for $x > 0$.

(b) Find the tangent line to the curve $y = \sqrt{x}$ at $x = 4$.

Derivative of the Square Root Function

$$\frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}}, \quad x > 0$$

Solution

(a) We use the alternative formula to calculate f':

$$\begin{aligned}
f'(x) &= \lim_{z\to x} \frac{f(z) - f(x)}{z - x} \\
&= \lim_{z\to x} \frac{\sqrt{z} - \sqrt{x}}{z - x} \\
&= \lim_{z\to x} \frac{\sqrt{z} - \sqrt{x}}{\left(\sqrt{z} - \sqrt{x}\right)\left(\sqrt{z} + \sqrt{x}\right)} \\
&= \lim_{z\to x} \frac{1}{\sqrt{z} + \sqrt{x}} = \frac{1}{2\sqrt{x}}.
\end{aligned}$$

(b) The slope of the curve at $x = 4$ is

$$f'(4) = \frac{1}{2\sqrt{4}} = \frac{1}{4}.$$

The tangent is the line through the point (4, 2) with slope 1/4 (Figure 3.5):

$$y = 2 + \frac{1}{4}(x - 4)$$

$$y = \frac{1}{4}x + 1.$$

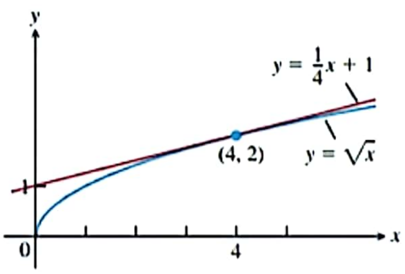


FIGURE 3.5 The curve $y = \sqrt{x}$ and its tangent at (4, 2). The tangent's slope is found by evaluating the derivative at $x = 4$ (Example 2).

Notations

There are many ways to denote the derivative of a function $y = f(x)$, where the independent variable is x and the dependent variable is y. Some common alternative notations for the derivative are

$$f'(x) = y' = \frac{dy}{dx} = \frac{df}{dx} = \frac{d}{dx}f(x) = D(f)(x) = D_x f(x).$$

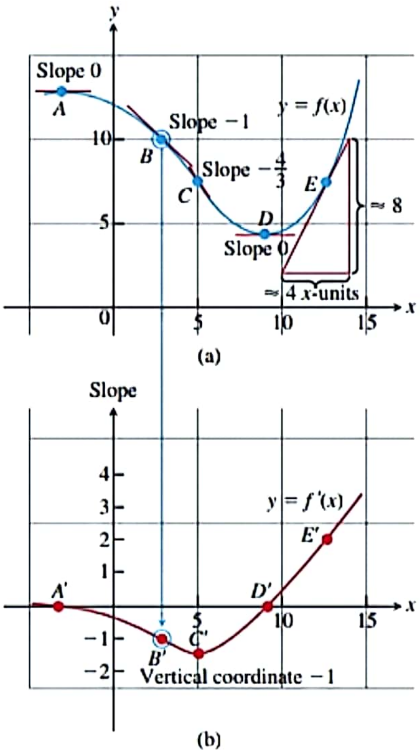


FIGURE 3.6 We made the graph of $y = f'(x)$ in (b) by plotting slopes from the graph of $y = f(x)$ in (a). The vertical coordinate of B' is the slope at B and so on. The slope at E is approximately $8/4 = 2$. In (b) we see that the rate of change of f is negative for x between A' and D'; the rate of change is positive for x to the right of D'.

The symbols d/dx and D indicate the operation of differentiation. We read dy/dx as "the derivative of y with respect to x," and df/dx and $(d/dx)f(x)$ as "the derivative of f with respect to x." The "prime" notations y' and f' come from notations that Newton used for derivatives. The d/dx notations are similar to those used by Leibniz. The symbol dy/dx should not be regarded as a ratio (until we introduce the idea of "differentials" in Section 3.11).

To indicate the value of a derivative at a specified number $x = a$, we use the notation

$$f'(a) = \left.\frac{dy}{dx}\right|_{x=a} = \left.\frac{df}{dx}\right|_{x=a} = \left.\frac{d}{dx}f(x)\right|_{x=a}.$$

For instance, in Example 2

$$f'(4) = \left.\frac{d}{dx}\sqrt{x}\right|_{x=4} = \left.\frac{1}{2\sqrt{x}}\right|_{x=4} = \frac{1}{2\sqrt{4}} = \frac{1}{4}.$$

Graphing the Derivative

We can often make a reasonable plot of the derivative of $y = f(x)$ by estimating the slopes on the graph of f. That is, we plot the points $(x, f'(x))$ in the xy-plane and connect them with a smooth curve, which represents $y = f'(x)$.

EXAMPLE 3 Graph the derivative of the function $y = f(x)$ in Figure 3.6a.

Solution We sketch the tangents to the graph of f at frequent intervals and use their slopes to estimate the values of $f'(x)$ at these points. We plot the corresponding $(x, f'(x))$ pairs and connect them with a smooth curve as sketched in Figure 3.6b. ■

What can we learn from the graph of $y = f'(x)$? At a glance we can see

1. where the rate of change of f is positive, negative, or zero;
2. the rough size of the growth rate at any x and its size in relation to the size of $f(x)$;
3. where the rate of change itself is increasing or decreasing.

Differentiable on an Interval; One-Sided Derivatives

A function $y = f(x)$ is **differentiable on an open interval** (finite or infinite) if it has a derivative at each point of the interval. It is **differentiable on a closed interval** $[a, b]$ if it is differentiable on the interior (a, b) and if the limits

$$\lim_{h\to 0^+}\frac{f(a+h) - f(a)}{h} \qquad \textbf{Right-hand derivative at } a$$

$$\lim_{h\to 0^-}\frac{f(b+h) - f(b)}{h} \qquad \textbf{Left-hand derivative at } b$$

exist at the endpoints (Figure 3.7).

Right-hand and left-hand derivatives may be defined at any point of a function's domain. Because of Theorem 6, Section 2.4, a function has a derivative at a point if and only if it has left-hand and right-hand derivatives there, and these one-sided derivatives are equal.

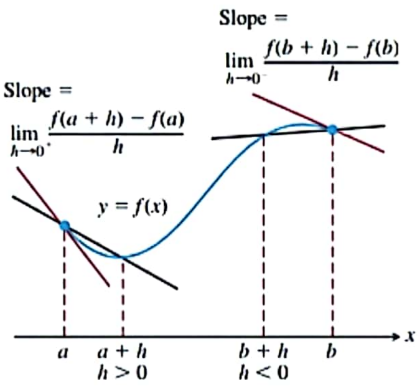


FIGURE 3.7 Derivatives at endpoints of a closed interval are one-sided limits.

EXAMPLE 4 Show that the function $y = |x|$ is differentiable on $(-\infty, 0)$ and $(0, \infty)$ but has no derivative at $x = 0$.

Solution From Section 3.1, the derivative of $y = mx + b$ is the slope m. Thus, to the right of the origin,

$$\frac{d}{dx}(|x|) = \frac{d}{dx}(x) = \frac{d}{dx}(1 \cdot x) = 1. \qquad \frac{d}{dx}(mx+b) = m,\ |x| = x$$

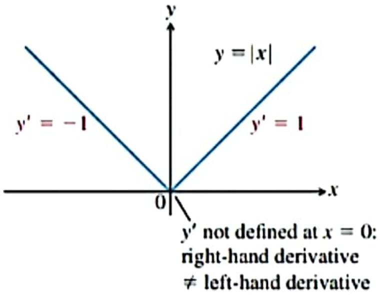


FIGURE 3.8 The function $y = |x|$ is not differentiable at the origin where the graph has a "corner" (Example 4).

To the left,

$$\frac{d}{dx}(|x|) = \frac{d}{dx}(-x) = \frac{d}{dx}(-1 \cdot x) = -1 \qquad |x| = -x$$

(Figure 3.8). There is no derivative at the origin because the one-sided derivatives differ there:

$$\begin{aligned}
\text{Right-hand derivative of } |x| \text{ at zero} &= \lim_{h\to0^+} \frac{|0+h| - |0|}{h} = \lim_{h\to0^+} \frac{|h|}{h} \\
&= \lim_{h\to0^+} \frac{h}{h} \qquad |h| = h \text{ when } h > 0 \\
&= \lim_{h\to0^+} 1 = 1 \\
\text{Left-hand derivative of } |x| \text{ at zero} &= \lim_{h\to0^-} \frac{|0+h| - |0|}{h} = \lim_{h\to0^-} \frac{|h|}{h} \\
&= \lim_{h\to0^-} \frac{-h}{h} \qquad |h| = -h \text{ when } h < 0 \\
&= \lim_{h\to0^-} -1 = -1.
\end{aligned}$$

EXAMPLE 5 In Example 2 we found that for $x > 0$,

$$\frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}}.$$

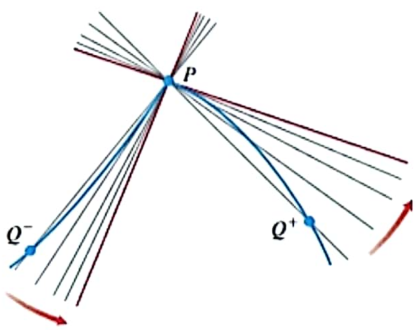
We apply the definition to examine if the derivative exists at $x = 0$:

$$\lim_{h\to0^+} \frac{\sqrt{0+h} - \sqrt{0}}{h} = \lim_{h\to0^+} \frac{1}{\sqrt{h}} = \infty.$$

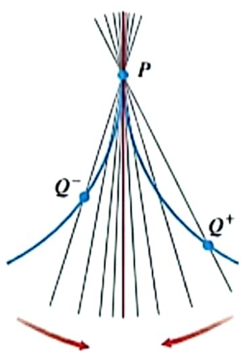
Since the (right-hand) limit is not finite, there is no derivative at $x = 0$. Since the slopes of the secant lines joining the origin to the points $(h, \sqrt{h})$ on a graph of $y = \sqrt{x}$ approach ∞, the graph has a *vertical tangent* at the origin. (See Figure 1.17 on page 9.)

When Does a Function *Not* Have a Derivative at a Point?

A function has a derivative at a point x_0 if the slopes of the secant lines through $P(x_0, f(x_0))$ and a nearby point Q on the graph approach a finite limit as Q approaches P. Whenever the secants fail to take up a limiting position or become vertical as Q approaches P, the derivative does not exist. Thus differentiability is a "smoothness" condition on the graph of f. A function can fail to have a derivative at a point for many reasons, including the existence of points where the graph has



1. a *corner*, where the one-sided derivatives differ.



2. a *cusp*, where the slope of PQ approaches ∞ from one side and $-\infty$ from the other.

3.3 Differentiation Rules

This section introduces several rules that allow us to differentiate constant functions, power functions, polynomials, exponential functions, rational functions, and certain combinations of them, simply and directly, without having to take limits each time.

Powers, Multiples, Sums, and Differences

A simple rule of differentiation is that the derivative of every constant function is zero.

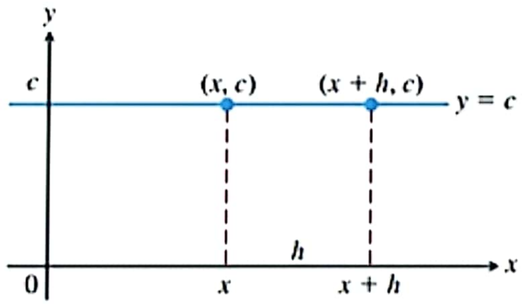


FIGURE 3.9 The rule $(d/dx)(c) = 0$ is another way to say that the values of constant functions never change and that the slope of a horizontal line is zero at every point.

> **Derivative of a Constant Function**
>
> If f has the constant value $f(x) = c$, then
>
> $$\frac{df}{dx} = \frac{d}{dx}(c) = 0.$$

Proof We apply the definition of the derivative to $f(x) = c$, the function whose outputs have the constant value c (Figure 3.9). At every value of x, we find that

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{c - c}{h} = \lim_{h \to 0} 0 = 0. \qquad \blacksquare$$

From Section 3.1, we know that

$$\frac{d}{dx}\left(\frac{1}{x}\right) = -\frac{1}{x^2}, \quad \text{or} \quad \frac{d}{dx}(x^{-1}) = -x^{-2}.$$

From Example 2 of the last section we also know that

$$\frac{d}{dx}\left(\sqrt{x}\right) = \frac{1}{2\sqrt{x}}, \quad \text{or} \quad \frac{d}{dx}(x^{1/2}) = \frac{1}{2}x^{-1/2}.$$

These two examples illustrate a general rule for differentiating a power x^n. We first prove the rule when n is a positive integer.

> **Derivative of a Positive Integer Power**
>
> If n is a positive integer, then
>
> $$\frac{d}{dx}x^n = nx^{n-1}.$$

HISTORICAL BIOGRAPHY
Richard Courant
(1888–1972)

Proof of the Positive Integer Power Rule The formula

$$z^n - x^n = (z - x)(z^{n-1} + z^{n-2}x + \cdots + zx^{n-2} + x^{n-1})$$

can be verified by multiplying out the right-hand side. Then from the alternative formula for the definition of the derivative,

$$\begin{aligned} f'(x) &= \lim_{z \to x} \frac{f(z) - f(x)}{z - x} = \lim_{z \to x} \frac{z^n - x^n}{z - x} \\ &= \lim_{z \to x}(z^{n-1} + z^{n-2}x + \cdots + zx^{n-2} + x^{n-1}) \qquad n \text{ terms} \\ &= nx^{n-1}. \end{aligned}$$

■

The Power Rule is actually valid for all real numbers n. We have seen examples for a negative integer and fractional power, but n could be an irrational number as well. To apply the Power Rule, we subtract 1 from the original exponent n and multiply the result by n. Here we state the general version of the rule, but postpone its proof until Section 3.8.

> **Power Rule (General Version)**
>
> If n is any real number, then
>
> $$\frac{d}{dx}x^n = nx^{n-1},$$
>
> for all x where the powers x^n and x^{n-1} are defined.

EXAMPLE 1 Differentiate the following powers of x.

(a) x^3 **(b)** $x^{2/3}$ **(c)** $x^{\sqrt{2}}$ **(d)** $\frac{1}{x^4}$ **(e)** $x^{-4/3}$ **(f)** $\sqrt{x^{2+\pi}}$

Solution

Applying the Power Rule
Subtract 1 from the exponent and multiply the result by the original exponent.

(a) $\frac{d}{dx}(x^3) = 3x^{3-1} = 3x^2$

(b) $\frac{d}{dx}(x^{2/3}) = \frac{2}{3}x^{(2/3)-1} = \frac{2}{3}x^{-1/3}$

(c) $\frac{d}{dx}(x^{\sqrt{2}}) = \sqrt{2}x^{\sqrt{2}-1}$

(d) $\frac{d}{dx}\left(\frac{1}{x^4}\right) = \frac{d}{dx}(x^{-4}) = -4x^{-4-1} = -4x^{-5} = -\frac{4}{x^5}$

(e) $\frac{d}{dx}(x^{-4/3}) = -\frac{4}{3}x^{-(4/3)-1} = -\frac{4}{3}x^{-7/3}$

(f) $\frac{d}{dx}\left(\sqrt{x^{2+\pi}}\right) = \frac{d}{dx}\left(x^{1+(\pi/2)}\right) = \left(1 + \frac{\pi}{2}\right)x^{1+(\pi/2)-1} = \frac{1}{2}(2 + \pi)\sqrt{x^{\pi}}$ ■

The next rule says that when a differentiable function is multiplied by a constant, its derivative is multiplied by the same constant.

Derivative Constant Multiple Rule

If u is a differentiable function of x, and c is a constant, then

$$\frac{d}{dx}(cu) = c\frac{du}{dx}.$$

In particular, if n is any real number, then

$$\frac{d}{dx}(cx^n) = cnx^{n-1}.$$

Proof

$$\begin{aligned}\frac{d}{dx}cu &= \lim_{h\to 0}\frac{cu(x+h)-cu(x)}{h} && \text{Derivative definition with } f(x) = cu(x)\\ &= c\lim_{h\to 0}\frac{u(x+h)-u(x)}{h} && \text{Constant Multiple Limit Property}\\ &= c\frac{du}{dx} && u \text{ is differentiable.}\end{aligned}$$

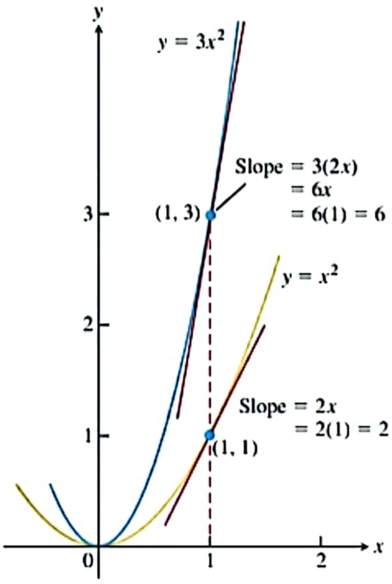


FIGURE 3.10 The graphs of $y = x^2$ and $y = 3x^2$. Tripling the y-coordinate triples the slope (Example 2).

EXAMPLE 2

(a) The derivative formula

$$\frac{d}{dx}(3x^2) = 3 \cdot 2x = 6x$$

says that if we rescale the graph of $y = x^2$ by multiplying each y-coordinate by 3, then we multiply the slope at each point by 3 (Figure 3.10).

(b) Negative of a function

The derivative of the negative of a differentiable function u is the negative of the function's derivative. The Constant Multiple Rule with $c = -1$ gives

$$\frac{d}{dx}(-u) = \frac{d}{dx}(-1 \cdot u) = -1 \cdot \frac{d}{dx}(u) = -\frac{du}{dx}.$$

Denoting Functions by u and v

The functions we are working with when we need a differentiation formula are likely to be denoted by letters like f and g. We do not want to use these same letters when stating general differentiation rules, so instead we use letters like u and v that are not likely to be already in use.

The next rule says that the derivative of the sum of two differentiable functions is the sum of their derivatives.

Derivative Sum Rule

If u and v are differentiable functions of x, then their sum $u + v$ is differentiable at every point where u and v are both differentiable. At such points,

$$\frac{d}{dx}(u+v) = \frac{du}{dx} + \frac{dv}{dx}.$$

For example, if $y = x^4 + 12x$, then y is the sum of $u(x) = x^4$ and $v(x) = 12x$. We then have

$$\frac{dy}{dx} = \frac{d}{dx}(x^4) + \frac{d}{dx}(12x) = 4x^3 + 12.$$

Proof We apply the definition of the derivative to $f(x) = u(x) + v(x)$:

$$\begin{aligned}\frac{d}{dx}[u(x) + v(x)] &= \lim_{h\to 0}\frac{[u(x+h) + v(x+h)] - [u(x) + v(x)]}{h}\\ &= \lim_{h\to 0}\left[\frac{u(x+h) - u(x)}{h} + \frac{v(x+h) - v(x)}{h}\right]\\ &= \lim_{h\to 0}\frac{u(x+h) - u(x)}{h} + \lim_{h\to 0}\frac{v(x+h) - v(x)}{h} = \frac{du}{dx} + \frac{dv}{dx}.\end{aligned}$$

■

Combining the Sum Rule with the Constant Multiple Rule gives the **Difference Rule**, which says that the derivative of a *difference* of differentiable functions is the difference of their derivatives:

$$\frac{d}{dx}(u - v) = \frac{d}{dx}[u + (-1)v] = \frac{du}{dx} + (-1)\frac{dv}{dx} = \frac{du}{dx} - \frac{dv}{dx}.$$

The Sum Rule also extends to finite sums of more than two functions. If $u_1, u_2, \ldots, u_n$ are differentiable at x, then so is $u_1 + u_2 + \cdots + u_n$, and

$$\frac{d}{dx}(u_1 + u_2 + \cdots + u_n) = \frac{du_1}{dx} + \frac{du_2}{dx} + \cdots + \frac{du_n}{dx}.$$

For instance, to see that the rule holds for three functions we compute

$$\frac{d}{dx}(u_1 + u_2 + u_3) = \frac{d}{dx}((u_1 + u_2) + u_3) = \frac{d}{dx}(u_1 + u_2) + \frac{du_3}{dx} = \frac{du_1}{dx} + \frac{du_2}{dx} + \frac{du_3}{dx}.$$

A proof by mathematical induction for any finite number of terms is given in Appendix 2.

EXAMPLE 3 Find the derivative of the polynomial $y = x^3 + \frac{4}{3}x^2 - 5x + 1$.

Solution

$$\begin{aligned}\frac{dy}{dx} &= \frac{d}{dx}x^3 + \frac{d}{dx}\left(\frac{4}{3}x^2\right) - \frac{d}{dx}(5x) + \frac{d}{dx}(1) \qquad \text{Sum and Difference Rules}\\ &= 3x^2 + \frac{4}{3}\cdot 2x - 5 + 0 = 3x^2 + \frac{8}{3}x - 5\end{aligned}$$

■

We can differentiate any polynomial term by term, the way we differentiated the polynomial in Example 3. All polynomials are differentiable at all values of x.

EXAMPLE 4 Does the curve $y = x^4 - 2x^2 + 2$ have any horizontal tangents? If so, where?

Solution The horizontal tangents, if any, occur where the slope dy/dx is zero. We have

$$\frac{dy}{dx} = \frac{d}{dx}(x^4 - 2x^2 + 2) = 4x^3 - 4x.$$

Now solve the equation $\frac{dy}{dx} = 0$ for x:

$$\begin{aligned}4x^3 - 4x &= 0\\ 4x(x^2 - 1) &= 0\\ x &= 0, 1, -1.\end{aligned}$$

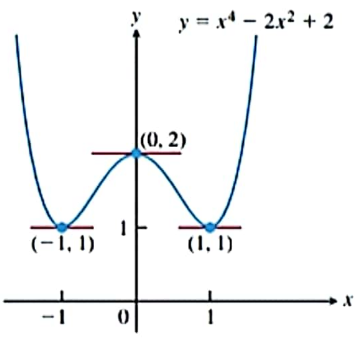


FIGURE 3.11 The curve in Example 4 and its horizontal tangents.

The curve $y = x^4 - 2x^2 + 2$ has horizontal tangents at $x = 0, 1$, and -1. The corresponding points on the curve are $(0, 2)$, $(1, 1)$, and $(-1, 1)$. See Figure 3.11. We will see in Chapter 4 that finding the values of x where the derivative of a function is equal to zero is an important and useful procedure. ■

Derivatives of Exponential Functions

We briefly reviewed exponential functions in Section 1.5. When we apply the definition of the derivative to $f(x) = a^x$, we get

$$\begin{aligned}
\frac{d}{dx}(a^x) &= \lim_{h\to 0}\frac{a^{x+h} - a^x}{h} && \text{Derivative definition}\\
&= \lim_{h\to 0}\frac{a^x \cdot a^h - a^x}{h} && a^{x+h} = a^x \cdot a^h\\
&= \lim_{h\to 0} a^x \cdot \frac{a^h - 1}{h} && \text{Factoring out } a^x\\
&= a^x \cdot \lim_{h\to 0}\frac{a^h - 1}{h} && a^x \text{ is constant as } h \to 0.\\
&= \underbrace{\left(\lim_{h\to 0}\frac{a^h - 1}{h}\right)}_{\text{a fixed number } L} \cdot a^x. && (1)
\end{aligned}$$

Thus we see that the derivative of a^x is a constant multiple L of a^x. The constant L is a limit unlike any we have encountered before. Note, however, that it equals the derivative of $f(x) = a^x$ at $x = 0$:

$$f'(0) = \lim_{h\to 0}\frac{a^h - a^0}{h} = \lim_{h\to 0}\frac{a^h - 1}{h} = L.$$

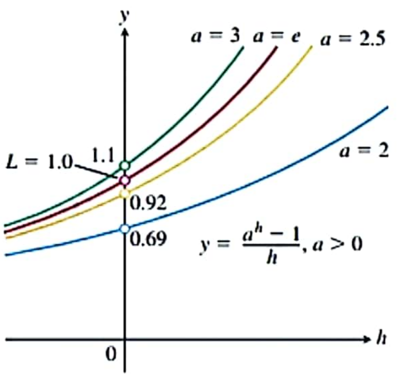


FIGURE 3.12 The position of the curve $y = (a^h - 1)/h$, $a > 0$, varies continuously with a. The limit L of y as $h \to 0$ changes with different values of a. The number for which $L = 1$ as $h \to 0$ is the number e between $a = 2$ and $a = 3$.

The limit L is therefore the slope of the graph of $f(x) = a^x$ where it crosses the y-axis. In Chapter 7, where we carefully develop the logarithmic and exponential functions, we prove that the limit L exists and has the value $\ln a$. For now we investigate values of L by graphing the function $y = (a^h - 1)/h$ and studying its behavior as h approaches 0.

Figure 3.12 shows the graphs of $y = (a^h - 1)/h$ for four different values of a. The limit L is approximately 0.69 if $a = 2$, about 0.92 if $a = 2.5$, and about 1.1 if $a = 3$. It appears that the value of L is 1 at some number a chosen between 2.5 and 3. That number is given by $a = e \approx 2.718281828$. With this choice of base we obtain the natural exponential function $f(x) = e^x$ as in Section 1.5, and see that it satisfies the property

$$f'(0) = \lim_{h\to 0}\frac{e^h - 1}{h} = 1 \qquad (2)$$

because it is the exponential function whose graph has slope 1 when it crosses the y-axis. That the limit is 1 implies an important relationship between the natural exponential function e^x and its derivative:

$$\begin{aligned}
\frac{d}{dx}(e^x) &= \lim_{h\to 0}\left(\frac{e^h - 1}{h}\right)\cdot e^x && \text{Eq. (1) with } a = e\\
&= 1 \cdot e^x = e^x. && \text{Eq. (2)}
\end{aligned}$$

Therefore the natural exponential function is its own derivative.

Derivative of the Natural Exponential Function

$$\frac{d}{dx}(e^x) = e^x$$

EXAMPLE 5 Find an equation for a line that is tangent to the graph of $y = e^x$ and goes through the origin.

Solution Since the line passes through the origin, its equation is of the form $y = mx$, where m is the slope. If it is tangent to the graph at the point (a, e^a), the slope is $m = (e^a - 0)/(a - 0)$. The slope of the natural exponential at $x = a$ is e^a. Because these

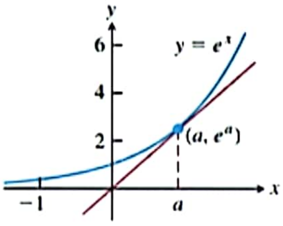


FIGURE 3.13 The line through the origin is tangent to the graph of $y = e^x$ when $a = 1$ (Example 5).

slopes are the same, we then have that $e^a = e^a/a$. It follows that $a = 1$ and $m = e$, so the equation of the tangent line is $y = ex$. See Figure 3.13. ■

We might ask if there are functions *other* than the natural exponential function that are their own derivatives. The answer is that the only functions that satisfy the property that $f'(x) = f(x)$ are functions that are constant multiples of the natural exponential function, $f(x) = c \cdot e^x$, c any constant. We prove this fact in Section 7.2. Note from the Constant Multiple Rule that indeed

$$\frac{d}{dx}(c \cdot e^x) = c \cdot \frac{d}{dx}(e^x) = c \cdot e^x.$$

Products and Quotients

While the derivative of the sum of two functions is the sum of their derivatives, the derivative of the product of two functions is *not* the product of their derivatives. For instance,

$$\frac{d}{dx}(x \cdot x) = \frac{d}{dx}(x^2) = 2x, \qquad \text{while} \qquad \frac{d}{dx}(x) \cdot \frac{d}{dx}(x) = 1 \cdot 1 = 1.$$

The derivative of a product of two functions is the sum of *two* products, as we now explain.

Derivative Product Rule

If u and v are differentiable at x, then so is their product uv, and

$$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}.$$

Equation (3) is equivalent to saying that

$$(fg)' = f'g + fg'.$$

This form of the Product Rule is useful and applies to *dot products* and *cross products* of *vector-valued* functions, studied in Chapter 13.

The derivative of the product uv is u times the derivative of v plus v times the derivative of u. In *prime notation*, $(uv)' = uv' + vu'$. In function notation,

$$\frac{d}{dx}[f(x)g(x)] = f(x)g'(x) + g(x)f'(x). \tag{3}$$

EXAMPLE 6 Find the derivative of **(a)** $y = \frac{1}{x}(x^2 + e^x)$, **(b)** $y = e^{2x}$.

Solution

(a) We apply the Product Rule with $u = 1/x$ and $v = x^2 + e^x$:

$$\begin{aligned}\frac{d}{dx}\left[\frac{1}{x}(x^2 + e^x)\right] &= \frac{1}{x}(2x + e^x) + (x^2 + e^x)\left(-\frac{1}{x^2}\right) \\ &= 2 + \frac{e^x}{x} - 1 - \frac{e^x}{x^2} \\ &= 1 + (x - 1)\frac{e^x}{x^2}.\end{aligned}$$

$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}$, and $\frac{d}{dx}\left(\frac{1}{x}\right) = -\frac{1}{x^2}$

(b) $\frac{d}{dx}(e^{2x}) = \frac{d}{dx}(e^x \cdot e^x) = e^x \cdot \frac{d}{dx}(e^x) + e^x \cdot \frac{d}{dx}(e^x) = 2e^x \cdot e^x = 2e^{2x}$ ■

EXAMPLE 7 Find the derivative of $y = (x^2 + 1)(x^3 + 3)$.

Solution

(a) From the Product Rule with $u = x^2 + 1$ and $v = x^3 + 3$, we find

$$\begin{aligned}\frac{d}{dx}[(x^2 + 1)(x^3 + 3)] &= (x^2 + 1)(3x^2) + (x^3 + 3)(2x) \\ &= 3x^4 + 3x^2 + 2x^4 + 6x \\ &= 5x^4 + 3x^2 + 6x.\end{aligned}$$

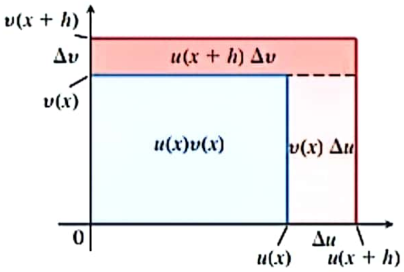
$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}$

(b) This particular product can be differentiated as well (perhaps better) by multiplying out the original expression for y and differentiating the resulting polynomial:

$$y = (x^2 + 1)(x^3 + 3) = x^5 + x^3 + 3x^2 + 3$$
$$\frac{dy}{dx} = 5x^4 + 3x^2 + 6x.$$

This is in agreement with our first calculation. ■

Picturing the Product Rule
Suppose $u(x)$ and $v(x)$ are positive and increase when x increases, and $h > 0$.



Then the change in the product uv is the difference in areas of the larger and smaller "squares," which is the sum of the upper and right-hand reddish-shaded rectangles. That is,

$$\begin{aligned}\Delta(uv) &= u(x + h)v(x + h) - u(x)v(x)\\ &= u(x + h)\Delta v + v(x)\Delta u.\end{aligned}$$

Division by h gives

$$\frac{\Delta(uv)}{h} = u(x + h)\frac{\Delta v}{h} + v(x)\frac{\Delta u}{h}.$$

The limit as $h \to 0^+$ gives the Product Rule.

Proof of the Derivative Product Rule

$$\frac{d}{dx}(uv) = \lim_{h\to 0}\frac{u(x + h)v(x + h) - u(x)v(x)}{h}$$

To change this fraction into an equivalent one that contains difference quotients for the derivatives of u and v, we subtract and add $u(x + h)v(x)$ in the numerator:

$$\begin{aligned}\frac{d}{dx}(uv) &= \lim_{h\to 0}\frac{u(x + h)v(x + h) - u(x + h)v(x) + u(x + h)v(x) - u(x)v(x)}{h}\\ &= \lim_{h\to 0}\left[u(x + h)\frac{v(x + h) - v(x)}{h} + v(x)\frac{u(x + h) - u(x)}{h}\right]\\ &= \lim_{h\to 0}u(x + h)\cdot\lim_{h\to 0}\frac{v(x + h) - v(x)}{h} + v(x)\cdot\lim_{h\to 0}\frac{u(x + h) - u(x)}{h}.\end{aligned}$$

As h approaches zero, $u(x + h)$ approaches $u(x)$ because u, being differentiable at x, is continuous at x. The two fractions approach the values of dv/dx at x and du/dx at x. Therefore,

$$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}.$$ ■

The derivative of the quotient of two functions is given by the Quotient Rule.

> **Derivative Quotient Rule**
>
> If u and v are differentiable at x and if $v(x) \neq 0$, then the quotient u/v is differentiable at x, and
>
> $$\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v\dfrac{du}{dx} - u\dfrac{dv}{dx}}{v^2}.$$

In function notation,

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)f'(x) - f(x)g'(x)}{g^2(x)}.$$

EXAMPLE 8 Find the derivative of **(a)** $y = \dfrac{t^2 - 1}{t^3 + 1}$, **(b)** $y = e^{-x}$.

Solution

(a) We apply the Quotient Rule with $u = t^2 - 1$ and $v = t^3 + 1$:

$$\begin{aligned}\frac{dy}{dt} &= \frac{(t^3 + 1)\cdot 2t - (t^2 - 1)\cdot 3t^2}{(t^3 + 1)^2}\\ &= \frac{2t^4 + 2t - 3t^4 + 3t^2}{(t^3 + 1)^2}\\ &= \frac{-t^4 + 3t^2 + 2t}{(t^3 + 1)^2}.\end{aligned}$$

$$\frac{d}{dt}\left(\frac{u}{v}\right) = \frac{v(du/dt) - u(dv/dt)}{v^2}$$

(b) $\dfrac{d}{dx}(e^{-x}) = \dfrac{d}{dx}\left(\dfrac{1}{e^x}\right) = \dfrac{e^x \cdot 0 - 1 \cdot e^x}{(e^x)^2} = \dfrac{-1}{e^x} = -e^{-x}$

Proof of the Derivative Quotient Rule

$$\frac{d}{dx}\left(\frac{u}{v}\right) = \lim_{h \to 0} \frac{\dfrac{u(x+h)}{v(x+h)} - \dfrac{u(x)}{v(x)}}{h}$$
$$= \lim_{h \to 0} \frac{v(x)u(x+h) - u(x)v(x+h)}{hv(x+h)v(x)}$$

To change the last fraction into an equivalent one that contains the difference quotients for the derivatives of u and v, we subtract and add $v(x)u(x)$ in the numerator. We then get

$$\frac{d}{dx}\left(\frac{u}{v}\right) = \lim_{h \to 0} \frac{v(x)u(x+h) - v(x)u(x) + v(x)u(x) - u(x)v(x+h)}{hv(x+h)v(x)}$$
$$= \lim_{h \to 0} \frac{v(x)\dfrac{u(x+h) - u(x)}{h} - u(x)\dfrac{v(x+h) - v(x)}{h}}{v(x+h)v(x)}.$$

Taking the limits in the numerator and denominator now gives the Quotient Rule. Exercise 74 outlines another proof.

The choice of which rules to use in solving a differentiation problem can make a difference in how much work you have to do. Here is an example.

EXAMPLE 9 Find the derivative of

$$y = \frac{(x-1)(x^2-2x)}{x^4}.$$

Solution Using the Quotient Rule here will result in a complicated expression with many terms. Instead, use some algebra to simplify the expression. First expand the numerator and divide by x^4:

$$y = \frac{(x-1)(x^2-2x)}{x^4} = \frac{x^3 - 3x^2 + 2x}{x^4} = x^{-1} - 3x^{-2} + 2x^{-3}.$$

Then use the Sum and Power Rules:

$$\frac{dy}{dx} = -x^{-2} - 3(-2)x^{-3} + 2(-3)x^{-4}$$
$$= -\frac{1}{x^2} + \frac{6}{x^3} - \frac{6}{x^4}.$$

Second- and Higher-Order Derivatives

If $y = f(x)$ is a differentiable function, then its derivative $f'(x)$ is also a function. If f' is also differentiable, then we can differentiate f' to get a new function of x denoted by f''. So $f'' = (f')'$. The function f'' is called the **second derivative** of f because it is the derivative of the first derivative. It is written in several ways:

$$f''(x) = \frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{dy'}{dx} = y'' = D^2(f)(x) = D_x^2 f(x).$$

The symbol D^2 means the operation of differentiation is performed twice.

If $y = x^6$, then $y' = 6x^5$ and we have

$$y'' = \frac{dy'}{dx} = \frac{d}{dx}(6x^5) = 30x^4.$$

Thus $D^2(x^6) = 30x^4$.

How to Read the Symbols for Derivatives

y'	"y prime"
y''	"y double prime"
$\dfrac{d^2y}{dx^2}$	"d squared y dx squared"
y'''	"y triple prime"
$y^{(n)}$	"y super n"
$\dfrac{d^ny}{dx^n}$	"d to the n of y by dx to the n"
D^n	"D to the n"

If y'' is differentiable, its derivative, $y''' = dy''/dx = d^3y/dx^3$, is the **third derivative** of y with respect to x. The names continue as you imagine, with

$$y^{(n)} = \frac{d}{dx}y^{(n-1)} = \frac{d^ny}{dx^n} = D^ny$$

denoting the ***n*th derivative** of y with respect to x for any positive integer n.

We can interpret the second derivative as the rate of change of the slope of the tangent to the graph of $y = f(x)$ at each point. You will see in the next chapter that the second derivative reveals whether the graph bends upward or downward from the tangent line as we move off the point of tangency. In the next section, we interpret both the second and third derivatives in terms of motion along a straight line.

EXAMPLE 10 The first four derivatives of $y = x^3 - 3x^2 + 2$ are

First derivative: $y' = 3x^2 - 6x$

Second derivative: $y'' = 6x - 6$

Third derivative: $y''' = 6$

Fourth derivative: $y^{(4)} = 0.$

All polynomial functions have derivatives of all orders. In this example, the fifth and later derivatives are all zero. ■

3.5 Derivatives of Trigonometric Functions

Many phenomena of nature are approximately periodic (electromagnetic fields, heart rhythms, tides, weather). The derivatives of sines and cosines play a key role in describing periodic changes. This section shows how to differentiate the six basic trigonometric functions.

Derivative of the Sine Function

To calculate the derivative of $f(x) = \sin x$, for x measured in radians, we combine the limits in Example 5a and Theorem 7 in Section 2.4 with the angle sum identity for the sine function:

$$\sin(x + h) = \sin x \cos h + \cos x \sin h.$$

If $f(x) = \sin x$, then

$$\begin{aligned}
f'(x) &= \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} = \lim_{h \to 0} \frac{\sin(x + h) - \sin x}{h} && \text{Derivative definition} \\
&= \lim_{h \to 0} \frac{(\sin x \cos h + \cos x \sin h) - \sin x}{h} \\
&= \lim_{h \to 0} \frac{\sin x(\cos h - 1) + \cos x \sin h}{h} \\
&= \lim_{h \to 0} \left(\sin x \cdot \frac{\cos h - 1}{h} \right) + \lim_{h \to 0} \left(\cos x \cdot \frac{\sin h}{h} \right) \\
&= \sin x \cdot \underbrace{\lim_{h \to 0} \frac{\cos h - 1}{h}}_{\text{limit } 0} + \cos x \cdot \underbrace{\lim_{h \to 0} \frac{\sin h}{h}}_{\text{limit } 1} = \sin x \cdot 0 + \cos x \cdot 1 = \cos x. && \text{Example 5a and Theorem 7, Section 2.4}
\end{aligned}$$

The derivative of the sine function is the cosine function:

$$\frac{d}{dx}(\sin x) = \cos x.$$

EXAMPLE 1 We find derivatives of the sine function involving differences, products, and quotients.

(a) $y = x^2 - \sin x$:

$$\begin{aligned}
\frac{dy}{dx} &= 2x - \frac{d}{dx}(\sin x) && \text{Difference Rule} \\
&= 2x - \cos x
\end{aligned}$$

(b) $y = e^x \sin x$:

$$\begin{aligned}
\frac{dy}{dx} &= e^x \frac{d}{dx}(\sin x) + \frac{d}{dx}(e^x) \sin x && \text{Product Rule} \\
&= e^x \cos x + e^x \sin x \\
&= e^x (\cos x + \sin x)
\end{aligned}$$

(c) $y = \dfrac{\sin x}{x}$:

$$\begin{aligned}
\frac{dy}{dx} &= \frac{x \cdot \frac{d}{dx}(\sin x) - \sin x \cdot 1}{x^2} && \text{Quotient Rule} \\
&= \frac{x \cos x - \sin x}{x^2}
\end{aligned}$$

Derivative of the Cosine Function

With the help of the angle sum formula for the cosine function,

$$\cos(x + h) = \cos x \cos h - \sin x \sin h,$$

we can compute the limit of the difference quotient:

$$\begin{aligned}
\frac{d}{dx}(\cos x) &= \lim_{h \to 0} \frac{\cos(x + h) - \cos x}{h} && \text{Derivative definition} \\
&= \lim_{h \to 0} \frac{(\cos x \cos h - \sin x \sin h) - \cos x}{h} && \text{Cosine angle sum identity} \\
&= \lim_{h \to 0} \frac{\cos x(\cos h - 1) - \sin x \sin h}{h} \\
&= \lim_{h \to 0} \cos x \cdot \frac{\cos h - 1}{h} - \lim_{h \to 0} \sin x \cdot \frac{\sin h}{h} \\
&= \cos x \cdot \lim_{h \to 0} \frac{\cos h - 1}{h} - \sin x \cdot \lim_{h \to 0} \frac{\sin h}{h} \\
&= \cos x \cdot 0 - \sin x \cdot 1 && \text{Example 5a and Theorem 7, Section 2.4} \\
&= -\sin x.
\end{aligned}$$

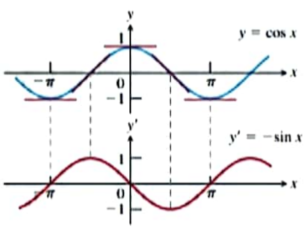


FIGURE 3.22 The curve $y' = -\sin x$ as the graph of the slopes of the tangents to the curve $y = \cos x$.

The derivative of the cosine function is the negative of the sine function:

$$\frac{d}{dx}(\cos x) = -\sin x.$$

Figure 3.22 shows a way to visualize this result in the same way we did for graphing derivatives in Section 3.2, Figure 3.6.

EXAMPLE 2 We find derivatives of the cosine function in combinations with other functions.

(a) $y = 5e^x + \cos x$:

$$\frac{dy}{dx} = \frac{d}{dx}(5e^x) + \frac{d}{dx}(\cos x) \qquad \text{Sum Rule}$$
$$= 5e^x - \sin x$$

(b) $y = \sin x \cos x$:

$$\frac{dy}{dx} = \sin x \frac{d}{dx}(\cos x) + \cos x \frac{d}{dx}(\sin x) \qquad \text{Product Rule}$$
$$= \sin x(-\sin x) + \cos x(\cos x)$$
$$= \cos^2 x - \sin^2 x$$

(c) $y = \dfrac{\cos x}{1 - \sin x}$:

$$\frac{dy}{dx} = \frac{(1 - \sin x)\frac{d}{dx}(\cos x) - \cos x \frac{d}{dx}(1 - \sin x)}{(1 - \sin x)^2} \qquad \text{Quotient Rule}$$
$$= \frac{(1 - \sin x)(-\sin x) - \cos x(0 - \cos x)}{(1 - \sin x)^2}$$
$$= \frac{1 - \sin x}{(1 - \sin x)^2} \qquad \sin^2 x + \cos^2 x = 1$$
$$= \frac{1}{1 - \sin x}$$

Simple Harmonic Motion

The motion of an object or weight bobbing freely up and down with no resistance on the end of a spring is an example of *simple harmonic motion*. The motion is periodic and repeats indefinitely, so we represent it using trigonometric functions. The next example describes a case in which there are no opposing forces such as friction to slow the motion.

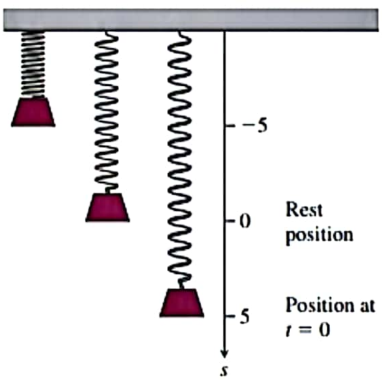


FIGURE 3.23 A weight hanging from a vertical spring and then displaced oscillates above and below its rest position (Example 3).

EXAMPLE 3 A weight hanging from a spring (Figure 3.23) is stretched down 5 units beyond its rest position and released at time $t = 0$ to bob up and down. Its position at any later time t is

$$s = 5 \cos t.$$

What are its velocity and acceleration at time t?

Solution We have

Position: $s = 5 \cos t$

Velocity: $v = \dfrac{ds}{dt} = \dfrac{d}{dt}(5 \cos t) = -5 \sin t$

Acceleration: $a = \dfrac{dv}{dt} = \dfrac{d}{dt}(-5 \sin t) = -5 \cos t.$

Notice how much we can learn from these equations:

1. As time passes, the weight moves down and up between $s = -5$ and $s = 5$ on the s-axis. The amplitude of the motion is 5. The period of the motion is 2π, the period of the cosine function.
2. The velocity $v = -5 \sin t$ attains its greatest magnitude, 5, when $\cos t = 0$, as the graphs show in Figure 3.24. Hence, the speed of the weight, $|v| = 5|\sin t|$, is greatest

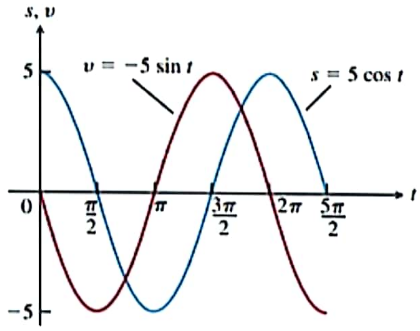


FIGURE 3.24 The graphs of the position and velocity of the weight in Example 3.

when $\cos t = 0$, that is, when $s = 0$ (the rest position). The speed of the weight is zero when $\sin t = 0$. This occurs when $s = 5\cos t = \pm 5$, at the endpoints of the interval of motion.

3. The weight is acted on by the spring and by gravity. When the weight is below the rest position, the combined forces pull it up, and when it is above the rest position, they pull it down. The weight's acceleration is always proportional to the negative of its displacement. This property of springs is called *Hooke's Law*, and is studied further in Section 6.5.

4. The acceleration, $a = -5\cos t$, is zero only at the rest position, where $\cos t = 0$ and the force of gravity and the force from the spring balance each other. When the weight is anywhere else, the two forces are unequal and acceleration is nonzero. The acceleration is greatest in magnitude at the points farthest from the rest position, where $\cos t = \pm 1$. ■

EXAMPLE 4 The jerk associated with the simple harmonic motion in Example 3 is

$$j = \frac{da}{dt} = \frac{d}{dt}(-5\cos t) = 5\sin t.$$

It has its greatest magnitude when $\sin t = \pm 1$, not at the extremes of the displacement but at the rest position, where the acceleration changes direction and sign. ■

Derivatives of the Other Basic Trigonometric Functions

Because $\sin x$ and $\cos x$ are differentiable functions of x, the related functions

$$\tan x = \frac{\sin x}{\cos x}, \qquad \cot x = \frac{\cos x}{\sin x}, \qquad \sec x = \frac{1}{\cos x}, \qquad \text{and} \qquad \csc x = \frac{1}{\sin x}$$

are differentiable at every value of x at which they are defined. Their derivatives, calculated from the Quotient Rule, are given by the following formulas. Notice the negative signs in the derivative formulas for the cofunctions.

The derivatives of the other trigonometric functions:

$$\frac{d}{dx}(\tan x) = \sec^2 x \qquad \frac{d}{dx}(\cot x) = -\csc^2 x$$

$$\frac{d}{dx}(\sec x) = \sec x \tan x \qquad \frac{d}{dx}(\csc x) = -\csc x \cot x$$

To show a typical calculation, we find the derivative of the tangent function. The other derivations are left to Exercise 60.

EXAMPLE 5 Find $d(\tan x)/dx$.

Solution We use the Derivative Quotient Rule to calculate the derivative:

$$\begin{aligned}\frac{d}{dx}(\tan x) = \frac{d}{dx}\left(\frac{\sin x}{\cos x}\right) &= \frac{\cos x \frac{d}{dx}(\sin x) - \sin x \frac{d}{dx}(\cos x)}{\cos^2 x} && \text{Quotient Rule}\\ &= \frac{\cos x \cos x - \sin x(-\sin x)}{\cos^2 x}\\ &= \frac{\cos^2 x + \sin^2 x}{\cos^2 x}\\ &= \frac{1}{\cos^2 x} = \sec^2 x.\end{aligned}$$

■

EXAMPLE 6 Find y'' if $y = \sec x$.

Solution Finding the second derivative involves a combination of trigonometric derivatives.

$$
\begin{aligned}
y &= \sec x \\
y' &= \sec x \tan x && \text{Derivative rule for secant function} \\
y'' &= \frac{d}{dx}(\sec x \tan x) \\
&= \sec x \frac{d}{dx}(\tan x) + \tan x \frac{d}{dx}(\sec x) && \text{Derivative Product Rule} \\
&= \sec x(\sec^2 x) + \tan x(\sec x \tan x) && \text{Derivative rules} \\
&= \sec^3 x + \sec x \tan^2 x
\end{aligned}
$$

The differentiability of the trigonometric functions throughout their domains gives another proof of their continuity at every point in their domains (Theorem 1, Section 3.2). So we can calculate limits of algebraic combinations and composites of trigonometric functions by direct substitution.

EXAMPLE 7 We can use direct substitution in computing limits provided there is no division by zero, which is algebraically undefined.

$$\lim_{x \to 0} \frac{\sqrt{2 + \sec x}}{\cos(\pi - \tan x)} = \frac{\sqrt{2 + \sec 0}}{\cos(\pi - \tan 0)} = \frac{\sqrt{2 + 1}}{\cos(\pi - 0)} = \frac{\sqrt{3}}{-1} = -\sqrt{3}$$

Exercises 3.5

Derivatives

In Exercises 1–18, find dy/dx.

1. $y = -10x + 3\cos x$
2. $y = \frac{3}{x} + 5\sin x$
3. $y = x^2 \cos x$
4. $y = \sqrt{x}\sec x + 3$
5. $y = \csc x - 4\sqrt{x} + \frac{7}{e^x}$
6. $y = x^2 \cot x - \frac{1}{x^2}$
7. $f(x) = \sin x \tan x$
8. $g(x) = \frac{\cos x}{\sin^2 x}$
9. $y = xe^{-x}\sec x$
10. $y = (\sin x + \cos x)\sec x$
11. $y = \frac{\cot x}{1 + \cot x}$
12. $y = \frac{\cos x}{1 + \sin x}$
13. $y = \frac{4}{\cos x} + \frac{1}{\tan x}$
14. $y = \frac{\cos x}{x} + \frac{x}{\cos x}$
15. $y = (\sec x + \tan x)(\sec x - \tan x)$
16. $y = x^2 \cos x - 2x \sin x - 2\cos x$
17. $f(x) = x^3 \sin x \cos x$
18. $g(x) = (2 - x)\tan^2 x$

In Exercises 19–22, find ds/dt.

19. $s = \tan t - e^{-t}$
20. $s = t^2 - \sec t + 5e^t$
21. $s = \frac{1 + \csc t}{1 - \csc t}$
22. $s = \frac{\sin t}{1 - \cos t}$

In Exercises 23–26, find $dr/d\theta$.

23. $r = 4 - \theta^2 \sin \theta$
24. $r = \theta \sin \theta + \cos \theta$
25. $r = \sec \theta \csc \theta$
26. $r = (1 + \sec \theta)\sin \theta$

In Exercises 27–32, find dp/dq.

27. $p = 5 + \frac{1}{\cot q}$
28. $p = (1 + \csc q)\cos q$
29. $p = \frac{\sin q + \cos q}{\cos q}$
30. $p = \frac{\tan q}{1 + \tan q}$
31. $p = \frac{q \sin q}{q^2 - 1}$
32. $p = \frac{3q + \tan q}{q \sec q}$
33. Find y'' if
    a. $y = \csc x$.
    b. $y = \sec x$.
34. Find $y^{(4)} = d^4 y/dx^4$ if
    a. $y = -2\sin x$.
    b. $y = 9\cos x$.

Tangent Lines

In Exercises 35–38, graph the curves over the given intervals, together with their tangents at the given values of x. Label each curve and tangent with its equation.

35. $y = \sin x, \quad -3\pi/2 \le x \le 2\pi$
    $x = -\pi, 0, 3\pi/2$

36. $y = \tan x, \quad -\pi/2 < x < \pi/2$
$x = -\pi/3, 0, \pi/3$

37. $y = \sec x, \quad -\pi/2 < x < \pi/2$
$x = -\pi/3, \pi/4$

38. $y = 1 + \cos x, \quad -3\pi/2 \le x \le 2\pi$
$x = -\pi/3, 3\pi/2$

T Do the graphs of the functions in Exercises 39–42 have any horizontal tangents in the interval $0 \le x \le 2\pi$? If so, where? If not, why not? Visualize your findings by graphing the functions with a grapher.

39. $y = x + \sin x$

40. $y = 2x + \sin x$

41. $y = x - \cot x$

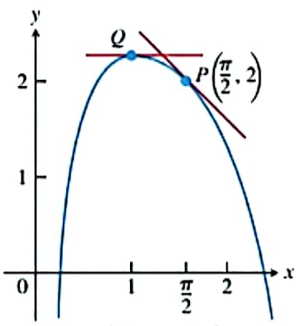
42. $y = x + 2\cos x$

43. Find all points on the curve $y = \tan x$, $-\pi/2 < x < \pi/2$, where the tangent line is parallel to the line $y = 2x$. Sketch the curve and tangent(s) together, labeling each with its equation.

44. Find all points on the curve $y = \cot x$, $0 < x < \pi$, where the tangent line is parallel to the line $y = -x$. Sketch the curve and tangent(s) together, labeling each with its equation.

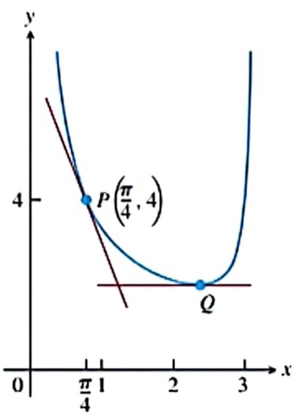
In Exercises 45 and 46, find an equation for **(a)** the tangent to the curve at P and **(b)** the horizontal tangent to the curve at Q.

45.



$y = 4 + \cot x - 2\csc x$

46.



$y = 1 + \sqrt{2}\csc x + \cot x$

Trigonometric Limits

Find the limits in Exercises 47–54.

47. $\lim_{x\to 2} \sin\left(\frac{1}{x} - \frac{1}{2}\right)$

48. $\lim_{x\to -\pi/6} \sqrt{1 + \cos(\pi \csc x)}$

49. $\lim_{\theta\to\pi/6} \frac{\sin\theta - \frac{1}{2}}{\theta - \frac{\pi}{6}}$

50. $\lim_{\theta\to\pi/4} \frac{\tan\theta - 1}{\theta - \frac{\pi}{4}}$

51. $\lim_{x\to 0} \sec\left[e^x + \pi\tan\left(\frac{\pi}{4\sec x}\right) - 1\right]$

52. $\lim_{x\to 0} \sin\left(\frac{\pi + \tan x}{\tan x - 2\sec x}\right)$

53. $\lim_{t\to 0} \tan\left(1 - \frac{\sin t}{t}\right)$

54. $\lim_{\theta\to 0} \cos\left(\frac{\pi\theta}{\sin\theta}\right)$

Theory and Examples

The equations in Exercises 55 and 56 give the position $s = f(t)$ of a body moving on a coordinate line (s in meters, t in seconds). Find the body's velocity, speed, acceleration, and jerk at time $t = \pi/4$ sec.

55. $s = 2 - 2\sin t$

56. $s = \sin t + \cos t$

57. Is there a value of c that will make

$$f(x) = \begin{cases} \frac{\sin^2 3x}{x^2}, & x \ne 0 \\ c, & x = 0 \end{cases}$$

continuous at $x = 0$? Give reasons for your answer.

58. Is there a value of b that will make

$$g(x) = \begin{cases} x + b, & x < 0 \\ \cos x, & x \ge 0 \end{cases}$$

continuous at $x = 0$? Differentiable at $x = 0$? Give reasons for your answers.

59. By computing the first few derivatives and looking for a pattern, find $d^{999}/dx^{999}(\cos x)$.

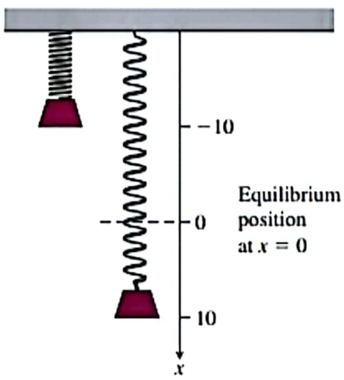
60. Derive the formula for the derivative with respect to x of

a. $\sec x$. **b.** $\csc x$. **c.** $\cot x$.

61. A weight is attached to a spring and reaches its equilibrium position ($x = 0$). It is then set in motion resulting in a displacement of

$$x = 10\cos t,$$

where x is measured in centimeters and t is measured in seconds. See the accompanying figure.



a. Find the spring's displacement when $t = 0$, $t = \pi/3$, and $t = 3\pi/4$.

b. Find the spring's velocity when $t = 0$, $t = \pi/3$, and $t = 3\pi/4$.

62. Assume that a particle's position on the x-axis is given by

$$x = 3\cos t + 4\sin t,$$

where x is measured in feet and t is measured in seconds.

a. Find the particle's position when $t = 0$, $t = \pi/2$, and $t = \pi$.

b. Find the particle's velocity when $t = 0$, $t = \pi/2$, and $t = \pi$.

T **63.** Graph $y = \cos x$ for $-\pi \le x \le 2\pi$. On the same screen, graph

$$y = \frac{\sin(x + h) - \sin x}{h}$$

for $h = 1, 0.5, 0.3,$ and 0.1. Then, in a new window, try $h = -1, -0.5,$ and -0.3. What happens as $h \to 0^+$? As $h \to 0^-$? What phenomenon is being illustrated here?

T **64.** Graph $y = -\sin x$ for $-\pi \le x \le 2\pi$. On the same screen, graph

$$y = \frac{\cos(x + h) - \cos x}{h}$$

for $h = 1, 0.5, 0.3,$ and 0.1. Then, in a new window, try $h = -1, -0.5,$ and -0.3. What happens as $h \to 0^+$? As $h \to 0^-$? What phenomenon is being illustrated here?

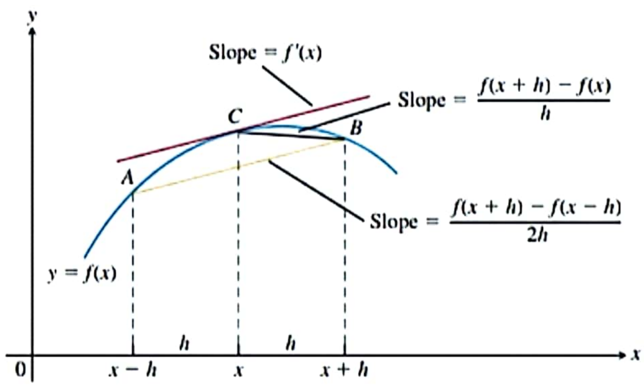
T **65. Centered difference quotients** The *centered difference quotient*

$$\frac{f(x + h) - f(x - h)}{2h}$$

is used to approximate $f'(x)$ in numerical work because (1) its limit as $h \to 0$ equals $f'(x)$ when $f'(x)$ exists, and (2) it usually gives a better approximation of $f'(x)$ for a given value of h than the difference quotient

$$\frac{f(x + h) - f(x)}{h}.$$

See the accompanying figure.



a. To see how rapidly the centered difference quotient for $f(x) = \sin x$ converges to $f'(x) = \cos x$, graph $y = \cos x$ together with

$$y = \frac{\sin(x + h) - \sin(x - h)}{2h}$$

over the interval $[-\pi, 2\pi]$ for $h = 1, 0.5,$ and 0.3. Compare the results with those obtained in Exercise 63 for the same values of h.

b. To see how rapidly the centered difference quotient for $f(x) = \cos x$ converges to $f'(x) = -\sin x$, graph $y = -\sin x$ together with

$$y = \frac{\cos(x + h) - \cos(x - h)}{2h}$$

over the interval $[-\pi, 2\pi]$ for $h = 1, 0.5,$ and 0.3. Compare the results with those obtained in Exercise 64 for the same values of h.

66. A caution about centered difference quotients (*Continuation of Exercise 65.*) The quotient

$$\frac{f(x + h) - f(x - h)}{2h}$$

may have a limit as $h \to 0$ when f has no derivative at x. As a case in point, take $f(x) = |x|$ and calculate

$$\lim_{h \to 0} \frac{|0 + h| - |0 - h|}{2h}.$$

As you will see, the limit exists even though $f(x) = |x|$ has no derivative at $x = 0$. *Moral:* Before using a centered difference quotient, be sure the derivative exists.

T **67. Slopes on the graph of the tangent function** Graph $y = \tan x$ and its derivative together on $(-\pi/2, \pi/2)$. Does the graph of the tangent function appear to have a smallest slope? A largest slope? Is the slope ever negative? Give reasons for your answers.

T **68. Slopes on the graph of the cotangent function** Graph $y = \cot x$ and its derivative together for $0 < x < \pi$. Does the graph of the cotangent function appear to have a smallest slope? A largest slope? Is the slope ever positive? Give reasons for your answers.

T **69. Exploring $(\sin kx)/x$** Graph $y = (\sin x)/x$, $y = (\sin 2x)/x$, and $y = (\sin 4x)/x$ together over the interval $-2 \le x \le 2$. Where does each graph appear to cross the y-axis? Do the graphs really intersect the axis? What would you expect the graphs of $y = (\sin 5x)/x$ and $y = (\sin(-3x))/x$ to do as $x \to 0$? Why? What about the graph of $y = (\sin kx)/x$ for other values of k? Give reasons for your answers.

T **70. Radians versus degrees: degree mode derivatives** What happens to the derivatives of $\sin x$ and $\cos x$ if x is measured in degrees instead of radians? To find out, take the following steps.

a. With your graphing calculator or computer grapher in *degree mode*, graph

$$f(h) = \frac{\sin h}{h}$$

and estimate $\lim_{h \to 0} f(h)$. Compare your estimate with $\pi/180$. Is there any reason to believe the limit *should* be $\pi/180$?

b. With your grapher still in degree mode, estimate

$$\lim_{h \to 0} \frac{\cos h - 1}{h}.$$

c. Now go back to the derivation of the formula for the derivative of $\sin x$ in the text and carry out the steps of the derivation using degree-mode limits. What formula do you obtain for the derivative?

d. Work through the derivation of the formula for the derivative of $\cos x$ using degree-mode limits. What formula do you obtain for the derivative?

e. The disadvantages of the degree-mode formulas become apparent as you start taking derivatives of higher order. Try it. What are the second and third degree-mode derivatives of $\sin x$ and $\cos x$?

3.6 The Chain Rule

How do we differentiate $F(x) = \sin(x^2 - 4)$? This function is the composite $f \circ g$ of two functions $y = f(u) = \sin u$ and $u = g(x) = x^2 - 4$ that we know how to differentiate. The answer, given by the *Chain Rule*, says that the derivative is the product of the derivatives of f and g. We develop the rule in this section.

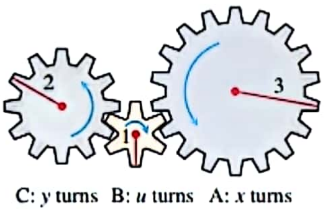


FIGURE 3.25 When gear A makes x turns, gear B makes u turns and gear C makes y turns. By comparing circumferences or counting teeth, we see that $y = u/2$ (C turns one-half turn for each B turn) and $u = 3x$ (B turns three times for A's one), so $y = 3x/2$. Thus, $dy/dx = 3/2 = (1/2)(3) = (dy/du)(du/dx)$.

Derivative of a Composite Function

The function $y = \frac{3}{2}x = \frac{1}{2}(3x)$ is the composite of the functions $y = \frac{1}{2}u$ and $u = 3x$. We have

$$\frac{dy}{dx} = \frac{3}{2}, \qquad \frac{dy}{du} = \frac{1}{2}, \qquad \text{and} \qquad \frac{du}{dx} = 3.$$

Since $\frac{3}{2} = \frac{1}{2} \cdot 3$, we see in this case that

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

If we think of the derivative as a rate of change, our intuition allows us to see that this relationship is reasonable. If $y = f(u)$ changes half as fast as u and $u = g(x)$ changes three times as fast as x, then we expect y to change $3/2$ times as fast as x. This effect is much like that of a multiple gear train (Figure 3.25). Let's look at another example.

EXAMPLE 1 The function

$$y = (3x^2 + 1)^2$$

is the composite of $y = f(u) = u^2$ and $u = g(x) = 3x^2 + 1$. Calculating derivatives, we see that

$$\begin{aligned} \frac{dy}{du} \cdot \frac{du}{dx} &= 2u \cdot 6x \\ &= 2(3x^2 + 1) \cdot 6x && \text{Substitute for } u \\ &= 36x^3 + 12x. \end{aligned}$$

Calculating the derivative from the expanded formula $(3x^2 + 1)^2 = 9x^4 + 6x^2 + 1$ gives the same result:

$$\begin{aligned} \frac{dy}{dx} &= \frac{d}{dx}(9x^4 + 6x^2 + 1) \\ &= 36x^3 + 12x. \end{aligned}$$

The derivative of the composite function $f(g(x))$ at x is the derivative of f at $g(x)$ times the derivative of g at x. This is known as the Chain Rule (Figure 3.26).

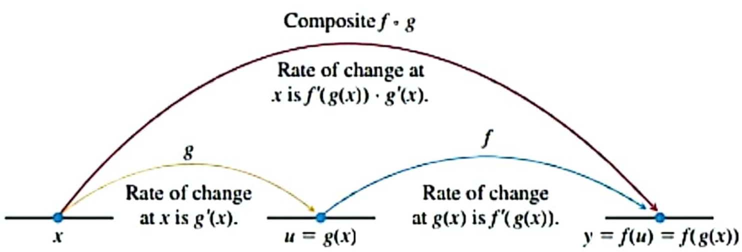


FIGURE 3.26 Rates of change multiply: The derivative of $f \circ g$ at x is the derivative of f at $g(x)$ times the derivative of g at x.

THEOREM 2—The Chain Rule If $f(u)$ is differentiable at the point $u = g(x)$ and $g(x)$ is differentiable at x, then the composite function $(f \circ g)(x) = f(g(x))$ is differentiable at x, and

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

In Leibniz's notation, if $y = f(u)$ and $u = g(x)$, then

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx},$$

where dy/du is evaluated at $u = g(x)$.

A Proof of One Case of the Chain Rule:

Let Δu be the change in u when x changes by Δx, so that

$$\Delta u = g(x + \Delta x) - g(x).$$

Then the corresponding change in y is

$$\Delta y = f(u + \Delta u) - f(u).$$

If $\Delta u \neq 0$, we can write the fraction $\Delta y / \Delta x$ as the product

$$\frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta u} \cdot \frac{\Delta u}{\Delta x} \tag{1}$$

and take the limit as $\Delta x \to 0$:

$$\begin{aligned}\frac{dy}{dx} &= \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta u} \cdot \frac{\Delta u}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta u} \cdot \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} \\ &= \lim_{\Delta u \to 0} \frac{\Delta y}{\Delta u} \cdot \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} && \text{(Note that } \Delta u \to 0 \text{ as } \Delta x \to 0 \text{ since } g \text{ is continuous.)} \\ &= \frac{dy}{du} \cdot \frac{du}{dx}.\end{aligned}$$

The problem with this argument is that if the function $g(x)$ oscillates rapidly near x, then Δu can be zero even when $\Delta x \neq 0$, so the cancelation of Δu in Equation (1) would be invalid. A complete proof requires a different approach that avoids this problem, and we give one such proof in Section 3.11. ■

EXAMPLE 2 An object moves along the x-axis so that its position at any time $t \geq 0$ is given by $x(t) = \cos(t^2 + 1)$. Find the velocity of the object as a function of t.

Solution We know that the velocity is dx/dt. In this instance, x is a composite function: $x = \cos(u)$ and $u = t^2 + 1$. We have

$$\frac{dx}{du} = -\sin(u) \qquad x = \cos(u)$$

$$\frac{du}{dt} = 2t. \qquad u = t^2 + 1$$

By the Chain Rule,

$$\begin{aligned}\frac{dx}{dt} &= \frac{dx}{du}\cdot\frac{du}{dt}\\ &= -\sin(u)\cdot 2t \qquad \frac{dx}{du}\text{ evaluated at } u\\ &= -\sin(t^2+1)\cdot 2t\\ &= -2t\sin(t^2+1).\end{aligned}$$

Ways to Write the Chain Rule

$(f\circ g)'(x) = f'(g(x))\cdot g'(x)$

$\frac{dy}{dx} = \frac{dy}{du}\cdot\frac{du}{dx}$

$\frac{dy}{dx} = f'(g(x))\cdot g'(x)$

$\frac{d}{dx}f(u) = f'(u)\frac{du}{dx}$

"Outside-Inside" Rule

A difficulty with the Leibniz notation is that it doesn't state specifically where the derivatives in the Chain Rule are supposed to be evaluated. So it sometimes helps to think about the Chain Rule using functional notation. If $y = f(g(x))$, then

$$\frac{dy}{dx} = f'(g(x))\cdot g'(x).$$

In words, differentiate the "outside" function f and evaluate it at the "inside" function $g(x)$ left alone; then multiply by the derivative of the "inside function."

EXAMPLE 3 Differentiate $\sin(x^2 + e^x)$ with respect to x.

Solution We apply the Chain Rule directly and find

$$\frac{d}{dx}\sin\underbrace{(x^2+e^x)}_{\text{inside}} = \cos\underbrace{(x^2+e^x)}_{\text{inside left alone}}\cdot\underbrace{(2x+e^x)}_{\text{derivative of the inside}}.$$

EXAMPLE 4 Differentiate $y = e^{\cos x}$.

Solution Here the inside function is $u = g(x) = \cos x$ and the outside function is the exponential function $f(x) = e^x$. Applying the Chain Rule, we get

$$\frac{dy}{dx} = \frac{d}{dx}(e^{\cos x}) = e^{\cos x}\frac{d}{dx}(\cos x) = e^{\cos x}(-\sin x) = -e^{\cos x}\sin x.$$

Generalizing Example 4, we see that the Chain Rule gives the formula

$$\boxed{\frac{d}{dx}e^u = e^u\frac{du}{dx}.}$$

For example,

$$\frac{d}{dx}(e^{kx}) = e^{kx}\cdot\frac{d}{dx}(kx) = ke^{kx}, \qquad \text{for any constant } k$$

and

$$\frac{d}{dx}\left(e^{x^2}\right) = e^{x^2}\cdot\frac{d}{dx}(x^2) = 2xe^{x^2}.$$

Repeated Use of the Chain Rule

We sometimes have to use the Chain Rule two or more times to find a derivative.

HISTORICAL BIOGRAPHY
Johann Bernoulli
(1667–1748)

EXAMPLE 5 Find the derivative of $g(t) = \tan(5 - \sin 2t)$.

Solution Notice here that the tangent is a function of $5 - \sin 2t$, whereas the sine is a function of $2t$, which is itself a function of t. Therefore, by the Chain Rule,

$$\begin{aligned}
g'(t) &= \frac{d}{dt}(\tan(5 - \sin 2t)) \\
&= \sec^2(5 - \sin 2t)\cdot\frac{d}{dt}(5 - \sin 2t) && \text{Derivative of } \tan u \text{ with } u = 5 - \sin 2t \\
&= \sec^2(5 - \sin 2t)\cdot\left(0 - \cos 2t\cdot\frac{d}{dt}(2t)\right) && \text{Derivative of } 5 - \sin u \text{ with } u = 2t \\
&= \sec^2(5 - \sin 2t)\cdot(-\cos 2t)\cdot 2 \\
&= -2(\cos 2t)\sec^2(5 - \sin 2t).
\end{aligned}$$

The Chain Rule with Powers of a Function

If f is a differentiable function of u and if u is a differentiable function of x, then substituting $y = f(u)$ into the Chain Rule formula

$$\frac{dy}{dx} = \frac{dy}{du}\cdot\frac{du}{dx}$$

leads to the formula

$$\frac{d}{dx}f(u) = f'(u)\frac{du}{dx}.$$

If n is any real number and f is a power function, $f(u) = u^n$, the Power Rule tells us that $f'(u) = nu^{n-1}$. If u is a differentiable function of x, then we can use the Chain Rule to extend this to the **Power Chain Rule**:

$$\frac{d}{dx}(u^n) = nu^{n-1}\frac{du}{dx}. \qquad \frac{d}{du}(u^n) = nu^{n-1}$$

EXAMPLE 6 The Power Chain Rule simplifies computing the derivative of a power of an expression.

(a)
$$\begin{aligned}
\frac{d}{dx}(5x^3 - x^4)^7 &= 7(5x^3 - x^4)^6\frac{d}{dx}(5x^3 - x^4) && \text{Power Chain Rule with } u = 5x^3 - x^4,\ n = 7 \\
&= 7(5x^3 - x^4)^6(5\cdot 3x^2 - 4x^3) \\
&= 7(5x^3 - x^4)^6(15x^2 - 4x^3)
\end{aligned}$$

(b)
$$\begin{aligned}
\frac{d}{dx}\left(\frac{1}{3x - 2}\right) &= \frac{d}{dx}(3x - 2)^{-1} \\
&= -1(3x - 2)^{-2}\frac{d}{dx}(3x - 2) && \text{Power Chain Rule with } u = 3x - 2,\ n = -1 \\
&= -1(3x - 2)^{-2}(3) \\
&= -\frac{3}{(3x - 2)^2}
\end{aligned}$$

In part (b) we could also find the derivative with the Derivative Quotient Rule.

(c)
$$\begin{aligned}
\frac{d}{dx}(\sin^5 x) &= 5\sin^4 x\cdot\frac{d}{dx}\sin x && \text{Power Chain Rule with } u = \sin x,\ n = 5, \text{ because } \sin^n x \text{ means } (\sin x)^n,\ n \neq -1. \\
&= 5\sin^4 x\cos x
\end{aligned}$$

$$\text{(d)}\quad \frac{d}{dx}\left(e^{\sqrt{3x+1}}\right) = e^{\sqrt{3x+1}} \cdot \frac{d}{dx}\left(\sqrt{3x+1}\right)$$

$$= e^{\sqrt{3x+1}} \cdot \frac{1}{2}(3x+1)^{-1/2} \cdot 3 \qquad \text{Power Chain Rule with } u = 3x+1,\ n = 1/2$$

$$= \frac{3}{2\sqrt{3x+1}} e^{\sqrt{3x+1}}$$

EXAMPLE 7 In Section 3.2, we saw that the absolute value function $y = |x|$ is not differentiable at $x = 0$. However, the function is differentiable at all other real numbers, as we now show. Since $|x| = \sqrt{x^2}$, we can derive the following formula:

Derivative of the Absolute Value Function

$$\frac{d}{dx}(|x|) = \frac{x}{|x|},\quad x \neq 0$$
$$= \begin{cases} 1, & x > 0 \\ -1, & x < 0 \end{cases}$$

$$\frac{d}{dx}(|x|) = \frac{d}{dx}\sqrt{x^2}$$

$$= \frac{1}{2\sqrt{x^2}} \cdot \frac{d}{dx}(x^2) \qquad \text{Power Chain Rule with } u = x^2,\ n = 1/2,\ x \neq 0$$

$$= \frac{1}{2|x|} \cdot 2x \qquad \sqrt{x^2} = |x|$$

$$= \frac{x}{|x|},\quad x \neq 0.$$

EXAMPLE 8 Show that the slope of every line tangent to the curve $y = 1/(1-2x)^3$ is positive.

Solution We find the derivative:

$$\frac{dy}{dx} = \frac{d}{dx}(1-2x)^{-3}$$

$$= -3(1-2x)^{-4} \cdot \frac{d}{dx}(1-2x) \qquad \text{Power Chain Rule with } u = (1-2x),\ n = -3$$

$$= -3(1-2x)^{-4} \cdot (-2)$$

$$= \frac{6}{(1-2x)^4}.$$

At any point (x, y) on the curve, the coordinate x is not $1/2$ and the slope of the tangent line is

$$\frac{dy}{dx} = \frac{6}{(1-2x)^4},$$

which is the quotient of two positive numbers.

EXAMPLE 9 The formulas for the derivatives of both $\sin x$ and $\cos x$ were obtained under the assumption that x is measured in radians, *not* degrees. The Chain Rule gives us new insight into the difference between the two. Since $180° = \pi$ radians, $x° = \pi x/180$ radians where $x°$ is the size of the angle measured in degrees.

By the Chain Rule,

$$\frac{d}{dx}\sin(x°) = \frac{d}{dx}\sin\left(\frac{\pi x}{180}\right) = \frac{\pi}{180}\cos\left(\frac{\pi x}{180}\right) = \frac{\pi}{180}\cos(x°).$$

See Figure 3.27. Similarly, the derivative of $\cos(x°)$ is $-(\pi/180)\sin(x°)$.

The factor $\pi/180$ would compound with repeated differentiation, showing an advantage for the use of radian measure in computations.

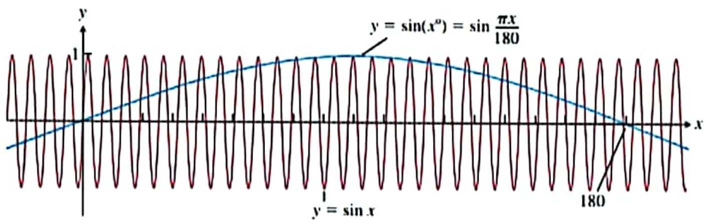


FIGURE 3.27 The function $\sin(x°)$ oscillates only $\pi/180$ times as often as $\sin x$ oscillates. Its maximum slope is $\pi/180$ at $x = 0$ (Example 9).

Related Rates Equations

Suppose we are pumping air into a spherical balloon. Both the volume and radius of the balloon are increasing over time. If V is the volume and r is the radius of the balloon at an instant of time, then

$$V = \frac{4}{3}\pi r^3.$$

Using the Chain Rule, we differentiate both sides with respect to t to find an equation relating the rates of change of V and r,

$$\frac{dV}{dt} = \frac{dV}{dr}\frac{dr}{dt} = 4\pi r^2 \frac{dr}{dt}.$$

So if we know the radius r of the balloon and the rate dV/dt at which the volume is increasing at a given instant of time, then we can solve this last equation for dr/dt to find how fast the radius is increasing at that instant. Note that it is easier to directly measure the rate of increase of the volume (the rate at which air is being pumped into the balloon) than it is to measure the increase in the radius. The related rates equation allows us to calculate dr/dt from dV/dt.

Very often the key to relating the variables in a related rates problem is drawing a picture that shows the geometric relations between them, as illustrated in the following example.

EXAMPLE 1 Water runs into a conical tank at the rate of 9 ft^3/min. The tank stands point down and has a height of 10 ft and a base radius of 5 ft. How fast is the water level rising when the water is 6 ft deep?

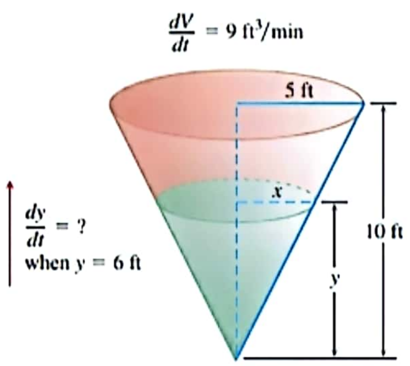


FIGURE 3.44 The geometry of the conical tank and the rate at which water fills the tank determine how fast the water level rises (Example 1).

Solution Figure 3.44 shows a partially filled conical tank. The variables in the problem are

V = volume (ft^3) of the water in the tank at time t (min)
x = radius (ft) of the surface of the water at time t
y = depth (ft) of the water in the tank at time t.

We assume that V, x, and y are differentiable functions of t. The constants are the dimensions of the tank. We are asked for dy/dt when

$$y = 6 \text{ ft} \quad \text{and} \quad \frac{dV}{dt} = 9 \text{ ft}^3/\text{min}.$$

The water forms a cone with volume

$$V = \frac{1}{3}\pi x^2 y.$$

This equation involves x as well as V and y. Because no information is given about x and dx/dt at the time in question, we need to eliminate x. The similar triangles in Figure 3.44 give us a way to express x in terms of y:

$$\frac{x}{y} = \frac{5}{10} \quad \text{or} \quad x = \frac{y}{2}.$$

Therefore, we find

$$V = \frac{1}{3}\pi\left(\frac{y}{2}\right)^2 y = \frac{\pi}{12}y^3$$

to give the derivative

$$\frac{dV}{dt} = \frac{\pi}{12}\cdot 3y^2\frac{dy}{dt} = \frac{\pi}{4}y^2\frac{dy}{dt}.$$

Finally, use $y = 6$ and $dV/dt = 9$ to solve for dy/dt.

$$9 = \frac{\pi}{4}(6)^2 \frac{dy}{dt}$$

$$\frac{dy}{dt} = \frac{1}{\pi} \approx 0.32$$

At the moment in question, the water level is rising at about 0.32 ft/min. ■

Related Rates Problem Strategy

1. *Draw a picture and name the variables and constants.* Use t for time. Assume that all variables are differentiable functions of t.
2. *Write down the numerical information* (in terms of the symbols you have chosen).
3. *Write down what you are asked to find* (usually a rate, expressed as a derivative).
4. *Write an equation that relates the variables.* You may have to combine two or more equations to get a single equation that relates the variable whose rate you want to the variables whose rates you know.
5. *Differentiate with respect to t.* Then express the rate you want in terms of the rates and variables whose values you know.
6. *Evaluate.* Use known values to find the unknown rate.

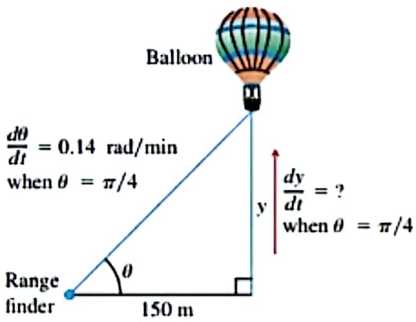


FIGURE 3.45 The rate of change of the balloon's height is related to the rate of change of the angle the range finder makes with the ground (Example 2).

EXAMPLE 2 A hot air balloon rising straight up from a level field is tracked by a range finder 150 m from the liftoff point. At the moment the range finder's elevation angle is $\pi/4$, the angle is increasing at the rate of 0.14 rad/min. How fast is the balloon rising at that moment?

Solution We answer the question in the six strategy steps.

1. *Draw a picture and name the variables and constants* (Figure 3.45). The variables in the picture are

   θ = the angle in radians the range finder makes with the ground.

   y = the height in meters of the balloon above the ground.

We let t represent time in minutes and assume that θ and y are differentiable functions of t.

The one constant in the picture is the distance from the range finder to the liftoff point (150 m). There is no need to give it a special symbol.

2. *Write down the additional numerical information.*

$$\frac{d\theta}{dt} = 0.14 \text{ rad/min} \quad \text{when} \quad \theta = \frac{\pi}{4}$$

3. *Write down what we are to find.* We want dy/dt when $\theta = \pi/4$.
4. *Write an equation that relates the variables y and θ.*

$$\frac{y}{150} = \tan\theta \quad \text{or} \quad y = 150 \tan\theta$$

5. *Differentiate with respect to t using the Chain Rule.* The result tells how dy/dt (which we want) is related to $d\theta/dt$ (which we know).

$$\frac{dy}{dt} = 150\,(\sec^2\theta)\frac{d\theta}{dt}$$

6. *Evaluate with $\theta = \pi/4$ and $d\theta/dt = 0.14$ to find dy/dt.*

$$\frac{dy}{dt} = 150(\sqrt{2})^2(0.14) = 42 \qquad \sec\frac{\pi}{4} = \sqrt{2}$$

At the moment in question, the balloon is rising at the rate of 42 m/min. ■

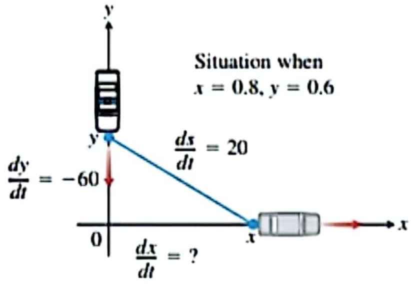


FIGURE 3.46 The speed of the car is related to the speed of the police cruiser and the rate of change of the distance s between them (Example 3).

EXAMPLE 3 A police cruiser, approaching a right-angled intersection from the north, is chasing a speeding car that has turned the corner and is now moving straight east. When the cruiser is 0.6 mi north of the intersection and the car is 0.8 mi to the east, the police determine with radar that the distance between them and the car is increasing at 20 mph. If the cruiser is moving at 60 mph at the instant of measurement, what is the speed of the car?

Solution We picture the car and cruiser in the coordinate plane, using the positive x-axis as the eastbound highway and the positive y-axis as the southbound highway (Figure 3.46). We let t represent time and set

$$x = \text{position of car at time } t$$
$$y = \text{position of cruiser at time } t$$
$$s = \text{distance between car and cruiser at time } t.$$

We assume that x, y, and s are differentiable functions of t.

We want to find dx/dt when

$$x = 0.8 \text{ mi}, \quad y = 0.6 \text{ mi}, \quad \frac{dy}{dt} = -60 \text{ mph}, \quad \frac{ds}{dt} = 20 \text{ mph}.$$

Note that dy/dt is negative because y is decreasing.

We differentiate the distance equation between the car and the cruiser,

$$s^2 = x^2 + y^2$$

(we could also use $s = \sqrt{x^2 + y^2}$), and obtain

$$2s\frac{ds}{dt} = 2x\frac{dx}{dt} + 2y\frac{dy}{dt}$$
$$\frac{ds}{dt} = \frac{1}{s}\left(x\frac{dx}{dt} + y\frac{dy}{dt}\right)$$
$$= \frac{1}{\sqrt{x^2 + y^2}}\left(x\frac{dx}{dt} + y\frac{dy}{dt}\right).$$

Finally, we use $x = 0.8$, $y = 0.6$, $dy/dt = -60$, $ds/dt = 20$, and solve for dx/dt.

$$20 = \frac{1}{\sqrt{(0.8)^2 + (0.6)^2}}\left(0.8\frac{dx}{dt} + (0.6)(-60)\right)$$
$$\frac{dx}{dt} = \frac{20\sqrt{(0.8)^2 + (0.6)^2} + (0.6)(60)}{0.8} = 70$$

At the moment in question, the car's speed is 70 mph.

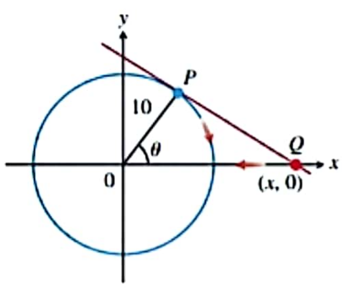


FIGURE 3.47 The particle P travels clockwise along the circle (Example 4).

EXAMPLE 4 A particle P moves clockwise at a constant rate along a circle of radius 10 m centered at the origin. The particle's initial position is (0, 10) on the y-axis, and its final destination is the point (10, 0) on the x-axis. Once the particle is in motion, the tangent line at P intersects the x-axis at a point Q (which moves over time). If it takes the particle 30 sec to travel from start to finish, how fast is the point Q moving along the x-axis when it is 20 m from the center of the circle?

Solution We picture the situation in the coordinate plane with the circle centered at the origin (see Figure 3.47). We let t represent time and let θ denote the angle from the x-axis to the radial line joining the origin to P. Since the particle travels from start to finish in 30 sec, it is traveling along the circle at a constant rate of $\pi/2$ radians in $1/2$ min, or π rad/min. In other words, $d\theta/dt = -\pi$, with t being measured in minutes. The negative sign appears because θ is decreasing over time.

Setting $x(t)$ to be the distance at time t from the point Q to the origin, we want to find dx/dt when

$$x = 20 \text{ m} \qquad \text{and} \qquad \frac{d\theta}{dt} = -\pi \text{ rad/min.}$$

To relate the variables x and θ, we see from Figure 3.47 that $x \cos\theta = 10$, or $x = 10 \sec\theta$. Differentiation of this last equation gives

$$\frac{dx}{dt} = 10 \sec\theta \tan\theta \frac{d\theta}{dt} = -10\pi \sec\theta \tan\theta.$$

Note that dx/dt is negative because x is decreasing (Q is moving toward the origin).

When $x = 20$, $\cos\theta = 1/2$ and $\sec\theta = 2$. Also, $\tan\theta = \sqrt{\sec^2\theta - 1} = \sqrt{3}$. It follows that

$$\frac{dx}{dt} = (-10\pi)(2)\left(\sqrt{3}\right) = -20\sqrt{3}\pi.$$

At the moment in question, the point Q is moving toward the origin at the speed of $20\sqrt{3}\pi \approx 109$ m/min. ■

EXAMPLE 5 A jet airliner is flying at a constant altitude of 12,000 ft above sea level as it approaches a Pacific island. The aircraft comes within the direct line of sight of a radar station located on the island, and the radar indicates the initial angle between sea level and its line of sight to the aircraft is 30°. How fast (in miles per hour) is the aircraft approaching the island when first detected by the radar instrument if it is turning upward (counterclockwise) at the rate of 2/3 deg/sec in order to keep the aircraft within its direct line of sight?

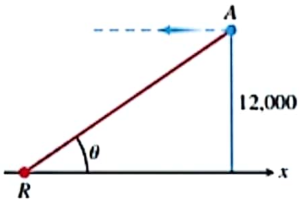


FIGURE 3.48 Jet airliner A traveling at constant altitude toward radar station R (Example 5).

Solution The aircraft A and radar station R are pictured in the coordinate plane, using the positive x-axis as the horizontal distance at sea level from R to A, and the positive y-axis as the vertical altitude above sea level. We let t represent time and observe that $y = 12{,}000$ is a constant. The general situation and line-of-sight angle θ are depicted in Figure 3.48. We want to find dx/dt when $\theta = \pi/6$ rad and $d\theta/dt = 2/3$ deg/sec.

From Figure 3.48, we see that

$$\frac{12{,}000}{x} = \tan\theta \qquad \text{or} \qquad x = 12{,}000 \cot\theta.$$

Using miles instead of feet for our distance units, the last equation translates to

$$x = \frac{12{,}000}{5280} \cot\theta.$$

Differentiation with respect to t gives

$$\frac{dx}{dt} = -\frac{1200}{528} \csc^2\theta \frac{d\theta}{dt}.$$

When $\theta = \pi/6$, $\sin^2\theta = 1/4$, so $\csc^2\theta = 4$. Converting $d\theta/dt = 2/3$ deg/sec to radians per hour, we find

$$\frac{d\theta}{dt} = \frac{2}{3}\left(\frac{\pi}{180}\right)(3600) \text{ rad/hr.} \qquad \text{1 hr = 3600 sec, 1 deg = } \pi/180 \text{ rad}$$

Substitution into the equation for dx/dt then gives

$$\frac{dx}{dt} = \left(-\frac{1200}{528}\right)(4)\left(\frac{2}{3}\right)\left(\frac{\pi}{180}\right)(3600) \approx -380.$$

The negative sign appears because the distance x is decreasing, so the aircraft is approaching the island at a speed of approximately 380 mi/hr when first detected by the radar. ■

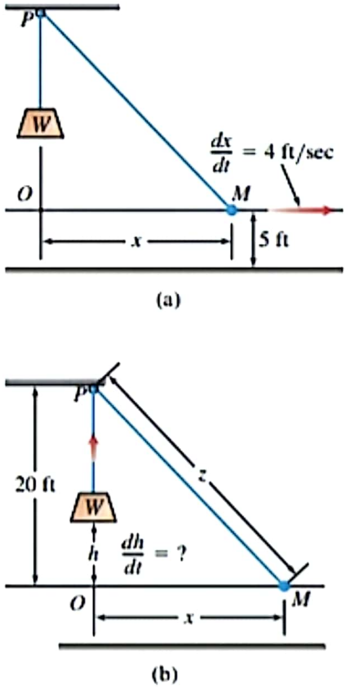


FIGURE 3.49 A worker at M walks to the right, pulling the weight W upward as the rope moves through the pulley P (Example 6).

EXAMPLE 6 Figure 3.49a shows a rope running through a pulley at P and bearing a weight W at one end. The other end is held 5 ft above the ground in the hand M of a worker. Suppose the pulley is 25 ft above ground, the rope is 45 ft long, and the worker is walking rapidly away from the vertical line PW at the rate of 4 ft/sec. How fast is the weight being raised when the worker's hand is 21 ft away from PW?

Solution We let OM be the horizontal line of length x ft from a point O directly below the pulley to the worker's hand M at any instant of time (Figure 3.49). Let h be the height of the weight W above O, and let z denote the length of rope from the pulley P to the worker's hand. We want to know dh/dt when $x = 21$ given that $dx/dt = 4$. Note that the height of P above O is 20 ft because O is 5 ft above the ground. We assume the angle at O is a right angle.

At any instant of time t we have the following relationships (see Figure 3.49b):

$$20 - h + z = 45 \qquad \text{Total length of rope is 45 ft.}$$

$$20^2 + x^2 = z^2. \qquad \text{Angle at } O \text{ is a right angle.}$$

If we solve for $z = 25 + h$ in the first equation, and substitute into the second equation, we have

$$20^2 + x^2 = (25 + h)^2. \tag{1}$$

Differentiating both sides with respect to t gives

$$2x\frac{dx}{dt} = 2(25 + h)\frac{dh}{dt},$$

and solving this last equation for dh/dt we find

$$\frac{dh}{dt} = \frac{x}{25 + h}\frac{dx}{dt}. \tag{2}$$

Since we know dx/dt, it remains only to find $25 + h$ at the instant when $x = 21$. From Equation (1),

$$20^2 + 21^2 = (25 + h)^2$$

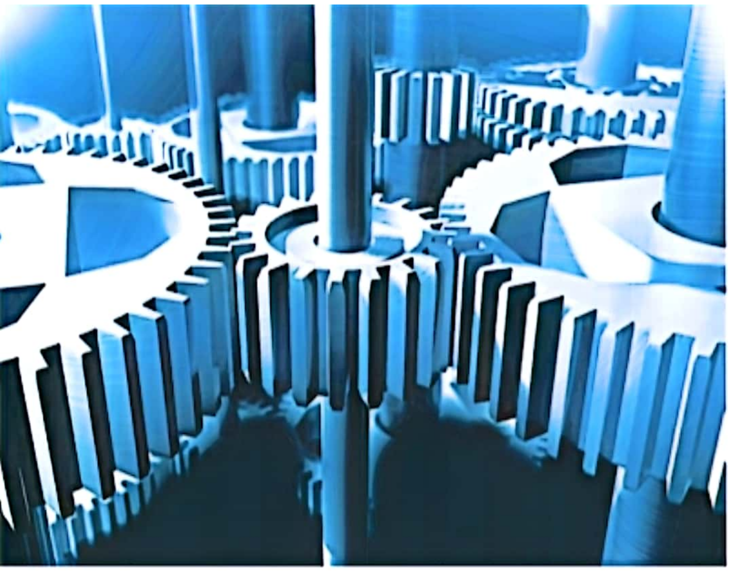
so that

$$(25 + h)^2 = 841, \qquad \text{or} \qquad 25 + h = 29.$$

Equation (2) now gives

$$\frac{dh}{dt} = \frac{21}{29}\cdot 4 = \frac{84}{29} \approx 2.9 \text{ ft/sec}$$

as the rate at which the weight is being raised when $x = 21$ ft. ■



4 Applications of Derivatives

OVERVIEW One of the most important applications of the derivative is its use as a tool for finding the optimal (best) solutions to problems. Optimization problems abound in mathematics, physical science and engineering, business and economics, and biology and medicine. For example, what are the height and diameter of the cylinder of largest volume that can be inscribed in a given sphere? What are the dimensions of the strongest rectangular wooden beam that can be cut from a cylindrical log of given diameter? Based on production costs and sales revenue, how many items should a manufacturer produce to maximize profit? How much does the trachea (windpipe) contract to expel air at the maximum speed during a cough? What is the branching angle at which blood vessels minimize the energy loss due to friction as blood flows through the branches?

In this chapter we use derivatives to find extreme values of functions, to determine and analyze the shapes of graphs, and to solve equations numerically. We also introduce the idea of recovering a function from its derivative. The key to many of these applications is the Mean Value Theorem, which paves the way to integral calculus.

4.1 Extreme Values of Functions

This section shows how to locate and identify extreme (maximum or minimum) values of a function from its derivative. Once we can do this, we can solve a variety of optimization problems (see Section 4.6). The domains of the functions we consider are intervals or unions of separate intervals.

DEFINITIONS Let f be a function with domain D. Then f has an **absolute maximum** value on D at a point c if

$$f(x) \leq f(c) \qquad \text{for all } x \text{ in } D$$

and an **absolute minimum** value on D at c if

$$f(x) \geq f(c) \qquad \text{for all } x \text{ in } D.$$

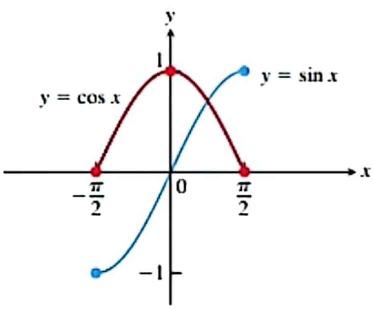


FIGURE 4.1 Absolute extrema for the sine and cosine functions on $[-\pi/2, \pi/2]$. These values can depend on the domain of a function.

Maximum and minimum values are called **extreme values** of the function f. Absolute maxima or minima are also referred to as **global** maxima or minima.

For example, on the closed interval $[-\pi/2, \pi/2]$ the function $f(x) = \cos x$ takes on an absolute maximum value of 1 (once) and an absolute minimum value of 0 (twice). On the same interval, the function $g(x) = \sin x$ takes on a maximum value of 1 and a minimum value of -1 (Figure 4.1).

Functions with the same defining rule or formula can have different extrema (maximum or minimum values), depending on the domain. We see this in the following example.

EXAMPLE 1 The absolute extrema of the following functions on their domains can be seen in Figure 4.2. Each function has the same defining equation, $y = x^2$, but the domains vary. Notice that a function might not have a maximum or minimum if the domain is unbounded or fails to contain an endpoint.

Function rule	Domain D	Absolute extrema on D
(a) $y = x^2$	$(-\infty, \infty)$	No absolute maximum Absolute minimum of 0 at $x = 0$
(b) $y = x^2$	$[0, 2]$	Absolute maximum of 4 at $x = 2$ Absolute minimum of 0 at $x = 0$
(c) $y = x^2$	$(0, 2]$	Absolute maximum of 4 at $x = 2$ No absolute minimum
(d) $y = x^2$	$(0, 2)$	No absolute extrema

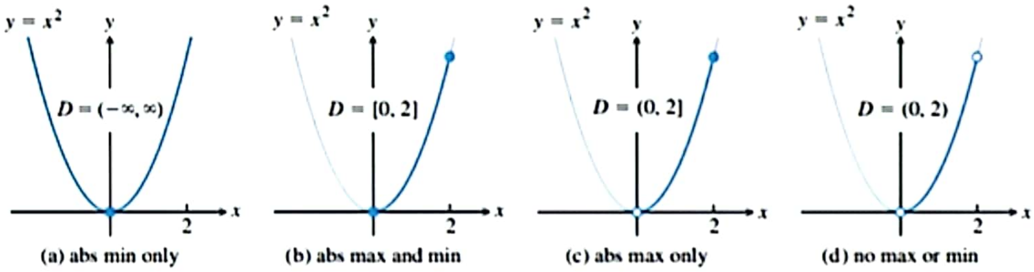


FIGURE 4.2 Graphs for Example 1.

HISTORICAL BIOGRAPHY
Daniel Bernoulli
(1700–1789)

Some of the functions in Example 1 did not have a maximum or a minimum value. The following theorem asserts that a function which is *continuous* over (or on) a finite *closed* interval $[a, b]$ has an absolute maximum and an absolute minimum value on the interval. We look for these extreme values when we graph a function.

THEOREM 1—The Extreme Value Theorem If f is continuous on a closed interval $[a, b]$, then f attains both an absolute maximum value M and an absolute minimum value m in $[a, b]$. That is, there are numbers x_1 and x_2 in $[a, b]$ with $f(x_1) = m$, $f(x_2) = M$, and $m \le f(x) \le M$ for every other x in $[a, b]$.

The proof of the Extreme Value Theorem requires a detailed knowledge of the real number system (see Appendix 7) and we will not give it here. Figure 4.3 illustrates possible locations for the absolute extrema of a continuous function on a closed interval $[a, b]$. As we observed for the function $y = \cos x$, it is possible that an absolute minimum (or absolute maximum) may occur at two or more different points of the interval.

The requirements in Theorem 1 that the interval be closed and finite, and that the function be continuous, are key ingredients. Without them, the conclusion of the theorem

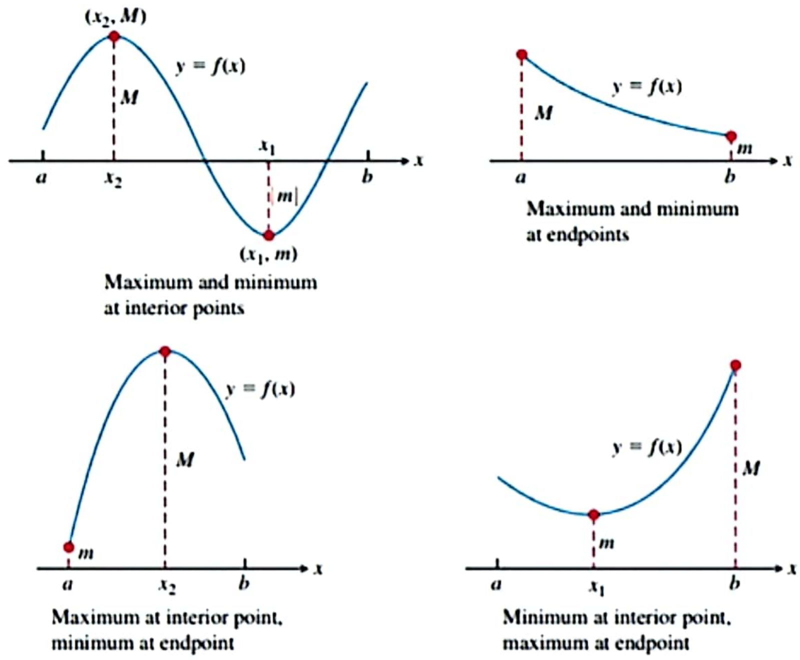


FIGURE 4.3 Some possibilities for a continuous function's maximum and minimum on a closed interval $[a, b]$.

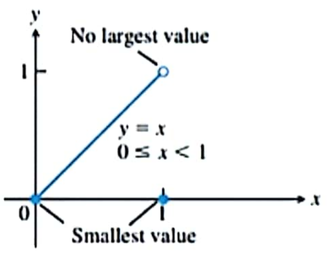


FIGURE 4.4 Even a single point of discontinuity can keep a function from having either a maximum or minimum value on a closed interval. The function

$$y = \begin{cases} x, & 0 \leq x < 1 \\ 0, & x = 1 \end{cases}$$

is continuous at every point of $[0, 1]$ except $x = 1$, yet its graph over $[0, 1]$ does not have a highest point.

need not hold. Example 1 shows that an absolute extreme value may not exist if the interval fails to be both closed and finite. The exponential function $y = e^x$ over $(-\infty, \infty)$ shows that neither extreme value need exist on an infinite interval. Figure 4.4 shows that the continuity requirement cannot be omitted.

Local (Relative) Extreme Values

Figure 4.5 shows a graph with five points where a function has extreme values on its domain $[a, b]$. The function's absolute minimum occurs at a even though at e the function's value is smaller than at any other point *nearby*. The curve rises to the left and falls to the right around c, making $f(c)$ a maximum locally. The function attains its absolute maximum at d. We now define what we mean by local extrema.

> **DEFINITIONS** A function f has a **local maximum** value at a point c within its domain D if $f(x) \leq f(c)$ for all $x \in D$ lying in some open interval containing c.
>
> A function f has a **local minimum** value at a point c within its domain D if $f(x) \geq f(c)$ for all $x \in D$ lying in some open interval containing c.

If the domain of f is the closed interval $[a, b]$, then f has a local maximum at the endpoint $x = a$, if $f(x) \leq f(a)$ for all x in some half-open interval $[a, a + \delta)$, $\delta > 0$. Likewise, f has a local maximum at an interior point $x = c$ if $f(x) \leq f(c)$ for all x in some open interval $(c - \delta, c + \delta)$, $\delta > 0$, and a local maximum at the endpoint $x = b$ if $f(x) \leq f(b)$ for all x in some half-open interval $(b - \delta, b]$, $\delta > 0$. The inequalities are reversed for local minimum values. In Figure 4.5, the function f has local maxima at c and d and local minima at a, e, and b. Local extrema are also called **relative extrema**. Some functions can have infinitely many local extrema, even over a finite interval. One example is the function $f(x) = \sin(1/x)$ on the interval $(0, 1]$. (We graphed this function in Figure 2.40.)

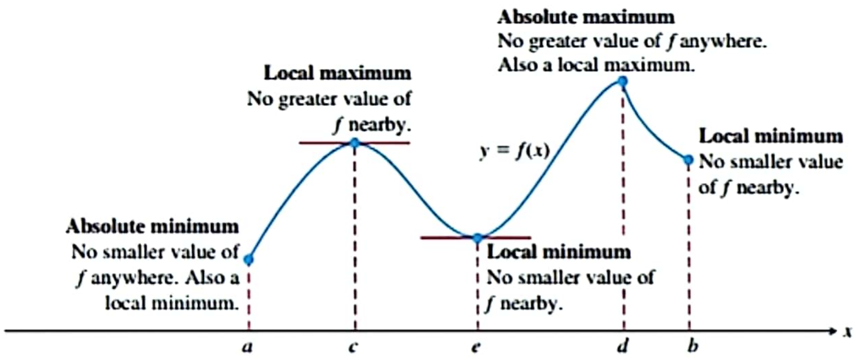


FIGURE 4.5 How to identify types of maxima and minima for a function with domain $a \le x \le b$.

An absolute maximum is also a local maximum. Being the largest value overall, it is also the largest value in its immediate neighborhood. Hence, *a list of all local maxima will automatically include the absolute maximum if there is one*. Similarly, *a list of all local minima will include the absolute minimum if there is one*.

Finding Extrema

The next theorem explains why we usually need to investigate only a few values to find a function's extrema.

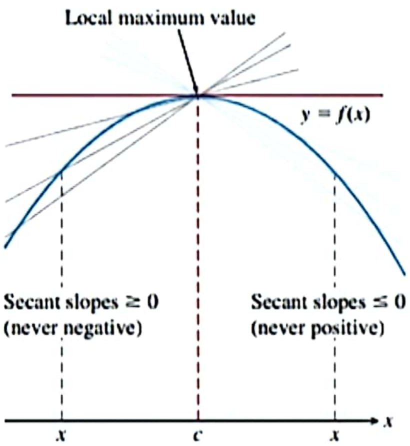


FIGURE 4.6 A curve with a local maximum value. The slope at c, simultaneously the limit of nonpositive numbers and nonnegative numbers, is zero.

THEOREM 2—The First Derivative Theorem for Local Extreme Values If f has a local maximum or minimum value at an interior point c of its domain, and if f' is defined at c, then

$$f'(c) = 0.$$

Proof To prove that $f'(c)$ is zero at a local extremum, we show first that $f'(c)$ cannot be positive and second that $f'(c)$ cannot be negative. The only number that is neither positive nor negative is zero, so that is what $f'(c)$ must be.

To begin, suppose that f has a local maximum value at $x = c$ (Figure 4.6) so that $f(x) - f(c) \le 0$ for all values of x near enough to c. Since c is an interior point of f's domain, $f'(c)$ is defined by the two-sided limit

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

This means that the right-hand and left-hand limits both exist at $x = c$ and equal $f'(c)$. When we examine these limits separately, we find that

$$f'(c) = \lim_{x \to c^+} \frac{f(x) - f(c)}{x - c} \le 0. \qquad \text{Because } (x - c) > 0 \text{ and } f(x) \le f(c) \tag{1}$$

Similarly,

$$f'(c) = \lim_{x \to c^-} \frac{f(x) - f(c)}{x - c} \ge 0. \qquad \text{Because } (x - c) < 0 \text{ and } f(x) \le f(c) \tag{2}$$

Together, Equations (1) and (2) imply $f'(c) = 0$.

This proves the theorem for local maximum values. To prove it for local minimum values, we simply use $f(x) \ge f(c)$, which reverses the inequalities in Equations (1) and (2). ■

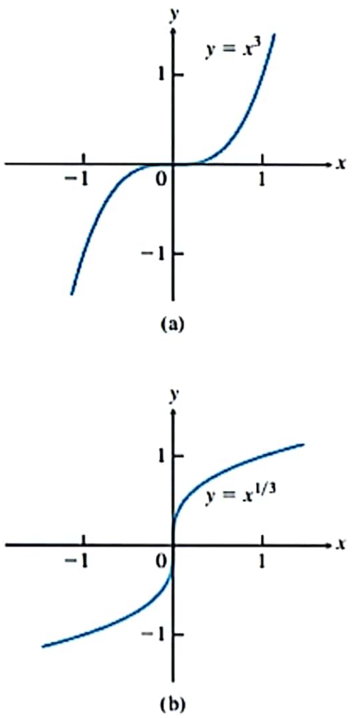


FIGURE 4.7 Critical points without extreme values. (a) $y' = 3x^2$ is 0 at $x = 0$, but $y = x^3$ has no extremum there. (b) $y' = (1/3)x^{-2/3}$ is undefined at $x = 0$, but $y = x^{1/3}$ has no extremum there.

Theorem 2 says that a function's first derivative is always zero at an interior point where the function has a local extreme value and the derivative is defined. If we recall that all the domains we consider are intervals or unions of separate intervals, the only places where a function f can possibly have an extreme value (local or global) are

1. interior points where $f' = 0$, — At $x = c$ and $x = e$ in Fig. 4.5
2. interior points where f' is undefined, — At $x = d$ in Fig. 4.5
3. endpoints of the domain of f. — At $x = a$ and $x = b$ in Fig. 4.5

The following definition helps us to summarize these results.

DEFINITION An interior point of the domain of a function f where f' is zero or undefined is a **critical point** of f.

Thus the only domain points where a function can assume extreme values are critical points and endpoints. However, be careful not to misinterpret what is being said here. A function may have a critical point at $x = c$ without having a local extreme value there. For instance, both of the functions $y = x^3$ and $y = x^{1/3}$ have critical points at the origin, but neither function has a local extreme value at the origin. Instead, each function has a *point of inflection* there (see Figure 4.7). We define and explore inflection points in Section 4.4.

Most problems that ask for extreme values call for finding the absolute extrema of a continuous function on a closed and finite interval. Theorem 1 assures us that such values exist; Theorem 2 tells us that they are taken on only at critical points and endpoints. Often we can simply list these points and calculate the corresponding function values to find what the largest and smallest values are, and where they are located. Of course, if the interval is not closed or not finite (such as $a < x < b$ or $a < x < \infty$), we have seen that absolute extrema need not exist. If an absolute maximum or minimum value does exist, it must occur at a critical point or at an included right- or left-hand endpoint of the interval.

How to Find the Absolute Extrema of a Continuous Function f on a Finite Closed Interval

1. Evaluate f at all critical points and endpoints.
2. Take the largest and smallest of these values.

EXAMPLE 2 Find the absolute maximum and minimum values of $f(x) = x^2$ on $[-2, 1]$.

Solution The function is differentiable over its entire domain, so the only critical point is where $f'(x) = 2x = 0$, namely $x = 0$. We need to check the function's values at $x = 0$ and at the endpoints $x = -2$ and $x = 1$:

Critical point value: $f(0) = 0$

Endpoint values: $f(-2) = 4$

$f(1) = 1.$

The function has an absolute maximum value of 4 at $x = -2$ and an absolute minimum value of 0 at $x = 0$. ■

EXAMPLE 3 Find the absolute maximum and minimum values of $f(x) = 10x(2 - \ln x)$ on the interval $[1, e^2]$.

4.2 The Mean Value Theorem

We know that constant functions have zero derivatives, but could there be a more complicated function whose derivative is always zero? If two functions have identical derivatives over an interval, how are the functions related? We answer these and other questions in this chapter by applying the Mean Value Theorem. First we introduce a special case, known as Rolle's Theorem, which is used to prove the Mean Value Theorem.

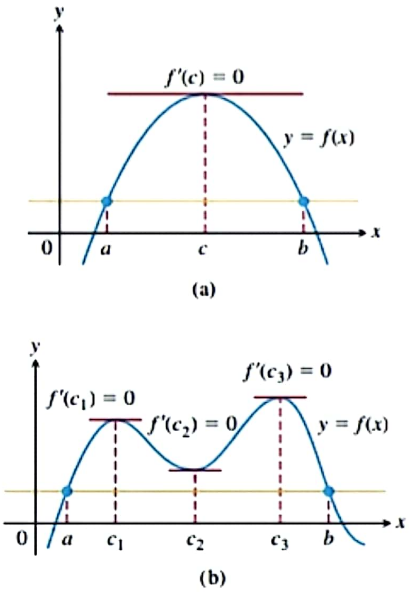


FIGURE 4.10 Rolle's Theorem says that a differentiable curve has at least one horizontal tangent between any two points where it crosses a horizontal line. It may have just one (a), or it may have more (b).

Rolle's Theorem

As suggested by its graph, if a differentiable function crosses a horizontal line at two different points, there is at least one point between them where the tangent to the graph is horizontal and the derivative is zero (Figure 4.10). We now state and prove this result.

> **THEOREM 3—Rolle's Theorem** Suppose that $y = f(x)$ is continuous over the closed interval $[a, b]$ and differentiable at every point of its interior (a, b). If $f(a) = f(b)$, then there is at least one number c in (a, b) at which $f'(c) = 0$.

Proof Being continuous, f assumes absolute maximum and minimum values on $[a, b]$ by Theorem 1. These can occur only

1. at interior points where f' is zero,
2. at interior points where f' does not exist,
3. at endpoints of the function's domain, in this case a and b.

By hypothesis, f has a derivative at every interior point. That rules out possibility (2), leaving us with interior points where $f' = 0$ and with the two endpoints a and b.

If either the maximum or the minimum occurs at a point c between a and b, then $f'(c) = 0$ by Theorem 2 in Section 4.1, and we have found a point for Rolle's Theorem.

If both the absolute maximum and the absolute minimum occur at the endpoints, then because $f(a) = f(b)$ it must be the case that f is a constant function with $f(x) = f(a) = f(b)$ for every $x \in [a, b]$. Therefore $f'(x) = 0$ and the point c can be taken anywhere in the interior (a, b). ■

The hypotheses of Theorem 3 are essential. If they fail at even one point, the graph may not have a horizontal tangent (Figure 4.11).

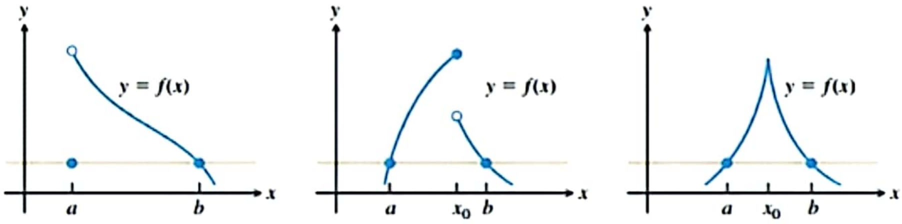
Rolle's Theorem may be combined with the Intermediate Value Theorem to show when there is only one real solution of an equation $f(x) = 0$, as we illustrate in the next example.

HISTORICAL BIOGRAPHY
Michel Rolle
(1652–1719)

EXAMPLE 1 Show that the equation

$$x^3 + 3x + 1 = 0$$

has exactly one real solution.



(a) Discontinuous at an endpoint of $[a, b]$

(b) Discontinuous at an interior point of $[a, b]$

(c) Continuous on $[a, b]$ but not differentiable at an interior point

FIGURE 4.11 There may be no horizontal tangent if the hypotheses of Rolle's Theorem do not hold.

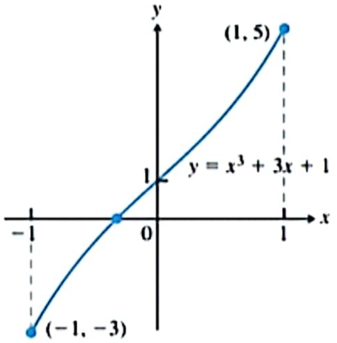


FIGURE 4.12 The only real zero of the polynomial $y = x^3 + 3x + 1$ is the one shown here where the curve crosses the x-axis between -1 and 0 (Example 1).

Solution We define the continuous function

$$f(x) = x^3 + 3x + 1.$$

Since $f(-1) = -3$ and $f(0) = 1$, the Intermediate Value Theorem tells us that the graph of f crosses the x-axis somewhere in the open interval $(-1, 0)$. (See Figure 4.12.) Now, if there were even two points $x = a$ and $x = b$ where $f(x)$ was zero, Rolle's Theorem would guarantee the existence of a point $x = c$ in between them where f' was zero. However, the derivative

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

is never zero (because it is always positive). Therefore, f has no more than one zero. ■

Our main use of Rolle's Theorem is in proving the Mean Value Theorem.

The Mean Value Theorem

The Mean Value Theorem, which was first stated by Joseph-Louis Lagrange, is a slanted version of Rolle's Theorem (Figure 4.13). The Mean Value Theorem guarantees that there is a point where the tangent line is parallel to the secant joining A and B.

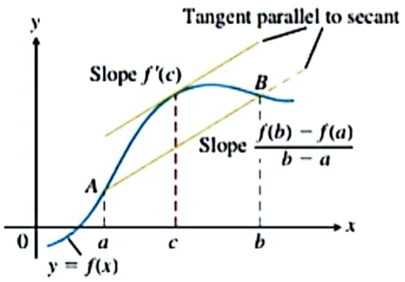


FIGURE 4.13 Geometrically, the Mean Value Theorem says that somewhere between a and b the curve has at least one tangent parallel to the secant joining A and B.

> **THEOREM 4—The Mean Value Theorem** Suppose $y = f(x)$ is continuous over a closed interval $[a, b]$ and differentiable on the interval's interior (a, b). Then there is at least one point c in (a, b) at which
>
> $$\frac{f(b) - f(a)}{b - a} = f'(c). \tag{1}$$

Proof We picture the graph of f and draw a line through the points $A(a, f(a))$ and $B(b, f(b))$. (See Figure 4.14.) The secant line is the graph of the function

$$g(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \tag{2}$$

(point-slope equation). The vertical difference between the graphs of f and g at x is

$$\begin{aligned} h(x) &= f(x) - g(x) \\ &= f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a). \end{aligned} \tag{3}$$

Figure 4.15 shows the graphs of f, g, and h together.

HISTORICAL BIOGRAPHY
Joseph-Louis Lagrange
(1736–1813)

5

Integrals

OVERVIEW A great achievement of classical geometry was obtaining formulas for the areas and volumes of triangles, spheres, and cones. In this chapter we develop a method to calculate the areas and volumes of very general shapes. This method, called *integration*, is a way to calculate much more than areas and volumes. The *definite integral* is the key tool in calculus for defining and calculating many important quantities, such as areas, volumes, lengths of curved paths, probabilities, averages, energy consumption, the weights of various objects, and the forces against a dam's floodgates, just to mention a few. Many of these applications are studied in subsequent chapters.

As with the derivative, the definite integral also arises as a limit, this time of increasingly fine approximations to the quantity of interest. The idea behind the integral is that we can effectively compute such quantities by breaking them into small pieces, and then summing the contributions from each piece. We then consider what happens when more and more, smaller and smaller pieces are taken in the summation process. As the number of terms contributing to the sum approaches infinity and we take the limit of these sums in a way described in Section 5.3, the result is a definite integral. By considering the rate of change of the area under a graph, we prove that definite integrals are connected to antiderivatives, a connection that gives one of the most important relationships in calculus.

5.1 Area and Estimating with Finite Sums

The basis for formulating definite integrals is the construction of appropriate approximations by finite sums. In this section we consider three examples of this construction process: finding the area under a graph, the distance traveled by a moving object, and the average value of a function. Although we need to define precisely what we mean by the area of a general region in the plane, or the average value of a function over a closed interval, we do have intuitive ideas of what these notions mean. So in this section we begin our approach to integration by *approximating* these quantities with finite sums. We also consider what happens when we take more and more terms in the summation process. In subsequent sections we look at taking the limit of these sums as the number of terms goes to infinity, which then leads to precise definitions of the quantities being approximated here.

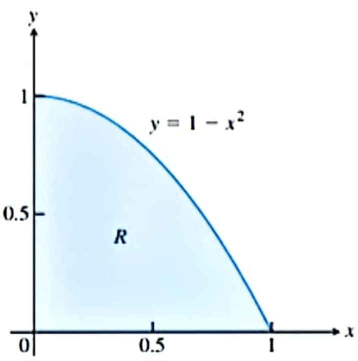


FIGURE 5.1 The area of the region R cannot be found by a simple formula.

Area

Suppose we want to find the area of the shaded region R that lies above the x-axis, below the graph of $y = 1 - x^2$, and between the vertical lines $x = 0$ and $x = 1$ (Figure 5.1). Unfortunately, there is no simple geometric formula for calculating the areas of general shapes having curved boundaries like the region R. How, then, can we find the area of R?

While we do not yet have a method for determining the exact area of R, we can approximate it in a simple way. Figure 5.2a shows two rectangles that together contain the

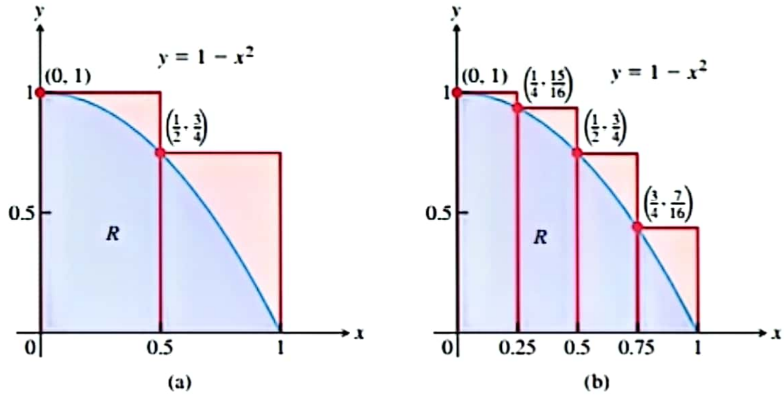


FIGURE 5.2 (a) We get an upper estimate of the area of R by using two rectangles containing R. (b) Four rectangles give a better upper estimate. Both estimates overshoot the true value for the area by the amount shaded in light red.

region R. Each rectangle has width $1/2$ and they have heights 1 and $3/4$, moving from left to right. The height of each rectangle is the maximum value of the function f in each subinterval. Because the function f is decreasing, the height is its value at the left endpoint of the subinterval of $[0, 1]$ forming the base of the rectangle. The total area of the two rectangles approximates the area A of the region R,

$$A \approx 1 \cdot \frac{1}{2} + \frac{3}{4} \cdot \frac{1}{2} = \frac{7}{8} = 0.875.$$

This estimate is larger than the true area A since the two rectangles contain R. We say that 0.875 is an **upper sum** because it is obtained by taking the height of each rectangle as the maximum (uppermost) value of $f(x)$ for a point x in the base interval of the rectangle. In Figure 5.2b, we improve our estimate by using four thinner rectangles, each of width $1/4$, which taken together contain the region R. These four rectangles give the approximation

$$A \approx 1 \cdot \frac{1}{4} + \frac{15}{16} \cdot \frac{1}{4} + \frac{3}{4} \cdot \frac{1}{4} + \frac{7}{16} \cdot \frac{1}{4} = \frac{25}{32} = 0.78125,$$

which is still greater than A since the four rectangles contain R.

Suppose instead we use four rectangles contained *inside* the region R to estimate the area, as in Figure 5.3a. Each rectangle has width $1/4$ as before, but the rectangles are shorter and

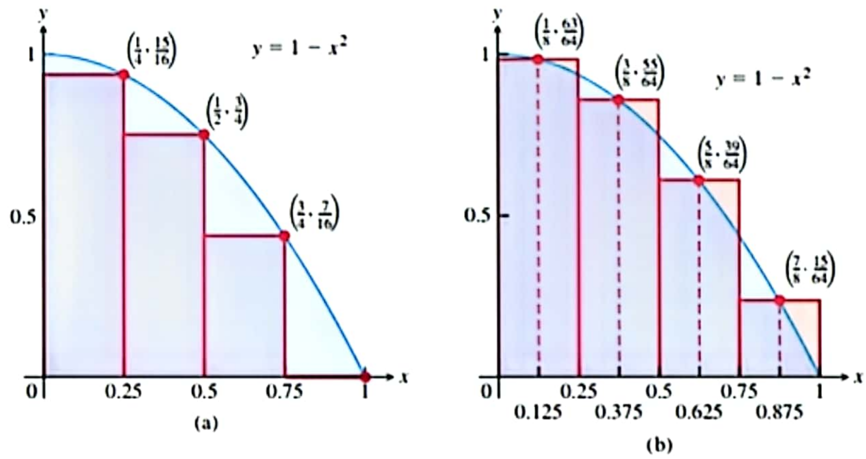


FIGURE 5.3 (a) Rectangles contained in R give an estimate for the area that undershoots the true value by the amount shaded in light blue. (b) The midpoint rule uses rectangles whose height is the value of $y = f(x)$ at the midpoints of their bases. The estimate appears closer to the true value of the area because the light red overshoot areas roughly balance the light blue undershoot areas.

lie entirely beneath the graph of f. The function $f(x) = 1 - x^2$ is decreasing on $[0, 1]$, so the height of each of these rectangles is given by the value of f at the right endpoint of the subinterval forming its base. The fourth rectangle has zero height and therefore contributes no area. Summing these rectangles with heights equal to the minimum value of $f(x)$ for a point x in each base subinterval gives a **lower sum** approximation to the area,

$$A \approx \frac{15}{16}\cdot\frac{1}{4} + \frac{3}{4}\cdot\frac{1}{4} + \frac{7}{16}\cdot\frac{1}{4} + 0\cdot\frac{1}{4} = \frac{17}{32} = 0.53125.$$

This estimate is smaller than the area A since the rectangles all lie inside of the region R. The true value of A lies somewhere between these lower and upper sums:

$$0.53125 < A < 0.78125.$$

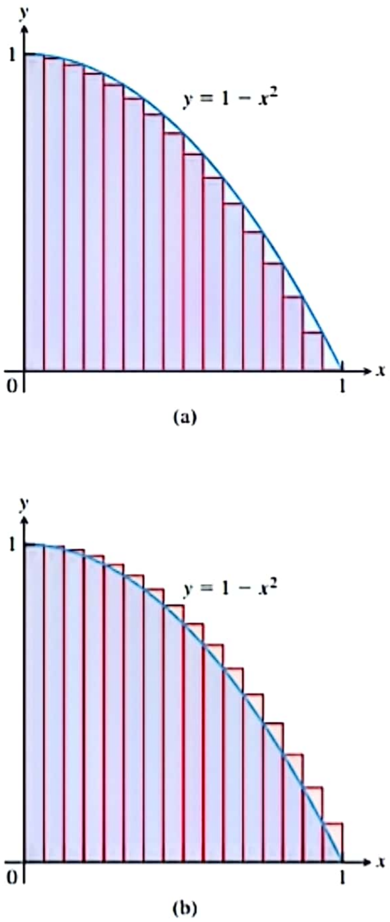


FIGURE 5.4 (a) A lower sum using 16 rectangles of equal width $\Delta x = 1/16$. (b) An upper sum using 16 rectangles.

By considering both lower and upper sum approximations, we get not only estimates for the area, but also a bound on the size of the possible error in these estimates, since the true value of the area lies somewhere between them. Here the error cannot be greater than the difference $0.78125 - 0.53125 = 0.25$.

Yet another estimate can be obtained by using rectangles whose heights are the values of f at the midpoints of their bases (Figure 5.3b). This method of estimation is called the **midpoint rule** for approximating the area. The midpoint rule gives an estimate that is between a lower sum and an upper sum, but it is not quite so clear whether it overestimates or underestimates the true area. With four rectangles of width $1/4$ as before, the midpoint rule estimates the area of R to be

$$A \approx \frac{63}{64}\cdot\frac{1}{4} + \frac{55}{64}\cdot\frac{1}{4} + \frac{39}{64}\cdot\frac{1}{4} + \frac{15}{64}\cdot\frac{1}{4} = \frac{172}{64}\cdot\frac{1}{4} = 0.671875.$$

In each of our computed sums, the interval $[a, b]$ over which the function f is defined was subdivided into n subintervals of equal width (also called length) $\Delta x = (b - a)/n$, and f was evaluated at a point in each subinterval: c_1 in the first subinterval, c_2 in the second subinterval, and so on. The finite sums then all take the form

$$f(c_1)\,\Delta x + f(c_2)\,\Delta x + f(c_3)\,\Delta x + \cdots + f(c_n)\,\Delta x.$$

By taking more and more rectangles, with each rectangle thinner than before, it appears that these finite sums give better and better approximations to the true area of the region R.

Figure 5.4a shows a lower sum approximation for the area of R using 16 rectangles of equal width. The sum of their areas is 0.634765625, which appears close to the true area, but is still smaller since the rectangles lie inside R.

Figure 5.4b shows an upper sum approximation using 16 rectangles of equal width. The sum of their areas is 0.697265625, which is somewhat larger than the true area because the rectangles taken together contain R. The midpoint rule for 16 rectangles gives a total area approximation of 0.6669921875, but it is not immediately clear whether this estimate is larger or smaller than the true area.

EXAMPLE 1 Table 5.1 shows the values of upper and lower sum approximations to the area of R, using up to 1000 rectangles. In Section 5.2 we will see how to get an exact value of the areas of regions such as R by taking a limit as the base width of each rectangle goes to zero and the number of rectangles goes to infinity. With the techniques developed there, we will be able to show that the area of R is exactly $2/3$. ■

Distance Traveled

Suppose we know the velocity function $v(t)$ of a car moving down a highway, without changing direction, and want to know how far it traveled between times $t = a$ and $t = b$. The position function $s(t)$ of the car has derivative $v(t)$. If we can find an antiderivative $F(t)$

TABLE 5.1 Finite approximations for the area of R

Number of subintervals	Lower sum	Midpoint sum	Upper sum
2	0.375	0.6875	0.875
4	0.53125	0.671875	0.78125
16	0.634765625	0.6669921875	0.697265625
50	0.6566	0.6667	0.6766
100	0.66165	0.666675	0.67165
1000	0.6661665	0.66666675	0.6671665

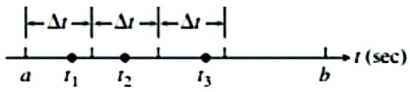
of $v(t)$ then we can find the car's position function $s(t)$ by setting $s(t) = F(t) + C$. The distance traveled can then be found by calculating the change in position, $s(b) - s(a) = F(b) - F(a)$. If the velocity function is known only by the readings at various times of a speedometer on the car, then we have no formula from which to obtain an antiderivative function for velocity. So what do we do in this situation?

When we don't know an antiderivative for the velocity function $v(t)$, we can approximate the distance traveled with finite sums in a way similar to our estimates for area discussed before. We subdivide the interval $[a, b]$ into short time intervals on each of which the velocity is considered to be fairly constant. Then we approximate the distance traveled on each time subinterval with the usual distance formula

$$\text{distance} = \text{velocity} \times \text{time}$$

and add the results across $[a, b]$.

Suppose the subdivided interval looks like



with the subintervals all of equal length Δt. Pick a number t_1 in the first interval. If Δt is so small that the velocity barely changes over a short time interval of duration Δt, then the distance traveled in the first time interval is about $v(t_1)\,\Delta t$. If t_2 is a number in the second interval, the distance traveled in the second time interval is about $v(t_2)\,\Delta t$. The sum of the distances traveled over all the time intervals is

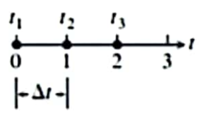
$$D \approx v(t_1)\,\Delta t + v(t_2)\,\Delta t + \cdots + v(t_n)\,\Delta t,$$

where n is the total number of subintervals.

EXAMPLE 2 The velocity function of a projectile fired straight into the air is $f(t) = 160 - 9.8t$ m/sec. Use the summation technique just described to estimate how far the projectile rises during the first 3 sec. How close do the sums come to the exact value of 435.9 m? (You will learn how to compute the exact value easily in Section 5.4.)

Solution We explore the results for different numbers of intervals and different choices of evaluation points. Notice that $f(t)$ is decreasing, so choosing left endpoints gives an upper sum estimate; choosing right endpoints gives a lower sum estimate.

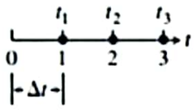
(a) *Three subintervals of length* 1, *with f evaluated at left endpoints giving an upper sum:*



With f evaluated at $t = 0$, 1, and 2, we have

$$\begin{aligned} D &\approx f(t_1)\,\Delta t + f(t_2)\,\Delta t + f(t_3)\,\Delta t \\ &= [160 - 9.8(0)](1) + [160 - 9.8(1)](1) + [160 - 9.8(2)](1) \\ &= 450.6. \end{aligned}$$

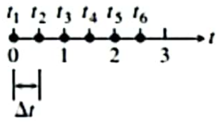
(b) *Three subintervals of length* 1, *with f evaluated at right endpoints giving a lower sum:*

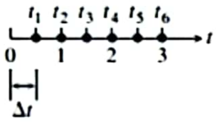


With f evaluated at $t = 1$, 2, and 3, we have

$$\begin{aligned} D &\approx f(t_1)\,\Delta t + f(t_2)\,\Delta t + f(t_3)\,\Delta t \\ &= [160 - 9.8(1)](1) + [160 - 9.8(2)](1) + [160 - 9.8(3)](1) \\ &= 421.2. \end{aligned}$$

(c) *With six subintervals of length* $1/2$, *we get*





These estimates give an upper sum using left endpoints: $D \approx 443.25$; and a lower sum using right endpoints: $D \approx 428.55$. These six-interval estimates are somewhat closer than the three-interval estimates. The results improve as the subintervals get shorter.

As we can see in Table 5.2, the left-endpoint upper sums approach the true value 435.9 from above, whereas the right-endpoint lower sums approach it from below. The true value lies between these upper and lower sums. The magnitude of the error in the closest entries is 0.23, a small percentage of the true value.

$$\begin{aligned} \text{Error magnitude} &= |\text{true value} - \text{calculated value}| \\ &= |435.9 - 435.67| = 0.23. \end{aligned}$$

$$\text{Error percentage} = \frac{0.23}{435.9} \approx 0.05\%.$$

It would be reasonable to conclude from the table's last entries that the projectile rose about 436 m during its first 3 sec of flight. ■

TABLE 5.2 Travel-distance estimates

Number of subintervals	Length of each subinterval	Upper sum	Lower sum
3	1	450.6	421.2
6	1/2	443.25	428.55
12	1/4	439.58	432.23
24	1/8	437.74	434.06
48	1/16	436.82	434.98
96	1/32	436.36	435.44
192	1/64	436.13	435.67

Displacement Versus Distance Traveled

If an object with position function $s(t)$ moves along a coordinate line without changing direction, we can calculate the total distance it travels from $t = a$ to $t = b$ by summing the distance traveled over small intervals, as in Example 2. If the object reverses direction one or more times during the trip, then we need to use the object's *speed* $|v(t)|$, which is the absolute value of its velocity function, $v(t)$, to find the total distance traveled. Using the velocity itself, as in Example 2, gives instead an estimate to the object's **displacement**, $s(b) - s(a)$, the difference between its initial and final positions.

To see why using the velocity function in the summation process gives an estimate to the displacement, partition the time interval $[a, b]$ into small enough equal subintervals Δt so that the object's velocity does not change very much from time t_{k-1} to t_k. Then $v(t_k)$ gives a good approximation of the velocity throughout the interval. Accordingly, the change in the object's position coordinate, which is its displacement during the time interval, is about

$$v(t_k)\,\Delta t.$$

The change is positive if $v(t_k)$ is positive and negative if $v(t_k)$ is negative.

In either case, the distance traveled by the object during the subinterval is about

$$|v(t_k)|\,\Delta t.$$

The **total distance traveled** over the time interval is approximately the sum

$$|v(t_1)|\,\Delta t + |v(t_2)|\,\Delta t + \cdots + |v(t_n)|\,\Delta t.$$

We revisit these ideas in Section 5.4.

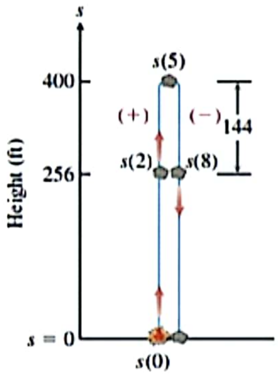


FIGURE 5.5 The rock in Example 3. The height $s = 256$ ft is reached at $t = 2$ and $t = 8$ sec. The rock falls 144 ft from its maximum height when $t = 8$.

EXAMPLE 3 In Example 4 in Section 3.4, we analyzed the motion of a heavy rock blown straight up by a dynamite blast. In that example, we found the velocity of the rock at any time during its motion to be $v(t) = 160 - 32t$ ft/sec. The rock was 256 ft above the ground 2 sec after the explosion, continued upward to reach a maximum height of 400 ft at 5 sec after the explosion, and then fell back down to reach the height of 256 ft again at $t = 8$ sec after the explosion. (See Figure 5.5.)

If we follow a procedure like that presented in Example 2, and use the velocity function $v(t)$ in the summation process over the time interval $[0, 8]$, we will obtain an estimate to the rock's 256 ft *height* above the ground at $t = 8$. The positive upward motion (which yields a positive distance change of 144 ft from the height of 256 ft to the maximum height) is canceled by the negative downward motion (giving a negative change of 144 ft from the maximum height down to 256 ft again), so the displacement or height above the ground is estimated from the velocity function.

On the other hand, if the absolute value $|v(t)|$ is used in the summation process, we will obtain an estimate to the *total distance* the rock has traveled: the maximum height reached of 400 ft plus the additional distance of 144 ft it has fallen back down from that maximum when it again reaches the height of 256 ft at $t = 8$ sec. That is, using the absolute value of the velocity function in the summation process over the time interval $[0, 8]$, we obtain an estimate to 544 ft, the total distance up and down that the rock has traveled in 8 sec. There is no cancelation of distance changes due to sign changes in the velocity function, so we estimate distance traveled rather than displacement when we use the absolute value of the velocity function (that is, the speed of the rock).

As an illustration of our discussion, we subdivide the interval $[0, 8]$ into sixteen subintervals of length $\Delta t = 1/2$ and take the right endpoint of each subinterval in our calculations. Table 5.3 shows the values of the velocity function at these endpoints.

TABLE 5.3 Velocity function

t	$v(t)$	t	$v(t)$
0	160	4.5	16
0.5	144	5.0	0
1.0	128	5.5	−16
1.5	112	6.0	−32
2.0	96	6.5	−48
2.5	80	7.0	−64
3.0	64	7.5	−80
3.5	48	8.0	−96
4.0	32		

Using $v(t)$ in the summation process, we estimate the displacement at $t = 8$:

$$(144 + 128 + 112 + 96 + 80 + 64 + 48 + 32 + 16$$
$$+ 0 - 16 - 32 - 48 - 64 - 80 - 96)\cdot\frac{1}{2} = 192$$
$$\text{Error magnitude} = 256 - 192 = 64$$

Using $|v(t)|$ in the summation process, we estimate the total distance traveled over the time interval $[0, 8]$:

$$(144 + 128 + 112 + 96 + 80 + 64 + 48 + 32 + 16$$
$$+ 0 + 16 + 32 + 48 + 64 + 80 + 96)\cdot\frac{1}{2} = 528$$
$$\text{Error magnitude} = 544 - 528 = 16$$

If we take more and more subintervals of $[0, 8]$ in our calculations, the estimates to the heights 256 ft and 544 ft improve, approaching them as shown in Table 5.4. ■

TABLE 5.4 Travel estimates for a rock blown straight up during the time interval $[0, 8]$

Number of subintervals	Length of each subinterval	Displacement	Total distance
16	1/2	192.0	528.0
32	1/4	224.0	536.0
64	1/8	240.0	540.0
128	1/16	248.0	542.0
256	1/32	252.0	543.0
512	1/64	254.0	543.5

Average Value of a Nonnegative Continuous Function

The average value of a collection of n numbers $x_1, x_2, \ldots, x_n$ is obtained by adding them together and dividing by n. But what is the average value of a continuous function f on an interval $[a, b]$? Such a function can assume infinitely many values. For example, the temperature at a certain location in a town is a continuous function that goes up and down each day. What does it mean to say that the average temperature in the town over the course of a day is 73 degrees?

When a function is constant, this question is easy to answer. A function with constant value c on an interval $[a, b]$ has average value c. When c is positive, its graph over $[a, b]$ gives a rectangle of height c. The average value of the function can then be interpreted geometrically as the area of this rectangle divided by its width $b - a$ (Figure 5.6a).

What if we want to find the average value of a nonconstant function, such as the function g in Figure 5.6b? We can think of this graph as a snapshot of the height of some water

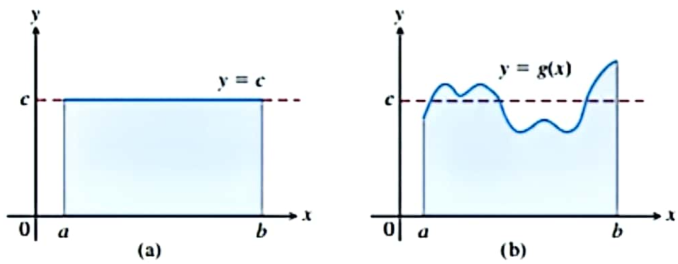


FIGURE 5.6 (a) The average value of $f(x) = c$ on $[a, b]$ is the area of the rectangle divided by $b - a$. (b) The average value of $g(x)$ on $[a, b]$ is the area beneath its graph divided by $b - a$.

5.3 The Definite Integral

In Section 5.2 we investigated the limit of a finite sum for a function defined over a closed interval $[a, b]$ using n subintervals of equal width (or length), $(b - a)/n$. In this section we consider the limit of more general Riemann sums as the norm of the partitions of $[a, b]$ approaches zero. For general Riemann sums, the subintervals of the partitions need not have equal widths. The limiting process then leads to the definition of the *definite integral* of a function over a closed interval $[a, b]$.

Definition of the Definite Integral

The definition of the definite integral is based on the idea that for certain functions, as the norm of the partitions of $[a, b]$ approaches zero, the values of the corresponding Riemann sums approach a limiting value J. What we mean by this limit is that a Riemann sum will be close to the number J provided that the norm of its partition is sufficiently small (so that all of its subintervals have thin enough widths). We introduce the symbol ϵ as a small positive number that specifies how close to J the Riemann sum must be, and the symbol δ as a second small positive number that specifies how small the norm of a partition must be in order for convergence to happen. We now define this limit precisely.

DEFINITION Let $f(x)$ be a function defined on a closed interval $[a, b]$. We say that a number J is the **definite integral of f over $[a, b]$** and that J is the limit of the Riemann sums $\sum_{k=1}^{n} f(c_k)\,\Delta x_k$ if the following condition is satisfied:

Given any number $\epsilon > 0$ there is a corresponding number $\delta > 0$ such that for every partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ with $\|P\| < \delta$ and any choice of c_k in $[x_{k-1}, x_k]$, we have

$$\left| \sum_{k=1}^{n} f(c_k)\,\Delta x_k - J \right| < \epsilon.$$

The definition involves a limiting process in which the norm of the partition goes to zero.

We have many choices for a partition P with norm going to zero, and many choices of points c_k for each partition. The definite integral exists when we always get the same limit J, no matter what choices are made. When the limit exists we write it as the definite integral

$$J = \lim_{\|P\| \to 0} \sum_{k=1}^{n} f(c_k)\,\Delta x_k.$$

The limit of any Riemann sum is always taken as the norm of the partitions approaches zero and the number of subintervals goes to infinity.

Leibniz introduced a notation for the definite integral that captures its construction as a limit of Riemann sums. He envisioned the finite sums $\sum_{k=1}^{n} f(c_k)\,\Delta x_k$ becoming an infinite sum of function values $f(x)$ multiplied by "infinitesimal" subinterval widths dx. The sum symbol $\sum$ is replaced in the limit by the integral symbol $\int$, whose origin is in the letter "S." The function values $f(c_k)$ are replaced by a continuous selection of function values $f(x)$. The subinterval widths Δx_k become the differential dx. It is as if we are summing all products of the form $f(x) \cdot dx$ as x goes from a to b. While this notation captures the process of constructing an integral, it is Riemann's definition that gives a precise meaning to the definite integral.

The symbol for the number J in the definition of the **definite integral** is

$$\int_a^b f(x)\,dx,$$

which is read as "the integral from a to b of f of x dee x" or sometimes as "the integral from a to b of f of x with respect to x." The component parts in the integral symbol also have names:

Upper limit of integration

The function is the integrand.

Integral sign

x is the variable of integration.

$$\underbrace{\int_a^b f(x)\,dx}_{\text{Integral of } f \text{ from } a \text{ to } b}$$

Lower limit of integration

When you find the value of the integral, you have evaluated the integral.

When the condition in the definition is satisfied, we say that the Riemann sums of f on $[a, b]$ **converge** to the definite integral $J = \int_a^b f(x)\,dx$ and that f is **integrable** over $[a, b]$.

In the cases where the subintervals all have equal width $\Delta x = (b - a)/n$, we can form each Riemann sum as

$$S_n = \sum_{k=1}^{n} f(c_k)\,\Delta x_k = \sum_{k=1}^{n} f(c_k)\left(\frac{b-a}{n}\right). \qquad \Delta x_k = \Delta x = (b-a)/n \text{ for all } k$$

where c_k is chosen in the kth subinterval. When the limit of these Riemann sums as $n \to \infty$ exists and is equal to J, then J is the definite integral of f over $[a, b]$, so

$$J = \int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{k=1}^{n} f(c_k)\left(\frac{b-a}{n}\right) \qquad \|P\| \to 0 \text{ means } n \to \infty.$$

If we pick the point c_k at the right endpoint of the kth subinterval, so $c_k = a + k\,\Delta x = a + k(b - a)/n$, then the formula for the definite integral becomes

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{k=1}^{n} f\left(a + k\frac{(b-a)}{n}\right)\left(\frac{b-a}{n}\right) \qquad (1)$$

Equation (1) gives one explicit formula that can be used to compute definite integrals. Other choices of partitions and locations of points c_k result in the same value for the definite integral when we take the limit as $n \to \infty$ provided that the norm of the partition approaches zero.

The value of the definite integral of a function over any particular interval depends on the function, not on the letter we choose to represent its independent variable. If we decide to use t or u instead of x, we simply write the integral as

$$\int_a^b f(t)\,dt \quad \text{or} \quad \int_a^b f(u)\,du \quad \text{instead of} \quad \int_a^b f(x)\,dx.$$

No matter how we write the integral, it is still the same number that is defined as a limit of Riemann sums. Since it does not matter what letter we use, the variable of integration is called a **dummy variable** representing the real numbers in the closed interval $[a, b]$.

Integrable and Nonintegrable Functions

Not every function defined over the closed interval $[a, b]$ is integrable there, even if the function is bounded. That is, the Riemann sums for some functions may not converge to the same limiting value, or to any value at all. A full development of exactly which functions defined over $[a, b]$ are integrable requires advanced mathematical analysis, but fortunately most functions that commonly occur in applications are integrable. In particular, every *continuous* function over $[a, b]$ is integrable over this interval, and so is every function having no more than a finite number of jump discontinuities on $[a, b]$. (See Figures 1.9 and 1.10. The latter functions are called *piecewise-continuous functions*, and they are defined in Additional Exercises 11–18 at the end of this chapter.) The following theorem, which is proved in more advanced courses, establishes these results.

THEOREM 1—Integrability of Continuous Functions If a function f is continuous over the interval $[a, b]$, or if f has at most finitely many jump discontinuities there, then the definite integral $\int_a^b f(x)\,dx$ exists and f is integrable over $[a, b]$.

The idea behind Theorem 1 for continuous functions is given in Exercises 86 and 87. Briefly, when f is continuous we can choose each c_k so that $f(c_k)$ gives the maximum value of f on the subinterval $[x_{k-1}, x_k]$, resulting in an upper sum. Likewise, we can choose c_k to give the minimum value of f on $[x_{k-1}, x_k]$ to obtain a lower sum. The upper and lower sums can be shown to converge to the same limiting value as the norm of the partition P tends to zero. Moreover, every Riemann sum is trapped between the values of the upper and lower sums, so every Riemann sum converges to the same limit as well. Therefore, the number J in the definition of the definite integral exists, and the continuous function f is integrable over $[a, b]$.

For integrability to fail, a function needs to be sufficiently discontinuous that the region between its graph and the x-axis cannot be approximated well by increasingly thin rectangles. Our first example shows a function that is not integrable over a closed interval.

EXAMPLE 1 The function

$$f(x) = \begin{cases} 1, & \text{if } x \text{ is rational} \\ 0, & \text{if } x \text{ is irrational} \end{cases}$$

has no Riemann integral over $[0, 1]$. Underlying this is the fact that between any two numbers there is both a rational number and an irrational number. Thus the function jumps up and down too erratically over $[0, 1]$ to allow the region beneath its graph and above the x-axis to be approximated by rectangles, no matter how thin they are. We show, in fact, that upper sum approximations and lower sum approximations converge to different limiting values.

If we pick a partition P of $[0, 1]$ and choose c_k to be the point giving the maximum value for f on $[x_{k-1}, x_k]$ then the corresponding Riemann sum is

$$U = \sum_{k=1}^{n} f(c_k)\,\Delta x_k = \sum_{k=1}^{n} (1)\,\Delta x_k = 1,$$

since each subinterval $[x_{k-1}, x_k]$ contains a rational number where $f(c_k) = 1$. Note that the lengths of the intervals in the partition sum to 1, $\sum_{k=1}^{n} \Delta x_k = 1$. So each such Riemann sum equals 1, and a limit of Riemann sums using these choices equals 1.

On the other hand, if we pick c_k to be the point giving the minimum value for f on $[x_{k-1}, x_k]$, then the Riemann sum is

$$L = \sum_{k=1}^{n} f(c_k)\,\Delta x_k = \sum_{k=1}^{n} (0)\,\Delta x_k = 0,$$

since each subinterval $[x_{k-1}, x_k]$ contains an irrational number c_k where $f(c_k) = 0$. The limit of Riemann sums using these choices equals zero. Since the limit depends on the choices of c_k, the function f is not integrable. ■

Theorem 1 says nothing about how to *calculate* definite integrals. A method of calculation will be developed in Section 5.4, through a connection to knowing an antiderivative of the integrand function f.

Properties of Definite Integrals

In defining $\int_a^b f(x)\,dx$ as a limit of sums $\sum_{k=1}^{n} f(c_k)\,\Delta x_k$, we moved from left to right across the interval $[a, b]$. What would happen if we instead move right to left, starting with $x_0 = b$ and ending at $x_n = a$? Each Δx_k in the Riemann sum would change its sign, with $x_k - x_{k-1}$ now negative instead of positive. With the same choices of c_k in each subinterval, the sign of any Riemann sum would change, as would the sign of the limit, the integral $\int_b^a f(x)\,dx$. Since we have not previously given a meaning to integrating backward, we are led to define

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx.$$

Although we have only defined the integral over an interval $[a, b]$ when $a < b$, it is convenient to have a definition for the integral over $[a, b]$ when $a = b$, that is, for the integral over an interval of zero width. Since $a = b$ gives $\Delta x = 0$, whenever $f(a)$ exists we define

$$\int_a^a f(x)\,dx = 0.$$

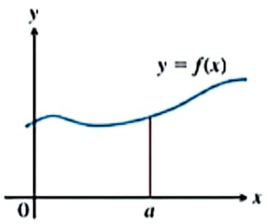
Theorem 2 states basic properties of integrals, given as rules that they satisfy, including the two just discussed. These rules, listed in Table 5.6, become very useful in the process of computing integrals. We will refer to them repeatedly to simplify our calculations. Rules 2 through 7 have geometric interpretations, shown in Figure 5.11. The graphs in these figures are of positive functions, but the rules apply to general integrable functions.

THEOREM 2 When f and g are integrable over the interval $[a, b]$, the definite integral satisfies the rules in Table 5.6.

While Rules 1 and 2 are definitions, Rules 3 to 7 of Table 5.6 must be proved. The following is a proof of Rule 6. Similar proofs can be given to verify the other properties in Table 5.6.

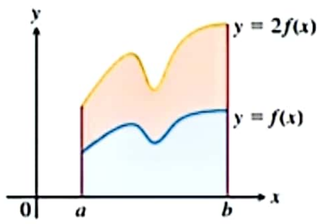
TABLE 5.6 Rules satisfied by definite integrals

1. *Order of Integration:*	$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$	A definition
2. *Zero Width Interval:*	$\int_a^a f(x)\,dx = 0$	A definition when $f(a)$ exists
3. *Constant Multiple:*	$\int_a^b kf(x)\,dx = k\int_a^b f(x)\,dx$	Any constant k
4. *Sum and Difference:*	$\int_a^b (f(x) \pm g(x))\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$	
5. *Additivity:*	$\int_a^b f(x)\,dx + \int_b^c f(x)\,dx = \int_a^c f(x)\,dx$	
6. *Max-Min Inequality:*	If f has maximum value $\max f$ and minimum value $\min f$ on $[a, b]$, then $\min f \cdot (b-a) \le \int_a^b f(x)\,dx \le \max f \cdot (b-a).$	
7. *Domination:*	$f(x) \ge g(x)$ on $[a, b] \Rightarrow \int_a^b f(x)\,dx \ge \int_a^b g(x)\,dx$	
	$f(x) \ge 0$ on $[a, b] \Rightarrow \int_a^b f(x)\,dx \ge 0$	(Special case)



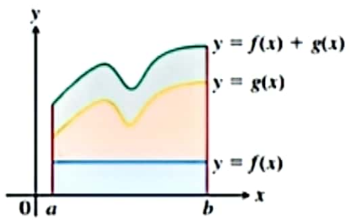
(a) *Zero Width Interval:*

$$\int_a^a f(x)\,dx = 0$$



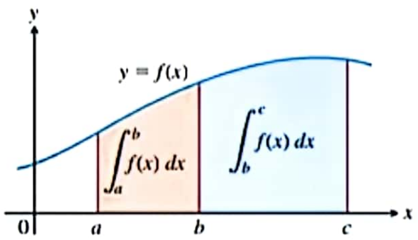
(b) *Constant Multiple:* ($k = 2$)

$$\int_a^b kf(x)\,dx = k\int_a^b f(x)\,dx$$



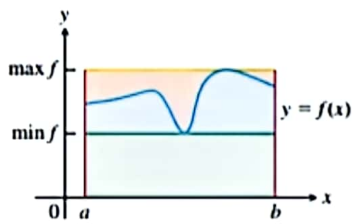
(c) *Sum: (areas add)*

$$\int_a^b (f(x) + g(x))\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$$



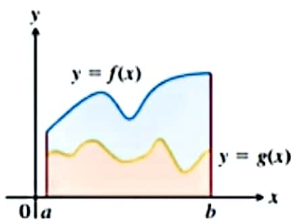
(d) *Additivity for Definite Integrals:*

$$\int_a^b f(x)\,dx + \int_b^c f(x)\,dx = \int_a^c f(x)\,dx$$



(e) *Max-Min Inequality:*

$$\min f \cdot (b-a) \le \int_a^b f(x)\,dx \le \max f \cdot (b-a)$$



(f) *Domination:*

$f(x) \ge g(x)$ on $[a, b]$

$$\Rightarrow \int_a^b f(x)\,dx \ge \int_a^b g(x)\,dx$$

FIGURE 5.11 Geometric interpretations of Rules 2–7 in Table 5.6.

5.6 Definite Integral Substitutions and the Area Between Curves

There are two methods for evaluating a definite integral by substitution. One method is to find an antiderivative using substitution and then to evaluate the definite integral by applying the Evaluation Theorem. The other method extends the process of substitution directly to *definite* integrals by changing the limits of integration. We apply the new formula introduced here to the problem of computing the area between two curves.

The Substitution Formula

The following formula shows how the limits of integration change when the variable of integration is changed by substitution.

THEOREM 7—Substitution in Definite Integrals If g' is continuous on the interval $[a, b]$ and f is continuous on the range of $g(x) = u$, then

$$\int_a^b f(g(x)) \cdot g'(x)\, dx = \int_{g(a)}^{g(b)} f(u)\, du.$$

Proof Let F denote any antiderivative of f. Then,

$$\int_a^b f(g(x)) \cdot g'(x)\, dx = F(g(x))\Big]_{x=a}^{x=b}$$

$\frac{d}{dx}F(g(x)) = F'(g(x))g'(x) = f(g(x))g'(x)$

$$= F(g(b)) - F(g(a))$$

$$= F(u)\Big]_{u=g(a)}^{u=g(b)}$$

$$= \int_{g(a)}^{g(b)} f(u)\, du. \qquad \text{Fundamental Theorem, Part 2} \qquad \blacksquare$$

To use the formula, make the same u-substitution $u = g(x)$ and $du = g'(x)\,dx$ you would use to evaluate the corresponding indefinite integral. Then integrate the transformed integral with respect to u from the value $g(a)$ (the value of u at $x = a$) to the value $g(b)$ (the value of u at $x = b$).

EXAMPLE 1 Evaluate $\displaystyle\int_{-1}^{1} 3x^2\sqrt{x^3 + 1}\, dx$.

Solution We have two choices.

Method 1: Transform the integral and evaluate the transformed integral with the transformed limits given in Theorem 7.

$$\int_{-1}^{1} 3x^2\sqrt{x^3+1}\, dx$$

Let $u = x^3 + 1$, $du = 3x^2\,dx$.
When $x = -1$, $u = (-1)^3 + 1 = 0$
When $x = 1$, $u = (1)^3 + 1 = 2$

$$= \int_0^2 \sqrt{u}\, du$$

$$= \frac{2}{3}u^{3/2}\Big]_0^2 \qquad \text{Evaluate the new definite integral.}$$

$$= \frac{2}{3}\left[2^{3/2} - 0^{3/2}\right] = \frac{2}{3}\left[2\sqrt{2}\right] = \frac{4\sqrt{2}}{3}$$

Method 2: Transform the integral as an indefinite integral, integrate, change back to x, and use the original x-limits.

$$\int 3x^2\sqrt{x^3+1}\,dx = \int \sqrt{u}\,du \qquad \text{Let } u = x^3+1,\ du = 3x^2\,dx.$$

$$= \frac{2}{3}u^{3/2} + C \qquad \text{Integrate with respect to } u.$$

$$= \frac{2}{3}(x^3+1)^{3/2} + C \qquad \text{Replace } u \text{ by } x^3+1.$$

$$\int_{-1}^{1} 3x^2\sqrt{x^3+1}\,dx = \frac{2}{3}(x^3+1)^{3/2}\Big]_{-1}^{1} \qquad \text{Use the integral just found, with limits of integration for } x.$$

$$= \frac{2}{3}\left[((1)^3+1)^{3/2} - ((-1)^3+1)^{3/2}\right]$$

$$= \frac{2}{3}\left[2^{3/2} - 0^{3/2}\right] = \frac{2}{3}\left[2\sqrt{2}\right] = \frac{4\sqrt{2}}{3}$$

Which method is better—evaluating the transformed definite integral with transformed limits using Theorem 7, or transforming the integral, integrating, and transforming back to use the original limits of integration? In Example 1, the first method seems easier, but that is not always the case. Generally, it is best to know both methods and to use whichever one seems better at the time.

EXAMPLE 2 We use the method of transforming the limits of integration.

(a) $$\int_{\pi/4}^{\pi/2} \cot\theta \csc^2\theta\,d\theta = \int_1^0 u\cdot(-du)$$

Let $u = \cot\theta$, $du = -\csc^2\theta\,d\theta$,
$-du = \csc^2\theta\,d\theta$.
When $\theta = \pi/4$, $u = \cot(\pi/4) = 1$.
When $\theta = \pi/2$, $u = \cot(\pi/2) = 0$.

$$= -\int_1^0 u\,du$$

$$= -\left[\frac{u^2}{2}\right]_1^0$$

$$= -\left[\frac{(0)^2}{2} - \frac{(1)^2}{2}\right] = \frac{1}{2}$$

(b) $$\int_{-\pi/4}^{\pi/4} \tan x\,dx = \int_{-\pi/4}^{\pi/4} \frac{\sin x}{\cos x}\,dx$$

$$= -\int_{\sqrt{2}/2}^{\sqrt{2}/2} \frac{du}{u}$$

Let $u = \cos x$, $du = -\sin x\,dx$.
When $x = -\pi/4$, $u = \sqrt{2}/2$.
When $x = \pi/4$, $u = \sqrt{2}/2$.

$$= -\ln|u|\Big]_{\sqrt{2}/2}^{\sqrt{2}/2} = 0 \qquad \text{Integrate, zero width interval}$$

Definite Integrals of Symmetric Functions

The Substitution Formula in Theorem 7 simplifies the calculation of definite integrals of even and odd functions (Section 1.1) over a symmetric interval $[-a, a]$ (Figure 5.23).

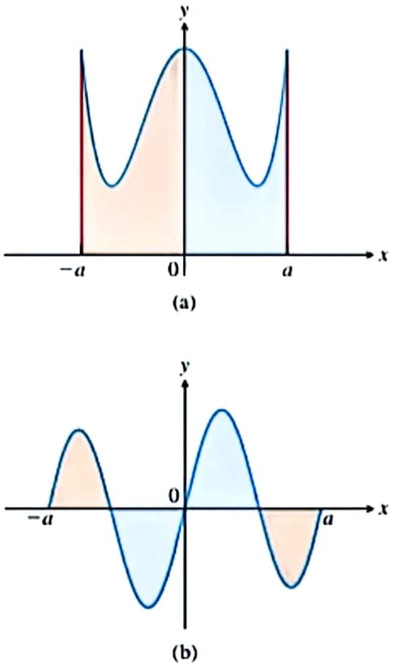


FIGURE 5.23 (a) For f an even function, the integral from $-a$ to a is twice the integral from 0 to a. (b) For f an odd function, the integral from $-a$ to a equals 0.

THEOREM 8 Let f be continuous on the symmetric interval $[-a, a]$.

(a) If f is even, then $\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx$.

(b) If f is odd, then $\int_{-a}^{a} f(x)\,dx = 0$.

Proof of Part (a)

$$
\begin{aligned}
\int_{-a}^{a} f(x)\,dx &= \int_{-a}^{0} f(x)\,dx + \int_{0}^{a} f(x)\,dx && \text{Additivity Rule for Definite Integrals} \\
&= -\int_{0}^{-a} f(x)\,dx + \int_{0}^{a} f(x)\,dx && \text{Order of Integration Rule} \\
&= -\int_{0}^{a} f(-u)(-du) + \int_{0}^{a} f(x)\,dx && \text{Let } u = -x,\ du = -dx. \text{ When } x = 0,\ u = 0. \text{ When } x = -a,\ u = a. \\
&= \int_{0}^{a} f(-u)\,du + \int_{0}^{a} f(x)\,dx \\
&= \int_{0}^{a} f(u)\,du + \int_{0}^{a} f(x)\,dx && f \text{ is even, so } f(-u) = f(u). \\
&= 2\int_{0}^{a} f(x)\,dx
\end{aligned}
$$

The proof of part (b) is entirely similar and you are asked to give it in Exercise 114. ■

The assertions of Theorem 8 remain true when f is an integrable function (rather than having the stronger property of being continuous).

EXAMPLE 3 Evaluate $\int_{-2}^{2} (x^4 - 4x^2 + 6)\,dx$.

Solution Since $f(x) = x^4 - 4x^2 + 6$ satisfies $f(-x) = f(x)$, it is even on the symmetric interval $[-2, 2]$, so

$$
\begin{aligned}
\int_{-2}^{2} (x^4 - 4x^2 + 6)\,dx &= 2\int_{0}^{2} (x^4 - 4x^2 + 6)\,dx \\
&= 2\left[\frac{x^5}{5} - \frac{4}{3}x^3 + 6x\right]_0^2 \\
&= 2\left(\frac{32}{5} - \frac{32}{3} + 12\right) = \frac{232}{15}.
\end{aligned}
$$
■

Areas Between Curves

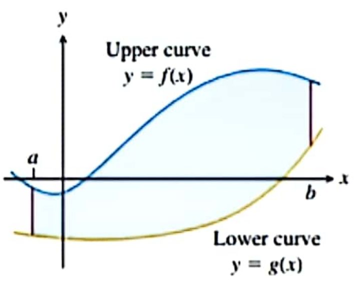


FIGURE 5.24 The region between the curves $y = f(x)$ and $y = g(x)$ and the lines $x = a$ and $x = b$.

Suppose we want to find the area of a region that is bounded above by the curve $y = f(x)$, below by the curve $y = g(x)$, and on the left and right by the lines $x = a$ and $x = b$ (Figure 5.24). The region might accidentally have a shape whose area we could find with geometry, but if f and g are arbitrary continuous functions, we usually have to find the area with an integral.

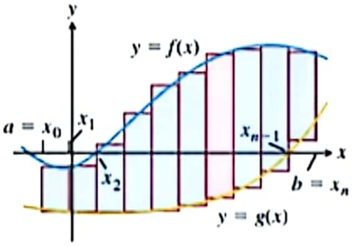


FIGURE 5.25 We approximate the region with rectangles perpendicular to the x-axis.

To see what the integral should be, we first approximate the region with n vertical rectangles based on a partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ (Figure 5.25). The area of the kth rectangle (Figure 5.26) is

$$\Delta A_k = \text{height} \times \text{width} = [f(c_k) - g(c_k)]\,\Delta x_k.$$

We then approximate the area of the region by adding the areas of the n rectangles:

$$A \approx \sum_{k=1}^{n} \Delta A_k = \sum_{k=1}^{n} [f(c_k) - g(c_k)]\,\Delta x_k. \qquad \text{Riemann sum}$$

As $\|P\| \to 0$, the sums on the right approach the limit $\int_a^b [f(x) - g(x)]\,dx$ because f and g are continuous. We take the area of the region to be the value of this integral. That is,

$$A = \lim_{\|P\|\to 0} \sum_{k=1}^{n} [f(c_k) - g(c_k)]\,\Delta x_k = \int_a^b [f(x) - g(x)]\,dx.$$

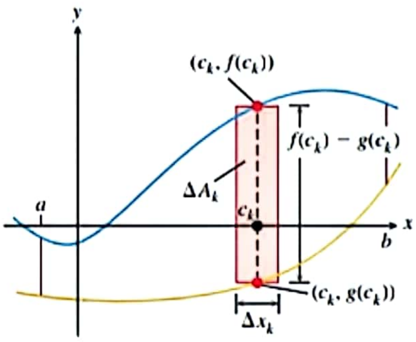


FIGURE 5.26 The area ΔA_k of the kth rectangle is the product of its height, $f(c_k) - g(c_k)$, and its width, Δx_k.

DEFINITION If f and g are continuous with $f(x) \geq g(x)$ throughout $[a, b]$, then the **area of the region between the curves $y = f(x)$ and $y = g(x)$ from a to b** is the integral of $(f - g)$ from a to b:

$$A = \int_a^b [f(x) - g(x)]\,dx.$$

When applying this definition it is helpful to graph the curves. The graph reveals which curve is the upper curve f and which is the lower curve g. It also helps you find the limits of integration if they are not given. You may need to find where the curves intersect to determine the limits of integration, and this may involve solving the equation $f(x) = g(x)$ for values of x. Then you can integrate the function $f - g$ for the area between the intersections.

EXAMPLE 4 Find the area of the region bounded above by the curve $y = 2e^{-x} + x$, below by the curve $y = e^x/2$, on the left by $x = 0$, and on the right by $x = 1$.

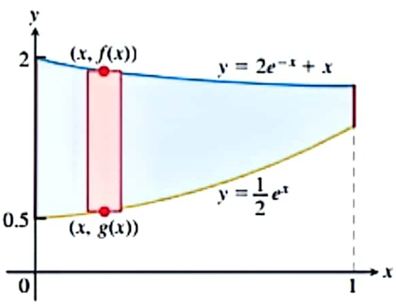


FIGURE 5.27 The region in Example 4 with a typical approximating rectangle.

Solution Figure 5.27 displays the graphs of the curves and the region whose area we want to find. The area between the curves over the interval $0 \leq x \leq 1$ is given by

$$A = \int_0^1 \left[(2e^{-x} + x) - \frac{1}{2}e^x\right] dx = \left[-2e^{-x} + \frac{1}{2}x^2 - \frac{1}{2}e^x\right]_0^1$$

$$= \left(-2e^{-1} + \frac{1}{2} - \frac{1}{2}e\right) - \left(-2 + 0 - \frac{1}{2}\right)$$

$$= 3 - \frac{2}{e} - \frac{e}{2} \approx 0.9051.$$

EXAMPLE 5 Find the area of the region enclosed by the parabola $y = 2 - x^2$ and the line $y = -x$.

Solution First we sketch the two curves (Figure 5.28). The limits of integration are found by solving $y = 2 - x^2$ and $y = -x$ simultaneously for x.

$$\begin{aligned} 2 - x^2 &= -x && \text{Equate } f(x) \text{ and } g(x). \\ x^2 - x - 2 &= 0 && \text{Rewrite.} \\ (x + 1)(x - 2) &= 0 && \text{Factor.} \\ x = -1, \qquad x &= 2. && \text{Solve.} \end{aligned}$$

The region runs from $x = -1$ to $x = 2$. The limits of integration are $a = -1$, $b = 2$.

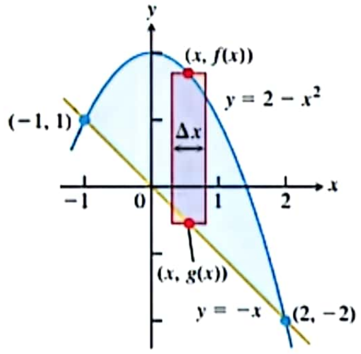


FIGURE 5.28 The region in Example 5 with a typical approximating rectangle.

The area between the curves is

$$A = \int_a^b [f(x) - g(x)]\,dx = \int_{-1}^{2} [(2 - x^2) - (-x)]\,dx$$

$$= \int_{-1}^{2} (2 + x - x^2)\,dx = \left[2x + \frac{x^2}{2} - \frac{x^3}{3}\right]_{-1}^{2}$$

$$= \left(4 + \frac{4}{2} - \frac{8}{3}\right) - \left(-2 + \frac{1}{2} + \frac{1}{3}\right) = \frac{9}{2}.$$

If the formula for a bounding curve changes at one or more points, we subdivide the region into subregions that correspond to the formula changes and apply the formula for the area between curves to each subregion.

EXAMPLE 6 Find the area of the region in the first quadrant that is bounded above by $y = \sqrt{x}$ and below by the x-axis and the line $y = x - 2$.

Solution The sketch (Figure 5.29) shows that the region's upper boundary is the graph of $f(x) = \sqrt{x}$. The lower boundary changes from $g(x) = 0$ for $0 \le x \le 2$ to $g(x) = x - 2$ for $2 \le x \le 4$ (both formulas agree at $x = 2$). We subdivide the region at $x = 2$ into subregions A and B, shown in Figure 5.29.

The limits of integration for region A are $a = 0$ and $b = 2$. The left-hand limit for region B is $a = 2$. To find the right-hand limit, we solve the equations $y = \sqrt{x}$ and $y = x - 2$ simultaneously for x:

$\sqrt{x} = x - 2$	Equate $f(x)$ and $g(x)$.
$x = (x - 2)^2 = x^2 - 4x + 4$	Square both sides.
$x^2 - 5x + 4 = 0$	Rewrite.
$(x - 1)(x - 4) = 0$	Factor.
$x = 1, \quad x = 4.$	Solve.

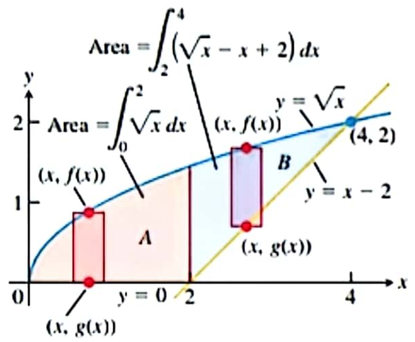


FIGURE 5.29 When the formula for a bounding curve changes, the area integral changes to become the sum of integrals to match, one integral for each of the shaded regions shown here for Example 6.

Only the value $x = 4$ satisfies the equation $\sqrt{x} = x - 2$. The value $x = 1$ is an extraneous root introduced by squaring. The right-hand limit is $b = 4$.

For $0 \le x \le 2$: $\quad f(x) - g(x) = \sqrt{x} - 0 = \sqrt{x}$

For $2 \le x \le 4$: $\quad f(x) - g(x) = \sqrt{x} - (x - 2) = \sqrt{x} - x + 2$

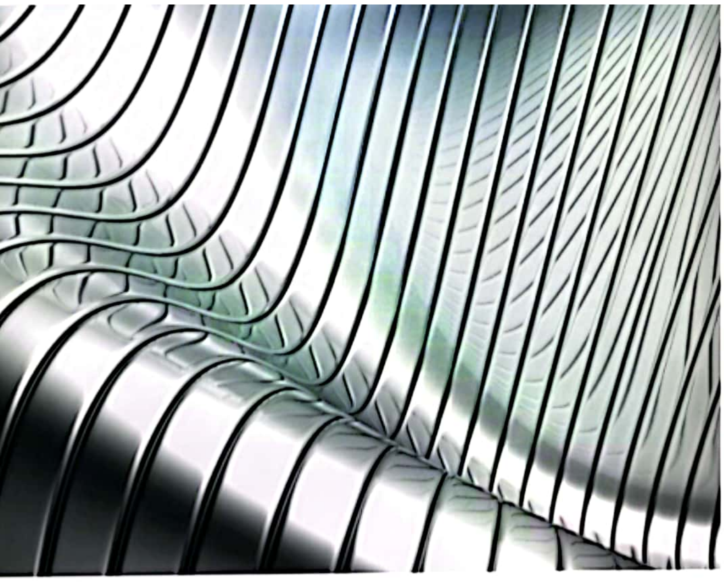
We add the areas of subregions A and B to find the total area:

$$\text{Total area} = \underbrace{\int_0^2 \sqrt{x}\,dx}_{\text{area of } A} + \underbrace{\int_2^4 (\sqrt{x} - x + 2)\,dx}_{\text{area of } B}$$

$$= \left[\frac{2}{3}x^{3/2}\right]_0^2 + \left[\frac{2}{3}x^{3/2} - \frac{x^2}{2} + 2x\right]_2^4$$

$$= \frac{2}{3}(2)^{3/2} - 0 + \left(\frac{2}{3}(4)^{3/2} - 8 + 8\right) - \left(\frac{2}{3}(2)^{3/2} - 2 + 4\right)$$

$$= \frac{2}{3}(8) - 2 = \frac{10}{3}.$$



6 Applications of Definite Integrals

OVERVIEW In Chapter 5 we saw that a continuous function over a closed interval has a definite integral, which is the limit of any Riemann sum for the function. We proved that we could evaluate definite integrals using the Fundamental Theorem of Calculus. We also found that the area under a curve and the area between two curves could be defined and computed as definite integrals.

In this chapter we extend the applications of definite integrals to defining and finding volumes, lengths of plane curves, and areas of surfaces of revolution. We also use integrals to solve physical problems involving the work done by a force, and to find the location of an object's center of mass. Each application comes from a process leading to an approximation by a Riemann sum, and then taking a limit to obtain an appropriate definite integral. These applications are important to mathematics, science, and engineering. We also use integrals to compute probabilities and their applications to the life sciences in Chapter 8.

6.1 Volumes Using Cross-Sections

In this section we define volumes of solids using the areas of their cross-sections. A **cross-section** of a solid S is the plane region formed by intersecting S with a plane (Figure 6.1). We present three different methods for obtaining the cross-sections appropriate to finding the volume of a particular solid: the method of slicing, the disk method, and the washer method.

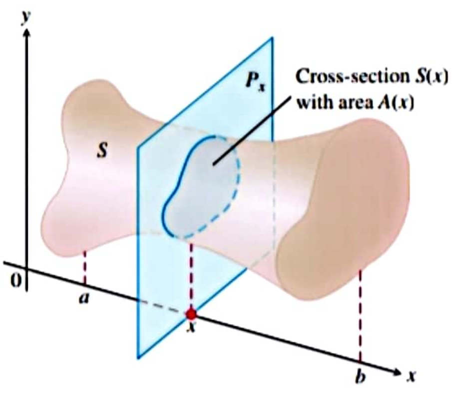


FIGURE 6.1 A cross-section $S(x)$ of the solid S formed by intersecting S with a plane P_x perpendicular to the x-axis through the point x in the interval $[a, b]$.

Suppose we want to find the volume of a solid S like the one in Figure 6.1. We begin by extending the definition of a cylinder from classical geometry to cylindrical solids with arbitrary bases (Figure 6.2). If the cylindrical solid has a known base area A and height h, then the volume of the cylindrical solid is

$$\text{Volume} = \text{area} \times \text{height} = A \cdot h.$$

This equation forms the basis for defining the volumes of many solids that are not cylinders, like the one in Figure 6.1. If the cross-section of the solid S at each point x in the interval $[a, b]$ is a region $S(x)$ of area $A(x)$, and A is a continuous function of x, we can

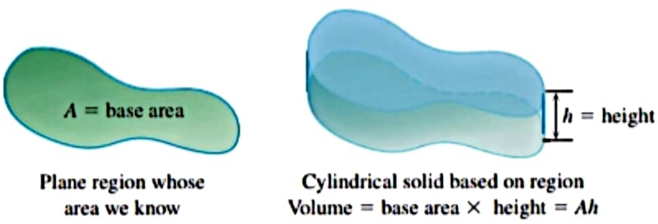


FIGURE 6.2 The volume of a cylindrical solid is always defined to be its base area times its height.

define and calculate the volume of the solid S as the definite integral of $A(x)$. We now show how this integral is obtained by the **method of slicing**.

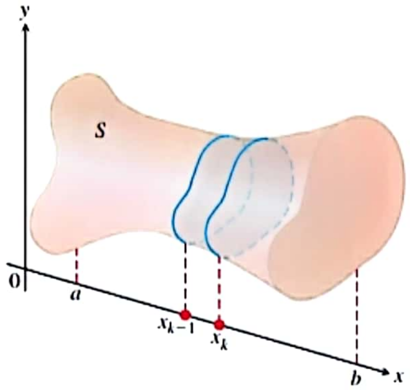


FIGURE 6.3 A typical thin slab in the solid S.

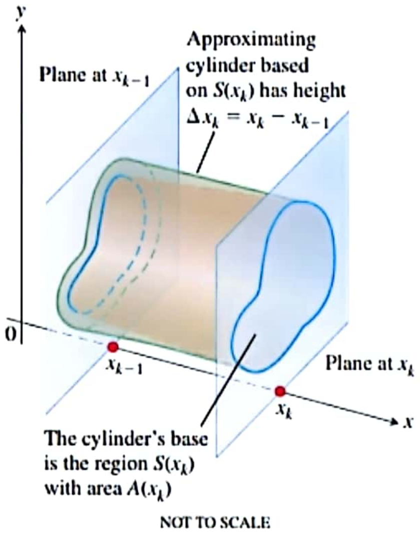


FIGURE 6.4 The solid thin slab in Figure 6.3 is shown enlarged here. It is approximated by the cylindrical solid with base $S(x_k)$ having area $A(x_k)$ and height $\Delta x_k = x_k - x_{k-1}$.

Slicing by Parallel Planes

We partition $[a, b]$ into subintervals of width (length) Δx_k and slice the solid, as we would a loaf of bread, by planes perpendicular to the x-axis at the partition points $a = x_0 < x_1 < \cdots < x_n = b$. The planes P_{x_k}, perpendicular to the x-axis at the partition points, slice S into thin "slabs" (like thin slices of a loaf of bread). A typical slab is shown in Figure 6.3. We approximate the slab between the plane at x_{k-1} and the plane at x_k by a cylindrical solid with base area $A(x_k)$ and height $\Delta x_k = x_k - x_{k-1}$ (Figure 6.4). The volume V_k of this cylindrical solid is $A(x_k) \cdot \Delta x_k$, which is approximately the same volume as that of the slab:

$$\text{Volume of the } k\text{th slab} \approx V_k = A(x_k)\,\Delta x_k.$$

The volume V of the entire solid S is therefore approximated by the sum of these cylindrical volumes,

$$V \approx \sum_{k=1}^{n} V_k = \sum_{k=1}^{n} A(x_k)\,\Delta x_k.$$

This is a Riemann sum for the function $A(x)$ on $[a, b]$. We expect the approximations from these sums to improve as the norm of the partition of $[a, b]$ goes to zero. Taking a partition of $[a, b]$ into n subintervals with $\|P\| \to 0$ gives

$$\lim_{n\to\infty} \sum_{k=1}^{n} A(x_k)\,\Delta x_k = \int_a^b A(x)\,dx.$$

So we define the limiting definite integral of the Riemann sum to be the volume of the solid S.

DEFINITION The **volume** of a solid of integrable cross-sectional area $A(x)$ from $x = a$ to $x = b$ is the integral of A from a to b,

$$V = \int_a^b A(x)\,dx.$$

This definition applies whenever $A(x)$ is integrable, and in particular when it is continuous. To apply the definition to calculate the volume of a solid using cross-sections perpendicular to the x-axis, take the following steps:

Calculating the Volume of a Solid

1. *Sketch the solid and a typical cross-section.*
2. *Find a formula for* $A(x)$, the area of a typical cross-section.
3. *Find the limits of integration.*
4. *Integrate* $A(x)$ to find the volume.

EXAMPLE 1 A pyramid 3 m high has a square base that is 3 m on a side. The cross-section of the pyramid perpendicular to the altitude x m down from the vertex is a square x m on a side. Find the volume of the pyramid.

Solution

1. *A sketch.* We draw the pyramid with its altitude along the x-axis and its vertex at the origin and include a typical cross-section (Figure 6.5).

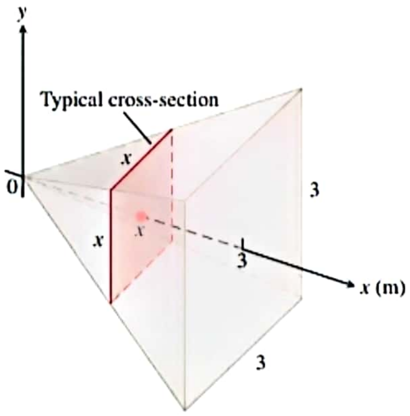


FIGURE 6.5 The cross-sections of the pyramid in Example 1 are squares.

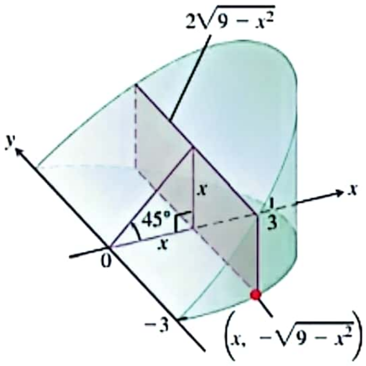


FIGURE 6.6 The wedge of Example 2, sliced perpendicular to the x-axis. The cross-sections are rectangles.

HISTORICAL BIOGRAPHY
Bonaventura Cavalieri
(1598–1647)

2. *A formula for* $A(x)$. The cross-section at x is a square x meters on a side, so its area is

$$A(x) = x^2.$$

3. *The limits of integration*. The squares lie on the planes from $x = 0$ to $x = 3$.
4. *Integrate to find the volume*:

$$V = \int_0^3 A(x)\,dx = \int_0^3 x^2\,dx = \frac{x^3}{3}\bigg]_0^3 = 9 \text{ m}^3.$$

EXAMPLE 2 A curved wedge is cut from a circular cylinder of radius 3 by two planes. One plane is perpendicular to the axis of the cylinder. The second plane crosses the first plane at a 45° angle at the center of the cylinder. Find the volume of the wedge.

Solution We draw the wedge and sketch a typical cross-section perpendicular to the x-axis (Figure 6.6). The base of the wedge in the figure is the semicircle with $x \geq 0$ that is cut from the circle $x^2 + y^2 = 9$ by the 45° plane when it intersects the y-axis. For any x in the interval $[0, 3]$, the y-values in this semicircular base vary from $y = -\sqrt{9 - x^2}$ to $y = \sqrt{9 - x^2}$. When we slice through the wedge by a plane perpendicular to the x-axis, we obtain a cross-section at x which is a rectangle of height x whose width extends across the semicircular base. The area of this cross-section is

$$\begin{aligned} A(x) &= (\text{height})(\text{width}) = (x)\left(2\sqrt{9 - x^2}\right) \\ &= 2x\sqrt{9 - x^2}. \end{aligned}$$

The rectangles run from $x = 0$ to $x = 3$, so we have

$$\begin{aligned} V &= \int_a^b A(x)\,dx = \int_0^3 2x\sqrt{9 - x^2}\,dx \\ &= -\frac{2}{3}(9 - x^2)^{3/2}\bigg]_0^3 \qquad \text{Let } u = 9 - x^2, \; du = -2x\,dx, \text{ integrate, and substitute back.} \\ &= 0 + \frac{2}{3}(9)^{3/2} \\ &= 18. \end{aligned}$$

EXAMPLE 3 Cavalieri's principle says that solids with equal altitudes and identical cross-sectional areas at each height have the same volume (Figure 6.7). This follows immediately from the definition of volume, because the cross-sectional area function $A(x)$ and the interval $[a, b]$ are the same for both solids.

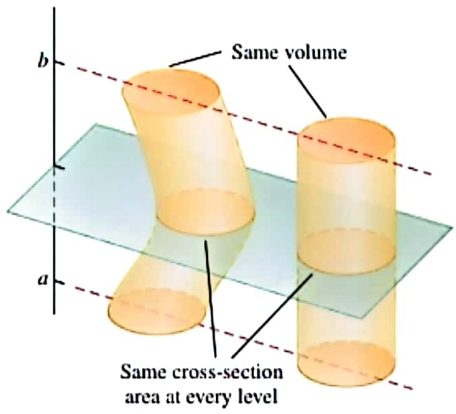
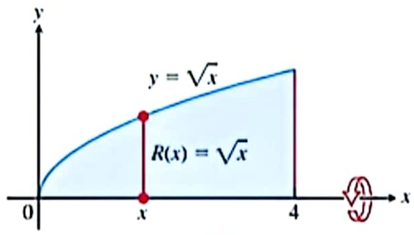


FIGURE 6.7 *Cavalieri's principle:* These solids have the same volume, which can be illustrated with stacks of coins.



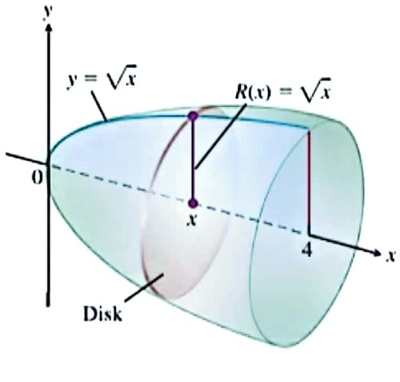


FIGURE 6.8 The region (a) and solid of revolution (b) in Example 4.

Solids of Revolution: The Disk Method

The solid generated by rotating (or revolving) a plane region about an axis in its plane is called a **solid of revolution**. To find the volume of a solid like the one shown in Figure 6.8, we need only observe that the cross-sectional area $A(x)$ is the area of a disk of radius $R(x)$, the distance of the planar region's boundary from the axis of revolution. The area is then

$$A(x) = \pi(\text{radius})^2 = \pi[R(x)]^2.$$

So the definition of volume in this case gives

Volume by Disks for Rotation About the x-axis

$$V = \int_a^b A(x)\,dx = \int_a^b \pi[R(x)]^2\,dx.$$

This method for calculating the volume of a solid of revolution is often called the **disk method** because a cross-section is a circular disk of radius $R(x)$.

EXAMPLE 4 The region between the curve $y = \sqrt{x}$, $0 \le x \le 4$, and the x-axis is revolved about the x-axis to generate a solid. Find its volume.

Solution We draw figures showing the region, a typical radius, and the generated solid (Figure 6.8). The volume is

$$\begin{aligned} V &= \int_a^b \pi[R(x)]^2\,dx \\ &= \int_0^4 \pi\left[\sqrt{x}\right]^2 dx \qquad \text{Radius } R(x) = \sqrt{x} \text{ for rotation around } x\text{-axis.} \\ &= \pi\int_0^4 x\,dx = \pi\frac{x^2}{2}\bigg]_0^4 = \pi\frac{(4)^2}{2} = 8\pi. \end{aligned}$$

EXAMPLE 5 The circle

$$x^2 + y^2 = a^2$$

is rotated about the x-axis to generate a sphere. Find its volume.

Solution We imagine the sphere cut into thin slices by planes perpendicular to the x-axis (Figure 6.9). The cross-sectional area at a typical point x between $-a$ and a is

$$A(x) = \pi y^2 = \pi(a^2 - x^2). \qquad R(x) = \sqrt{a^2 - x^2} \text{ for rotation around } x\text{-axis.}$$

Therefore, the volume is

$$V = \int_{-a}^{a} A(x)\,dx = \int_{-a}^{a} \pi(a^2 - x^2)\,dx = \pi\left[a^2x - \frac{x^3}{3}\right]_{-a}^{a} = \frac{4}{3}\pi a^3.$$

The axis of revolution in the next example is not the x-axis, but the rule for calculating the volume is the same: Integrate $\pi(\text{radius})^2$ between appropriate limits.

EXAMPLE 6 Find the volume of the solid generated by revolving the region bounded by $y = \sqrt{x}$ and the lines $y = 1$, $x = 4$ about the line $y = 1$.

Volume by Disks for Rotation About the y-axis

$$V = \int_c^d A(y)\,dy = \int_c^d \pi[R(y)]^2\,dy.$$

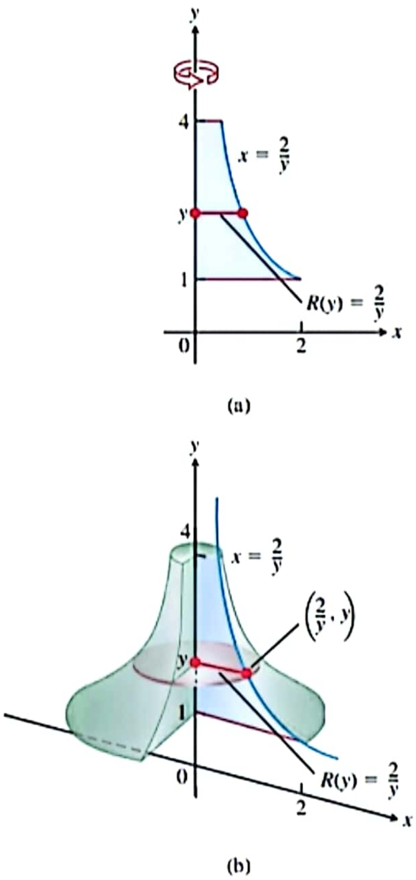


FIGURE 6.11 The region (a) and part of the solid of revolution (b) in Example 7.

EXAMPLE 7 Find the volume of the solid generated by revolving the region between the y-axis and the curve $x = 2/y$, $1 \le y \le 4$, about the y-axis.

Solution We draw figures showing the region, a typical radius, and the generated solid (Figure 6.11). The volume is

$$V = \int_1^4 \pi[R(y)]^2\,dy$$

$$= \int_1^4 \pi\left(\frac{2}{y}\right)^2 dy \qquad \text{Radius } R(y) = \frac{2}{y} \text{ for rotation around } y\text{-axis.}$$

$$= \pi\int_1^4 \frac{4}{y^2}\,dy = 4\pi\left[-\frac{1}{y}\right]_1^4 = 4\pi\left[\frac{3}{4}\right] = 3\pi.$$

EXAMPLE 8 Find the volume of the solid generated by revolving the region between the parabola $x = y^2 + 1$ and the line $x = 3$ about the line $x = 3$.

Solution We draw figures showing the region, a typical radius, and the generated solid (Figure 6.12). Note that the cross-sections are perpendicular to the line $x = 3$ and have y-coordinates from $y = -\sqrt{2}$ to $y = \sqrt{2}$. The volume is

$$V = \int_{-\sqrt{2}}^{\sqrt{2}} \pi[R(y)]^2\,dy \qquad y = \pm\sqrt{2} \text{ when } x = 3$$

$$= \int_{-\sqrt{2}}^{\sqrt{2}} \pi[2 - y^2]^2\,dy \qquad \text{Radius } R(y) = 3 - (y^2 + 1) \text{ for rotation around axis } x = 3.$$

$$= \pi\int_{-\sqrt{2}}^{\sqrt{2}} [4 - 4y^2 + y^4]\,dy \qquad \text{Expand integrand.}$$

$$= \pi\left[4y - \frac{4}{3}y^3 + \frac{y^5}{5}\right]_{-\sqrt{2}}^{\sqrt{2}} \qquad \text{Integrate.}$$

$$= \frac{64\pi\sqrt{2}}{15}.$$

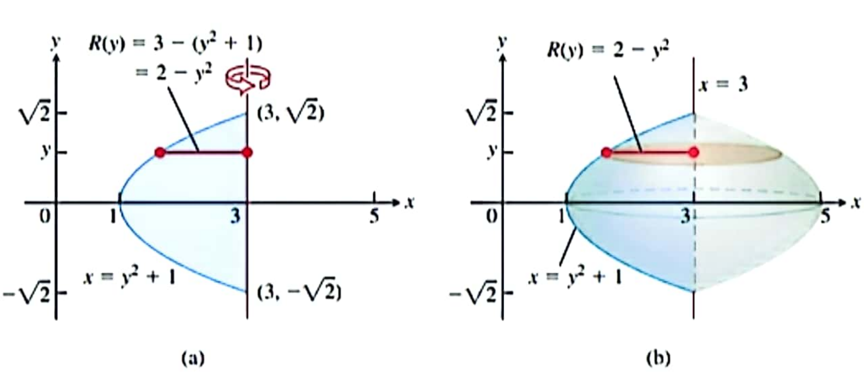


FIGURE 6.12 The region (a) and solid of revolution (b) in Example 8.

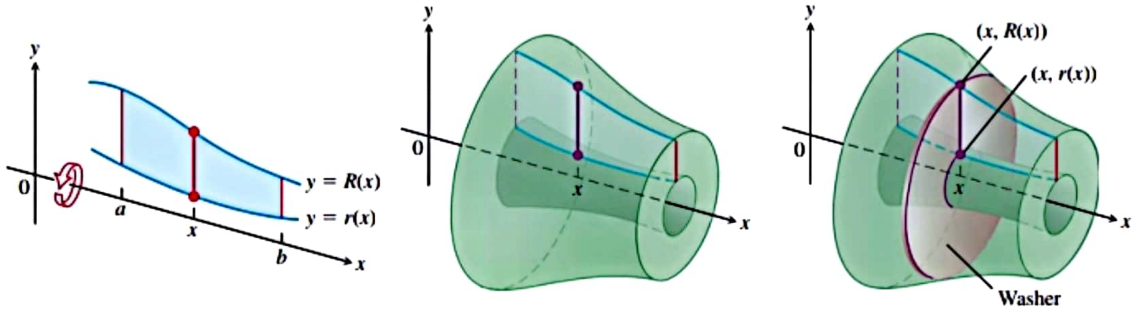


FIGURE 6.13 The cross-sections of the solid of revolution generated here are washers, not disks, so the integral $\int_a^b A(x)\,dx$ leads to a slightly different formula.

Solids of Revolution: The Washer Method

If the region we revolve to generate a solid does not border on or cross the axis of revolution, the solid has a hole in it (Figure 6.13). The cross-sections perpendicular to the axis of revolution are *washers* (the purplish circular surface in Figure 6.13) instead of disks. The dimensions of a typical washer are

Outer radius: $R(x)$

Inner radius: $r(x)$

The washer's area is

$$A(x) = \pi[R(x)]^2 - \pi[r(x)]^2 = \pi([R(x)]^2 - [r(x)]^2).$$

Consequently, the definition of volume in this case gives

Volume by Washers for Rotation About the x-axis

$$V = \int_a^b A(x)\,dx = \int_a^b \pi([R(x)]^2 - [r(x)]^2)\,dx.$$

This method for calculating the volume of a solid of revolution is called the **washer method** because a thin slab of the solid resembles a circular washer of outer radius $R(x)$ and inner radius $r(x)$.

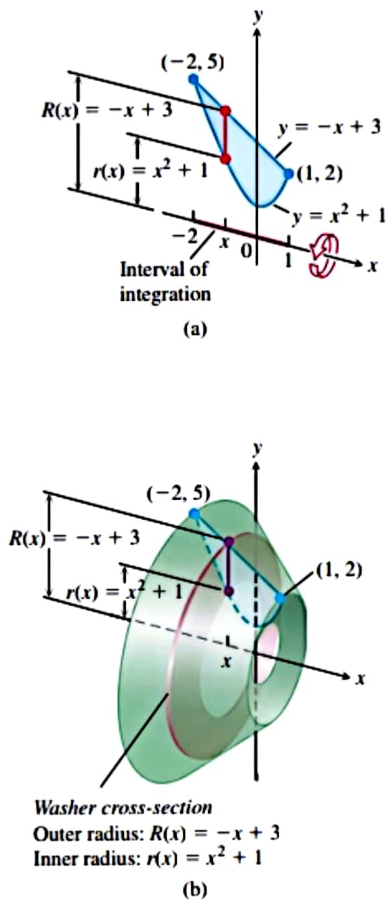


FIGURE 6.14 (a) The region in Example 9 spanned by a line segment perpendicular to the axis of revolution. (b) When the region is revolved about the x-axis, the line segment generates a washer.

EXAMPLE 9 The region bounded by the curve $y = x^2 + 1$ and the line $y = -x + 3$ is revolved about the x-axis to generate a solid. Find the volume of the solid.

Solution We use the four steps for calculating the volume of a solid as discussed early in this section.

1. Draw the region and sketch a line segment across it perpendicular to the axis of revolution (the red segment in Figure 6.14a).
2. Find the outer and inner radii of the washer that would be swept out by the line segment if it were revolved about the x-axis along with the region.

   These radii are the distances of the ends of the line segment from the axis of revolution (Figure 6.14).

Outer radius: $R(x) = -x + 3$

Inner radius: $r(x) = x^2 + 1$

3. Find the limits of integration by finding the x-coordinates of the intersection points of the curve and line in Figure 6.14a.

$$\begin{aligned} x^2 + 1 &= -x + 3 \\ x^2 + x - 2 &= 0 \\ (x + 2)(x - 1) &= 0 \\ x = -2, \quad x &= 1 \end{aligned}$$ Limits of integration

4. Evaluate the volume integral.

$$V = \int_a^b \pi([R(x)]^2 - [r(x)]^2)\,dx$$ Rotation around x-axis

$$= \int_{-2}^{1} \pi((-x + 3)^2 - (x^2 + 1)^2)\,dx$$ Values from Steps 2 and 3

$$= \pi \int_{-2}^{1} (8 - 6x - x^2 - x^4)\,dx$$ Simplify algebraically.

$$= \pi \left[8x - 3x^2 - \frac{x^3}{3} - \frac{x^5}{5} \right]_{-2}^{1} = \frac{117\pi}{5}$$ Integrate. ■

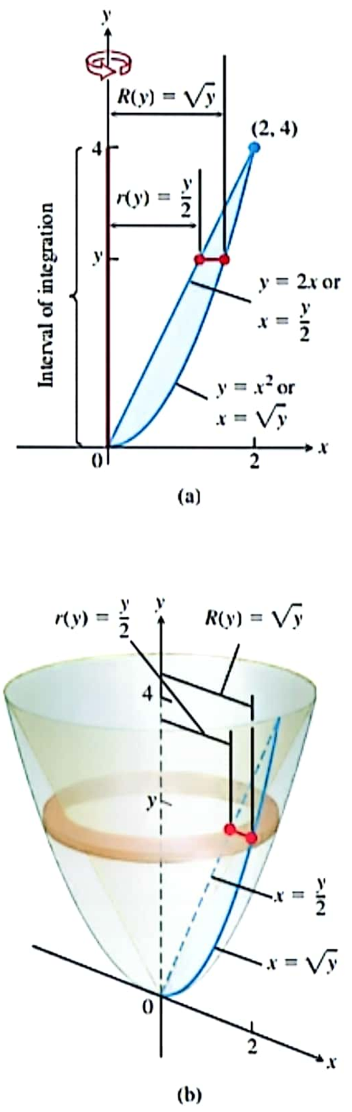


FIGURE 6.15 (a) The region being rotated about the y-axis, the washer radii, and limits of integration in Example 10. (b) The washer swept out by the line segment in part (a).

To find the volume of a solid formed by revolving a region about the y-axis, we use the same procedure as in Example 9, but integrate with respect to y instead of x. In this situation the line segment sweeping out a typical washer is perpendicular to the y-axis (the axis of revolution), and the outer and inner radii of the washer are functions of y.

EXAMPLE 10 The region bounded by the parabola $y = x^2$ and the line $y = 2x$ in the first quadrant is revolved about the y-axis to generate a solid. Find the volume of the solid.

Solution First we sketch the region and draw a line segment across it perpendicular to the axis of revolution (the y-axis). See Figure 6.15a.

The radii of the washer swept out by the line segment are $R(y) = \sqrt{y}$, $r(y) = y/2$ (Figure 6.15).

The line and parabola intersect at $y = 0$ and $y = 4$, so the limits of integration are $c = 0$ and $d = 4$. We integrate to find the volume:

$$V = \int_c^d \pi([R(y)]^2 - [r(y)]^2)\,dy$$ Rotation around y-axis

$$= \int_0^4 \pi\left([\sqrt{y}]^2 - \left[\frac{y}{2}\right]^2 \right) dy$$ Substitute for radii and limits of integration.

$$= \pi \int_0^4 \left(y - \frac{y^2}{4} \right) dy = \pi \left[\frac{y^2}{2} - \frac{y^3}{12} \right]_0^4 = \frac{8}{3}\pi.$$ ■

6.2 Volumes Using Cylindrical Shells

In Section 6.1 we defined the volume of a solid as the definite integral $V = \int_a^b A(x)\,dx$, where $A(x)$ is an integrable cross-sectional area of the solid from $x = a$ to $x = b$. The area $A(x)$ was obtained by slicing through the solid with a plane perpendicular to the x-axis. However, this method of slicing is sometimes awkward to apply, as we will illustrate in our first example. To overcome this difficulty, we use the same integral definition for volume, but obtain the area by slicing through the solid in a different way.

Slicing with Cylinders

Suppose we slice through the solid using circular cylinders of increasing radii, like cookie cutters. We slice straight down through the solid so that the axis of each cylinder is parallel to the y-axis. The vertical axis of each cylinder is the same line, but the radii of the cylinders increase with each slice. In this way the solid is sliced up into thin cylindrical shells of constant thickness that grow outward from their common axis, like circular tree rings. Unrolling a cylindrical shell shows that its volume is approximately that of a rectangular slab with area $A(x)$ and thickness Δx. This slab interpretation allows us to apply the same integral definition for volume as before. The following example provides some insight before we derive the general method.

EXAMPLE 1 The region enclosed by the x-axis and the parabola $y = f(x) = 3x - x^2$ is revolved about the vertical line $x = -1$ to generate a solid (Figure 6.16). Find the volume of the solid.

Solution Using the washer method from Section 6.1 would be awkward here because we would need to express the x-values of the left and right sides of the parabola in Figure 6.16a in terms of y. (These x-values are the inner and outer radii for a typical washer, requiring us to solve $y = 3x - x^2$ for x, which leads to complicated formulas.) Instead of rotating a horizontal strip of thickness Δy, we rotate a *vertical strip* of thickness Δx. This rotation produces a *cylindrical shell* of height y_k above a point x_k within the base of the vertical strip and of thickness Δx. An example of a cylindrical shell is shown as the orange-shaded region in Figure 6.17. We can think of the cylindrical shell shown in the figure as approximating a slice of the solid obtained by cutting straight down through it, parallel to the axis of revolution, all the way around close to the inside hole. We then cut another cylindrical slice around the enlarged hole, then another, and so on, obtaining n cylinders. The radii of the cylinders gradually increase, and the heights of the cylinders follow the contour of the parabola: shorter to taller, then back to shorter (Figure 6.16a).

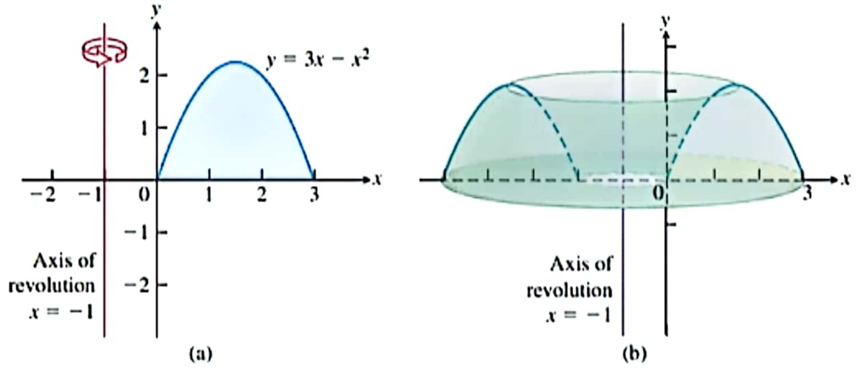


FIGURE 6.16 (a) The graph of the region in Example 1, before revolution. (b) The solid formed when the region in part (a) is revolved about the axis of revolution $x = -1$.

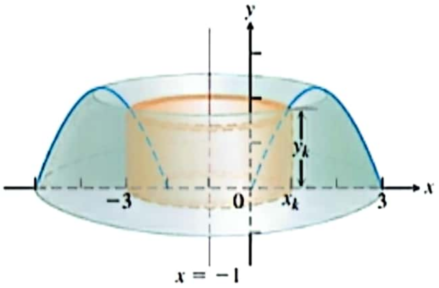


FIGURE 6.17 A cylindrical shell of height y_k obtained by rotating a vertical strip of thickness Δx_k about the line $x = -1$. The outer radius of the cylinder occurs at x_k, where the height of the parabola is $y_k = 3x_k - x_k^2$ (Example 1).

Each slice is sitting over a subinterval of the x-axis of length (width) Δx_k. Its radius is approximately $(1 + x_k)$, and its height is approximately $3x_k - x_k^2$. If we unroll the cylinder at x_k and flatten it out, it becomes (approximately) a rectangular slab with thickness Δx_k (Figure 6.18). The outer circumference of the kth cylinder is $2\pi \cdot \text{radius} = 2\pi(1 + x_k)$, and this is the length of the rolled-out rectangular slab. Its volume is approximated by that of a rectangular solid, the area of the rectangle times its thickness,

$$\begin{aligned}\Delta V_k &= \text{circumference} \times \text{height} \times \text{thickness}\\ &= 2\pi(1 + x_k) \cdot \left(3x_k - x_k^2\right) \cdot \Delta x_k.\end{aligned}$$

Summing together the volumes ΔV_k of the individual cylindrical shells over the interval $[0, 3]$ gives the Riemann sum

$$\sum_{k=1}^{n} \Delta V_k = \sum_{k=1}^{n} 2\pi(x_k + 1)\left(3x_k - x_k^2\right)\Delta x_k.$$

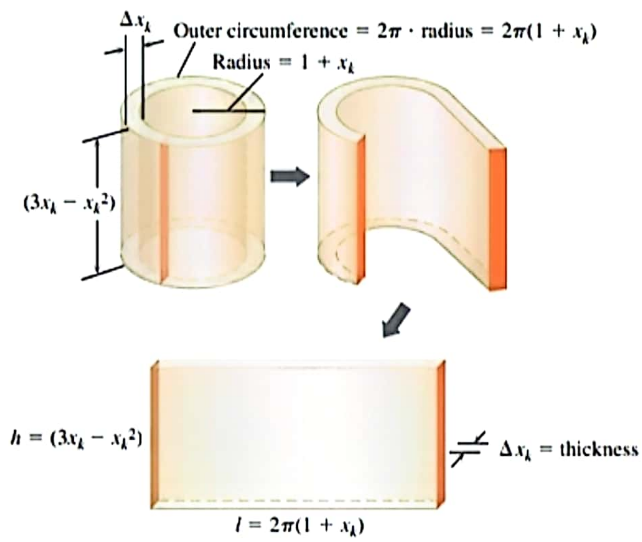


FIGURE 6.18 Cutting and unrolling a cylindrical shell gives a nearly rectangular solid (Example 1).

Taking the limit as the thickness $\Delta x_k \to 0$ and $n \to \infty$ gives the volume integral

$$\begin{aligned} V &= \lim_{n\to\infty} \sum_{k=1}^{n} 2\pi(x_k + 1)\left(3x_k - x_k^2\right) \Delta x_k \\ &= \int_0^3 2\pi(x+1)(3x - x^2)\,dx \\ &= \int_0^3 2\pi(3x^2 + 3x - x^3 - x^2)\,dx \\ &= 2\pi \int_0^3 (2x^2 + 3x - x^3)\,dx \\ &= 2\pi\left[\frac{2}{3}x^3 + \frac{3}{2}x^2 - \frac{1}{4}x^4\right]_0^3 = \frac{45\pi}{2}. \end{aligned}$$

We now generalize the procedure used in Example 1.

The Shell Method

Suppose the region bounded by the graph of a nonnegative continuous function $y = f(x)$ and the x-axis over the finite closed interval $[a, b]$ lies to the right of the vertical line $x = L$ (Figure 6.19a). We assume $a \geq L$, so the vertical line may touch the region, but not pass through it. We generate a solid S by rotating this region about the vertical line L.

Let P be a partition of the interval $[a, b]$ by the points $a = x_0 < x_1 < \cdots < x_n = b$, and let c_k be the midpoint of the kth subinterval $[x_{k-1}, x_k]$. We approximate the region in Figure 6.19a with rectangles based on this partition of $[a, b]$. A typical approximating rectangle has height $f(c_k)$ and width $\Delta x_k = x_k - x_{k-1}$. If this rectangle is rotated about the vertical line $x = L$, then a shell is swept out, as in Figure 6.19b. A formula from geometry tells us that the volume of the shell swept out by the rectangle is

$$\begin{aligned} \Delta V_k &= 2\pi \times \text{average shell radius} \times \text{shell height} \times \text{thickness} \\ &= 2\pi \cdot (c_k - L) \cdot f(c_k) \cdot \Delta x_k. \qquad R = x_k - L \text{ and } r = x_{k-1} - L \end{aligned}$$

The volume of a cylindrical shell of height h with inner radius r and outer radius R is

$$\pi R^2 h - \pi r^2 h = 2\pi\left(\frac{R + r}{2}\right)(h)(R - r).$$

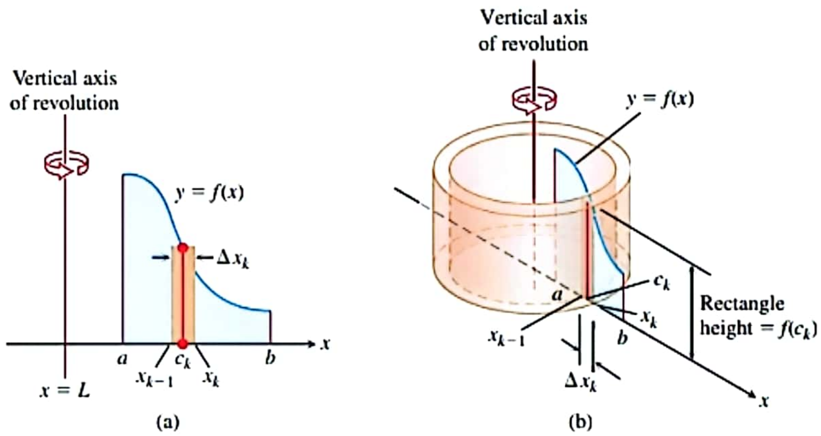


FIGURE 6.19 When the region shown in (a) is revolved about the vertical line $x = L$, a solid is produced which can be sliced into cylindrical shells. A typical shell is shown in (b).

We approximate the volume of the solid S by summing the volumes of the shells swept out by the n rectangles based on P:

$$V \approx \sum_{k=1}^{n} \Delta V_k.$$

The limit of this Riemann sum as each $\Delta x_k \to 0$ and $n \to \infty$ gives the volume of the solid as a definite integral:

$$V = \lim_{n\to\infty} \sum_{k=1}^{n} \Delta V_k = \int_a^b 2\pi(\text{shell radius})(\text{shell height})\, dx$$
$$= \int_a^b 2\pi(x - L)f(x)\, dx.$$

We refer to the variable of integration, here x, as the **thickness variable**. We use the first integral, rather than the second containing a formula for the integrand, to emphasize the *process* of the shell method. This will allow for rotations about a horizontal line L as well.

Shell Formula for Revolution About a Vertical Line

The volume of the solid generated by revolving the region between the x-axis and the graph of a continuous function $y = f(x) \geq 0$, $L \leq a \leq x \leq b$, about a vertical line $x = L$ is

$$V = \int_a^b 2\pi \begin{pmatrix}\text{shell}\\ \text{radius}\end{pmatrix} \begin{pmatrix}\text{shell}\\ \text{height}\end{pmatrix} dx.$$

EXAMPLE 2 The region bounded by the curve $y = \sqrt{x}$, the x-axis, and the line $x = 4$ is revolved about the y-axis to generate a solid. Find the volume of the solid.

Solution Sketch the region and draw a line segment across it *parallel* to the axis of revolution (Figure 6.20a). Label the segment's height (shell height) and distance from the axis of revolution (shell radius). (We drew the shell in Figure 6.20b, but you need not do that.)

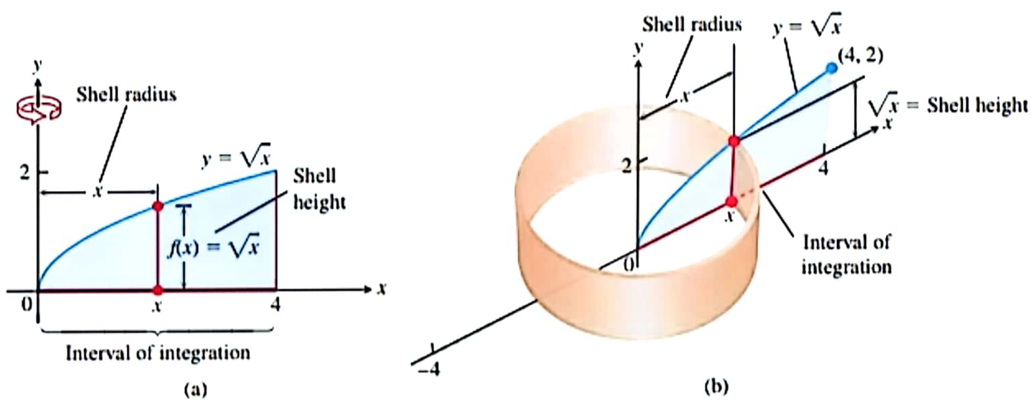


FIGURE 6.20 (a) The region, shell dimensions, and interval of integration in Example 2. (b) The shell swept out by the vertical segment in part (a) with a width Δx.

The shell thickness variable is x, so the limits of integration for the shell formula are $a = 0$ and $b = 4$ (Figure 6.20). The volume is then

$$V = \int_a^b 2\pi\begin{pmatrix}\text{shell}\\ \text{radius}\end{pmatrix}\begin{pmatrix}\text{shell}\\ \text{height}\end{pmatrix} dx$$
$$= \int_0^4 2\pi(x)(\sqrt{x})\, dx$$
$$= 2\pi\int_0^4 x^{3/2}\, dx = 2\pi\left[\frac{2}{5}x^{5/2}\right]_0^4 = \frac{128\pi}{5}.$$

So far, we have used vertical axes of revolution. For horizontal axes, we replace the x's with y's.

EXAMPLE 3 The region bounded by the curve $y = \sqrt{x}$, the x-axis, and the line $x = 4$ is revolved about the x-axis to generate a solid. Find the volume of the solid by the shell method.

Solution This is the solid whose volume was found by the disk method in Example 4 of Section 6.1. Now we find its volume by the shell method. First, sketch the region and draw a line segment across it *parallel* to the axis of revolution (Figure 6.21a). Label the segment's length (shell height) and distance from the axis of revolution (shell radius). (We drew the shell in Figure 6.21b, but you need not do that.)

In this case, the shell thickness variable is y, so the limits of integration for the shell formula method are $a = 0$ and $b = 2$ (along the y-axis in Figure 6.21). The volume of the solid is

$$V = \int_a^b 2\pi\begin{pmatrix}\text{shell}\\ \text{radius}\end{pmatrix}\begin{pmatrix}\text{shell}\\ \text{height}\end{pmatrix} dy$$
$$= \int_0^2 2\pi(y)(4 - y^2)\, dy$$
$$= 2\pi\int_0^2 (4y - y^3)\, dy$$
$$= 2\pi\left[2y^2 - \frac{y^4}{4}\right]_0^2 = 8\pi.$$

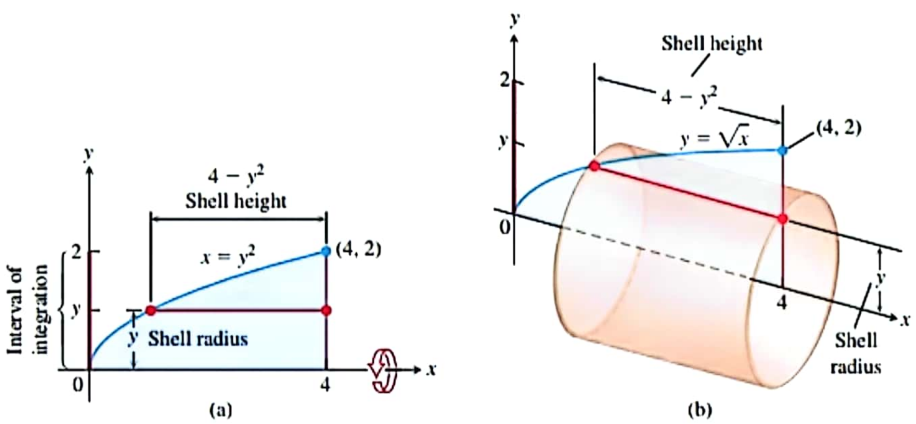


FIGURE 6.21 (a) The region, shell dimensions, and interval of integration in Example 3. (b) The shell swept out by the horizontal segment in part (a) with a width Δy.

6.3 Arc Length

We know what is meant by the length of a straight-line segment, but without calculus, we have no precise definition of the length of a general winding curve. If the curve is the graph of a continuous function defined over an interval, then we can find the length of the curve using a procedure similar to that we used for defining the area between the curve and the x-axis. This procedure results in a division of the curve from point A to point B into many pieces and joining successive points of division by straight-line segments. We then sum the lengths of all these line segments and define the length of the curve to be the limiting value of this sum as the number of segments goes to infinity.

Length of a Curve $y = f(x)$

Suppose the curve whose length we want to find is the graph of the function $y = f(x)$ from $x = a$ to $x = b$. In order to derive an integral formula for the length of the curve, we assume that f has a continuous derivative at every point of $[a, b]$. Such a function is called **smooth**, and its graph is a **smooth curve** because it does not have any breaks, corners, or cusps.

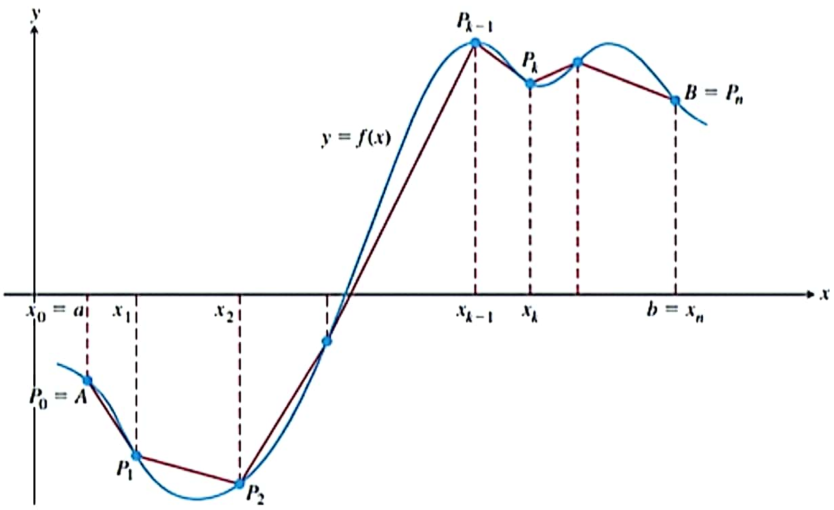


FIGURE 6.22 The length of the polygonal path $P_0P_1P_2 \cdots P_n$ approximates the length of the curve $y = f(x)$ from point A to point B.

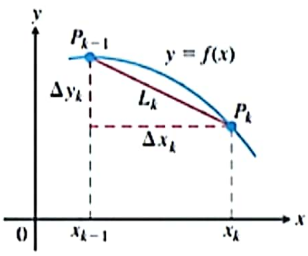


FIGURE 6.23 The arc $P_{k-1}P_k$ of the curve $y = f(x)$ is approximated by the straight-line segment shown here, which has length $L_k = \sqrt{(\Delta x_k)^2 + (\Delta y_k)^2}$.

We partition the interval $[a, b]$ into n subintervals with $a = x_0 < x_1 < x_2 < \cdots < x_n = b$. If $y_k = f(x_k)$, then the corresponding point $P_k(x_k, y_k)$ lies on the curve. Next we connect successive points P_{k-1} and P_k with straight-line segments that, taken together, form a polygonal path whose length approximates the length of the curve (Figure 6.22). If $\Delta x_k = x_k - x_{k-1}$ and $\Delta y_k = y_k - y_{k-1}$, then a representative line segment in the path has length (see Figure 6.23)

$$L_k = \sqrt{(\Delta x_k)^2 + (\Delta y_k)^2},$$

so the length of the curve is approximated by the sum

$$\sum_{k=1}^{n} L_k = \sum_{k=1}^{n} \sqrt{(\Delta x_k)^2 + (\Delta y_k)^2}. \qquad (1)$$

We expect the approximation to improve as the partition of $[a, b]$ becomes finer. Now, by the Mean Value Theorem, there is a point c_k, with $x_{k-1} < c_k < x_k$, such that

$$\Delta y_k = f'(c_k)\, \Delta x_k.$$

With this substitution for Δy_k, the sums in Equation (1) take the form

$$\sum_{k=1}^{n} L_k = \sum_{k=1}^{n} \sqrt{(\Delta x_k)^2 + (f'(c_k)\Delta x_k)^2} = \sum_{k=1}^{n} \sqrt{1 + [f'(c_k)]^2}\, \Delta x_k. \tag{2}$$

Because $\sqrt{1 + [f'(x)]^2}$ is continuous on $[a, b]$, the limit of the Riemann sum on the right-hand side of Equation (2) exists as the norm of the partition goes to zero, giving

$$\lim_{n\to\infty} \sum_{k=1}^{n} L_k = \lim_{n\to\infty} \sum_{k=1}^{n} \sqrt{1 + [f'(c_k)]^2}\, \Delta x_k = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx.$$

We define the value of this limiting integral to be the length of the curve.

DEFINITION If f' is continuous on $[a, b]$, then the **length (arc length)** of the curve $y = f(x)$ from the point $A = (a, f(a))$ to the point $B = (b, f(b))$ is the value of the integral

$$L = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx. \tag{3}$$

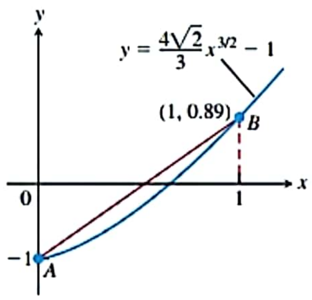


FIGURE 6.24 The length of the curve is slightly larger than the length of the line segment joining points A and B (Example 1).

EXAMPLE 1 Find the length of the curve (Figure 6.24)

$$y = \frac{4\sqrt{2}}{3}x^{3/2} - 1, \qquad 0 \le x \le 1.$$

Solution We use Equation (3) with $a = 0$, $b = 1$, and

$$y = \frac{4\sqrt{2}}{3}x^{3/2} - 1 \qquad x = 1, y \approx 0.89$$

$$\frac{dy}{dx} = \frac{4\sqrt{2}}{3} \cdot \frac{3}{2}x^{1/2} = 2\sqrt{2}x^{1/2}$$

$$\left(\frac{dy}{dx}\right)^2 = \left(2\sqrt{2}x^{1/2}\right)^2 = 8x.$$

The length of the curve over $x = 0$ to $x = 1$ is

$$L = \int_0^1 \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \int_0^1 \sqrt{1 + 8x}\, dx$$

Eq. (3) with $a = 0$, $b = 1$.

Let $u = 1 + 8x$, integrate, and replace u by $1 + 8x$.

$$= \frac{2}{3} \cdot \frac{1}{8}(1 + 8x)^{3/2}\Big]_0^1 = \frac{13}{6} \approx 2.17.$$

Notice that the length of the curve is slightly larger than the length of the straight-line segment joining the points $A = (0, -1)$ and $B = \left(1, 4\sqrt{2}/3 - 1\right)$ on the curve (see Figure 6.24):

$$2.17 > \sqrt{1^2 + (1.89)^2} \approx 2.14. \qquad \text{Decimal approximations}$$

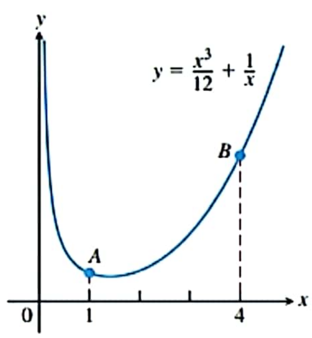


FIGURE 6.25 The curve in Example 2, where $A = (1, 13/12)$ and $B = (4, 67/12)$.

EXAMPLE 2 Find the length of the graph of

$$f(x) = \frac{x^3}{12} + \frac{1}{x}, \qquad 1 \le x \le 4.$$

Solution A graph of the function is shown in Figure 6.25. To use Equation (3), we find

$$f'(x) = \frac{x^2}{4} - \frac{1}{x^2}$$

so

$$1 + [f'(x)]^2 = 1 + \left(\frac{x^2}{4} - \frac{1}{x^2}\right)^2 = 1 + \left(\frac{x^4}{16} - \frac{1}{2} + \frac{1}{x^4}\right)$$
$$= \frac{x^4}{16} + \frac{1}{2} + \frac{1}{x^4} = \left(\frac{x^2}{4} + \frac{1}{x^2}\right)^2.$$

The length of the graph over $[1, 4]$ is

$$L = \int_1^4 \sqrt{1 + [f'(x)]^2}\, dx = \int_1^4 \left(\frac{x^2}{4} + \frac{1}{x^2}\right) dx$$
$$= \left[\frac{x^3}{12} - \frac{1}{x}\right]_1^4 = \left(\frac{64}{12} - \frac{1}{4}\right) - \left(\frac{1}{12} - 1\right) = \frac{72}{12} = 6.$$

EXAMPLE 3 Find the length of the curve

$$y = \frac{1}{2}(e^x + e^{-x}), \qquad 0 \le x \le 2.$$

Solution We use Equation (3) with $a = 0$, $b = 2$, and

$$y = \frac{1}{2}(e^x + e^{-x})$$
$$\frac{dy}{dx} = \frac{1}{2}(e^x - e^{-x})$$
$$\left(\frac{dy}{dx}\right)^2 = \frac{1}{4}(e^{2x} - 2 + e^{-2x})$$
$$1 + \left(\frac{dy}{dx}\right)^2 = \frac{1}{4}(e^{2x} + 2 + e^{-2x}) = \left[\frac{1}{2}(e^x + e^{-x})\right]^2.$$

The length of the curve from $x = 0$ to $x = 2$ is

$$L = \int_0^2 \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \int_0^2 \frac{1}{2}(e^x + e^{-x})\, dx$$

Eq. (3) with $a = 0$, $b = 2$.

$$= \frac{1}{2}\left[e^x - e^{-x}\right]_0^2 = \frac{1}{2}(e^2 - e^{-2}) \approx 3.63.$$

Dealing with Discontinuities in dy/dx

At a point on a curve where dy/dx fails to exist, dx/dy may exist. In this case, we may be able to find the curve's length by expressing x as a function of y and applying the following analogue of Equation (3):

Formula for the Length of $x = g(y)$, $c \le y \le d$

If g' is continuous on $[c, d]$, the length of the curve $x = g(y)$ from $A = (g(c), c)$ to $B = (g(d), d)$ is

$$L = \int_c^d \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy = \int_c^d \sqrt{1 + [g'(y)]^2}\, dy. \qquad (4)$$

EXAMPLE 4 Find the length of the curve $y = (x/2)^{2/3}$ from $x = 0$ to $x = 2$.

Solution The derivative

$$\frac{dy}{dx} = \frac{2}{3}\left(\frac{x}{2}\right)^{-1/3}\left(\frac{1}{2}\right) = \frac{1}{3}\left(\frac{2}{x}\right)^{1/3}$$

is not defined at $x = 0$, so we cannot find the curve's length with Equation (3).

We therefore rewrite the equation to express x in terms of y:

$$y = \left(\frac{x}{2}\right)^{2/3}$$

$$y^{3/2} = \frac{x}{2} \qquad \text{Raise both sides to the power } 3/2.$$

$$x = 2y^{3/2}. \qquad \text{Solve for } x.$$

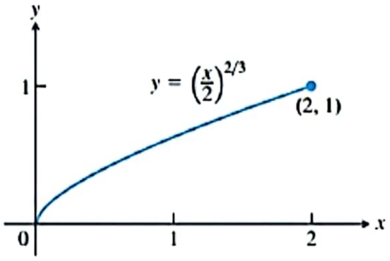


FIGURE 6.26 The graph of $y = (x/2)^{2/3}$ from $x = 0$ to $x = 2$ is also the graph of $x = 2y^{3/2}$ from $y = 0$ to $y = 1$ (Example 4).

From this we see that the curve whose length we want is also the graph of $x = 2y^{3/2}$ from $y = 0$ to $y = 1$ (Figure 6.26).

The derivative

$$\frac{dx}{dy} = 2\left(\frac{3}{2}\right)y^{1/2} = 3y^{1/2}$$

is continuous on $[0, 1]$. We may therefore use Equation (4) to find the curve's length:

$$L = \int_c^d \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\,dy = \int_0^1 \sqrt{1 + 9y}\,dy$$

Eq. (4) with $c = 0$, $d = 1$. Let $u = 1 + 9y$, $du/9 = dy$, integrate, and substitute back.

$$= \frac{1}{9}\cdot\frac{2}{3}(1 + 9y)^{3/2}\Big]_0^1$$

$$= \frac{2}{27}\left(10\sqrt{10} - 1\right) \approx 2.27.$$

The Differential Formula for Arc Length

If $y = f(x)$ and if f' is continuous on $[a, b]$, then by the Fundamental Theorem of Calculus we can define a new function

$$s(x) = \int_a^x \sqrt{1 + [f'(t)]^2}\,dt. \tag{5}$$

From Equation (3) and Figure 6.22, we see that this function $s(x)$ is continuous and measures the length along the curve $y = f(x)$ from the initial point $P_0(a, f(a))$ to the point $Q(x, f(x))$ for each $x \in [a, b]$. The function s is called the **arc length function** for $y = f(x)$. From the Fundamental Theorem, the function s is differentiable on (a, b) and

$$\frac{ds}{dx} = \sqrt{1 + [f'(x)]^2} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}.$$

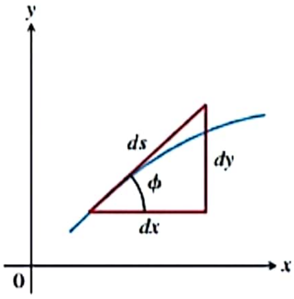
Then the differential of arc length is

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,dx. \tag{6}$$

A useful way to remember Equation (6) is to write

$$ds = \sqrt{dx^2 + dy^2}, \tag{7}$$

which can be integrated between appropriate limits to give the total length of a curve. From this point of view, all the arc length formulas are simply different expressions for the equation



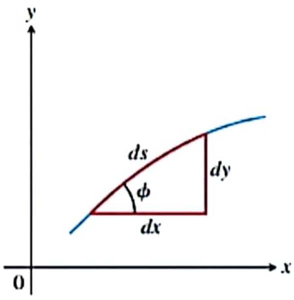


FIGURE 6.27 Diagrams for remembering the equation $ds = \sqrt{dx^2 + dy^2}$.

$L = \int ds$. Figure 6.27a gives the exact interpretation of ds corresponding to Equation (7). Figure 6.27b is not strictly accurate, but is to be thought of as a simplified approximation of Figure 6.27a. That is, $ds \approx \Delta s$.

EXAMPLE 5 Find the arc length function for the curve in Example 2, taking $A = (1, 13/12)$ as the starting point (see Figure 6.25).

Solution In the solution to Example 2, we found that

$$1 + [f'(x)]^2 = \left(\frac{x^2}{4} + \frac{1}{x^2}\right)^2.$$

Therefore the arc length function is given by

$$s(x) = \int_1^x \sqrt{1 + [f'(t)]^2}\,dt = \int_1^x \left(\frac{t^2}{4} + \frac{1}{t^2}\right) dt$$

$$= \left[\frac{t^3}{12} - \frac{1}{t}\right]_1^x = \frac{x^3}{12} - \frac{1}{x} + \frac{11}{12}.$$

To compute the arc length along the curve from $A = (1, 13/12)$ to $B = (4, 67/12)$, for instance, we simply calculate

$$s(4) = \frac{4^3}{12} - \frac{1}{4} + \frac{11}{12} = 6.$$

This is the same result we obtained in Example 2.

Exercises 6.3

Finding Lengths of Curves

Find the lengths of the curves in Exercises 1–14. If you have a grapher, you may want to graph these curves to see what they look like.

1. $y = (1/3)(x^2 + 2)^{3/2}$ from $x = 0$ to $x = 3$
2. $y = x^{3/2}$ from $x = 0$ to $x = 4$
3. $x = (y^3/3) + 1/(4y)$ from $y = 1$ to $y = 3$
4. $x = (y^{3/2}/3) - y^{1/2}$ from $y = 1$ to $y = 9$
5. $x = (y^4/4) + 1/(8y^2)$ from $y = 1$ to $y = 2$
6. $x = (y^3/6) + 1/(2y)$ from $y = 2$ to $y = 3$
7. $y = (3/4)x^{4/3} - (3/8)x^{2/3} + 5, \quad 1 \le x \le 8$
8. $y = (x^3/3) + x^2 + x + 1/(4x + 4), \quad 0 \le x \le 2$
9. $y = \ln x - \dfrac{x^2}{8}$ from $x = 1$ to $x = 2$
10. $y = \dfrac{x^2}{2} - \dfrac{\ln x}{4}$ from $x = 1$ to $x = 3$
11. $y = \dfrac{x^3}{3} + \dfrac{1}{4x}, \quad 1 \le x \le 3$
12. $y = \dfrac{x^5}{5} + \dfrac{1}{12x^3}, \quad \dfrac{1}{2} \le x \le 1$
13. $x = \displaystyle\int_0^y \sqrt{\sec^4 t - 1}\,dt, \quad -\pi/4 \le y \le \pi/4$
14. $y = \displaystyle\int_{-2}^x \sqrt{3t^4 - 1}\,dt, \quad -2 \le x \le -1$

T Finding Integrals for Lengths of Curves

In Exercises 15–22, do the following.

a. Set up an integral for the length of the curve.
b. Graph the curve to see what it looks like.
c. Use your grapher's or computer's integral evaluator to find the curve's length numerically.

15. $y = x^2, \quad -1 \le x \le 2$
16. $y = \tan x, \quad -\pi/3 \le x \le 0$
17. $x = \sin y, \quad 0 \le y \le \pi$
18. $x = \sqrt{1 - y^2}, \quad -1/2 \le y \le 1/2$
19. $y^2 + 2y = 2x + 1$ from $(-1, -1)$ to $(7, 3)$
20. $y = \sin x - x\cos x, \quad 0 \le x \le \pi$
21. $y = \displaystyle\int_0^x \tan t\,dt, \quad 0 \le x \le \pi/6$
22. $x = \displaystyle\int_0^y \sqrt{\sec^2 t - 1}\,dt, \quad -\pi/3 \le y \le \pi/4$

Theory and Examples

23. a. Find a curve with a positive derivative through the point $(1, 1)$ whose length integral (Equation 3) is

$$L = \int_1^4 \sqrt{1 + \frac{1}{4x}}\,dx.$$

b. How many such curves are there? Give reasons for your answer.

24. **a.** Find a curve with a positive derivative through the point (0, 1) whose length integral (Equation 4) is

$$L = \int_1^2 \sqrt{1 + \frac{1}{y^4}}\, dy.$$

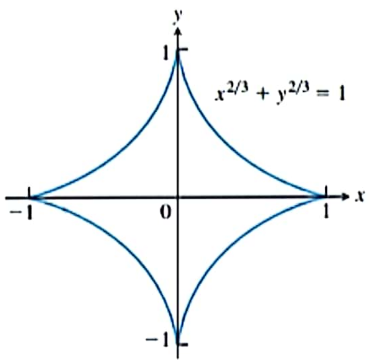
b. How many such curves are there? Give reasons for your answer.

25. Find the length of the curve

$$y = \int_0^x \sqrt{\cos 2t}\, dt$$

from $x = 0$ to $x = \pi/4$.

26. The length of an astroid The graph of the equation $x^{2/3} + y^{2/3} = 1$ is one of a family of curves called *astroids* (not "asteroids") because of their starlike appearance (see the accompanying figure). Find the length of this particular astroid by finding the length of half the first-quadrant portion, $y = (1 - x^{2/3})^{3/2}$, $\sqrt{2}/4 \le x \le 1$, and multiplying by 8.



27. Length of a line segment Use the arc length formula (Equation 3) to find the length of the line segment $y = 3 - 2x$, $0 \le x \le 2$. Check your answer by finding the length of the segment as the hypotenuse of a right triangle.

28. Circumference of a circle Set up an integral to find the circumference of a circle of radius r centered at the origin. You will learn how to evaluate the integral in Section 8.4.

29. If $9x^2 = y(y - 3)^2$, show that

$$ds^2 = \frac{(y + 1)^2}{4y}\, dy^2.$$

30. If $4x^2 - y^2 = 64$, show that

$$ds^2 = \frac{4}{y^2}(5x^2 - 16)\, dx^2.$$

31. Is there a smooth (continuously differentiable) curve $y = f(x)$ whose length over the interval $0 \le x \le a$ is always $\sqrt{2}a$? Give reasons for your answer.

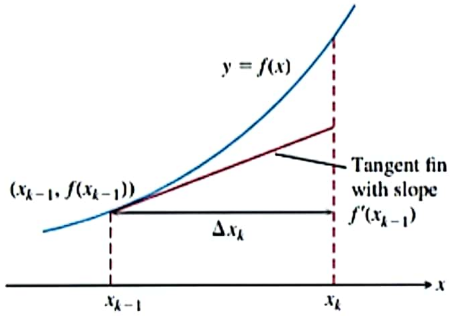
32. Using tangent fins to derive the length formula for curves Assume that f is smooth on $[a, b]$ and partition the interval $[a, b]$ in the usual way. In each subinterval $[x_{k-1}, x_k]$, construct the *tangent fin* at the point $(x_{k-1}, f(x_{k-1}))$, as shown in the accompanying figure.

a. Show that the length of the kth tangent fin over the interval $[x_{k-1}, x_k]$ equals $\sqrt{(\Delta x_k)^2 + (f'(x_{k-1})\,\Delta x_k)^2}$.

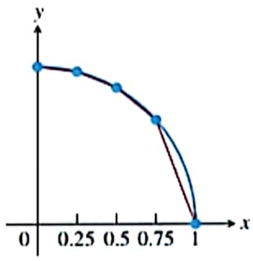
b. Show that

$$\lim_{n\to\infty} \sum_{k=1}^{n} (\text{length of } k\text{th tangent fin}) = \int_a^b \sqrt{1 + (f'(x))^2}\, dx,$$

which is the length L of the curve $y = f(x)$ from a to b.



33. Approximate the arc length of one-quarter of the unit circle (which is $\pi/2$) by computing the length of the polygonal approximation with $n = 4$ segments (see accompanying figure).



34. Distance between two points Assume that the two points (x_1, y_1) and (x_2, y_2) lie on the graph of the straight line $y = mx + b$. Use the arc length formula (Equation 3) to find the distance between the two points.

35. Find the arc length function for the graph of $f(x) = 2x^{3/2}$ using (0, 0) as the starting point. What is the length of the curve from (0, 0) to (1, 2)?

36. Find the arc length function for the curve in Exercise 8, using (0, 1/4) as the starting point. What is the length of the curve from (0, 1/4) to (1, 59/24)?

COMPUTER EXPLORATIONS

In Exercises 37–42, use a CAS to perform the following steps for the given graph of the function over the closed interval.

a. Plot the curve together with the polygonal path approximations for $n = 2, 4, 8$ partition points over the interval. (See Figure 6.22.)

b. Find the corresponding approximation to the length of the curve by summing the lengths of the line segments.

c. Evaluate the length of the curve using an integral. Compare your approximations for $n = 2, 4, 8$ with the actual length given by the integral. How does the actual length compare with the approximations as n increases? Explain your answer.

37. $f(x) = \sqrt{1 - x^2}, \quad -1 \le x \le 1$

38. $f(x) = x^{1/3} + x^{2/3}, \quad 0 \le x \le 2$

39. $f(x) = \sin(\pi x^2), \quad 0 \le x \le \sqrt{2}$

40. $f(x) = x^2 \cos x, \quad 0 \le x \le \pi$

41. $f(x) = \dfrac{x - 1}{4x^2 + 1}, \quad -\dfrac{1}{2} \le x \le 1$

42. $f(x) = x^3 - x^2, \quad -1 \le x \le 1$

6.4 Areas of Surfaces of Revolution

When you jump rope, the rope sweeps out a surface in the space around you similar to what is called a *surface of revolution*. The surface surrounds a volume of revolution, and many applications require that we know the area of the surface rather than the volume it encloses. In this section we define areas of surfaces of revolution. More general surfaces are treated in Chapter 16.

Defining Surface Area

If you revolve a region in the plane that is bounded by the graph of a function over an interval, it sweeps out a solid of revolution, as we saw earlier in the chapter. However, if you revolve only the bounding curve itself, it does not sweep out any interior volume but rather a surface that surrounds the solid and forms part of its boundary. Just as we were interested in defining and finding the length of a curve in the last section, we are now interested in defining and finding the area of a surface generated by revolving a curve about an axis.

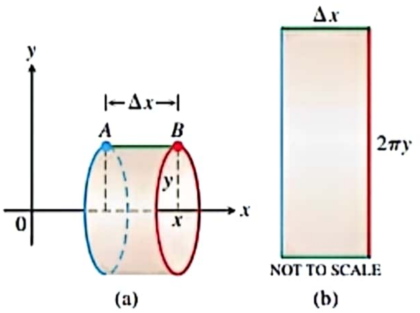


FIGURE 6.28 (a) A cylindrical surface generated by rotating the horizontal line segment AB of length Δx about the x-axis has area $2\pi y \Delta x$. (b) The cut and rolled-out cylindrical surface as a rectangle.

Before considering general curves, we begin by rotating horizontal and slanted line segments about the x-axis. If we rotate the horizontal line segment AB having length Δx about the x-axis (Figure 6.28a), we generate a cylinder with surface area $2\pi y \Delta x$. This area is the same as that of a rectangle with side lengths Δx and $2\pi y$ (Figure 6.28b). The length $2\pi y$ is the circumference of the circle of radius y generated by rotating the point (x, y) on the line AB about the x-axis.

Suppose the line segment AB has length L and is slanted rather than horizontal. Now when AB is rotated about the x-axis, it generates a frustum of a cone (Figure 6.29a). From classical geometry, the surface area of this frustum is $2\pi y^* L$, where $y^* = (y_1 + y_2)/2$ is the average height of the slanted segment AB above the x-axis. This surface area is the same as that of a rectangle with side lengths L and $2\pi y^*$ (Figure 6.29b).

Let's build on these geometric principles to define the area of a surface swept out by revolving more general curves about the x-axis. Suppose we want to find the area of the surface swept out by revolving the graph of a nonnegative continuous function $y = f(x)$, $a \le x \le b$, about the x-axis. We partition the closed interval $[a, b]$ in the usual way and use the points in the partition to subdivide the graph into short arcs. Figure 6.30 shows a typical arc PQ and the band it sweeps out as part of the graph of f.

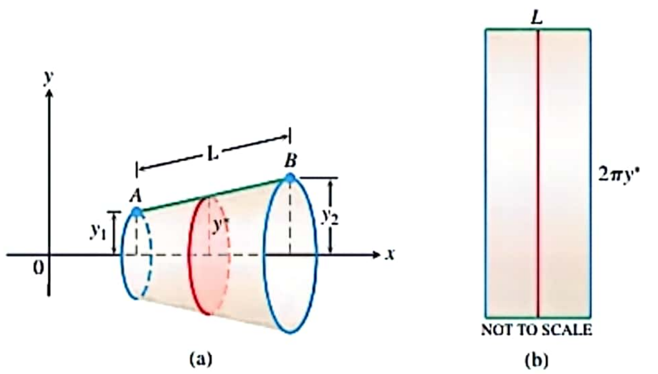


FIGURE 6.29 (a) The frustum of a cone generated by rotating the slanted line segment AB of length L about the x-axis has area $2\pi y^* L$. (b) The area of the rectangle for $y^* = \dfrac{y_1 + y_2}{2}$, the average height of AB above the x-axis.

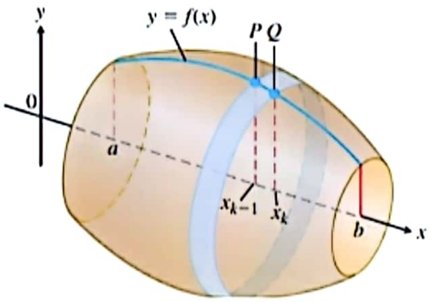


FIGURE 6.30 The surface generated by revolving the graph of a nonnegative function $y = f(x)$, $a \leq x \leq b$, about the x-axis. The surface is a union of bands like the one swept out by the arc PQ.

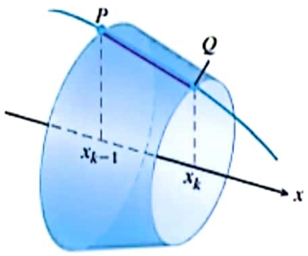


FIGURE 6.31 The line segment joining P and Q sweeps out a frustum of a cone.

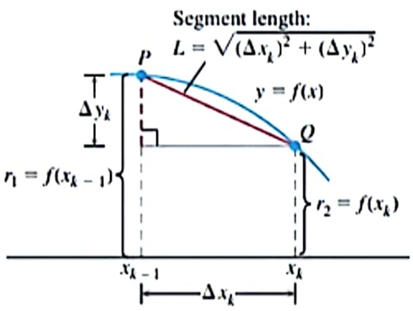


FIGURE 6.32 Dimensions associated with the arc and line segment PQ.

As the arc PQ revolves about the x-axis, the line segment joining P and Q sweeps out a frustum of a cone whose axis lies along the x-axis (Figure 6.31). The surface area of this frustum approximates the surface area of the band swept out by the arc PQ. The surface area of the frustum of the cone shown in Figure 6.31 is $2\pi y^*L$, where y^* is the average height of the line segment joining P and Q, and L is its length (just as before). Since $f \geq 0$, from Figure 6.32 we see that the average height of the line segment is $y^* = (f(x_{k-1}) + f(x_k))/2$, and the slant length is $L = \sqrt{(\Delta x_k)^2 + (\Delta y_k)^2}$. Therefore,

$$\begin{aligned}\text{Frustum surface area} &= 2\pi \cdot \frac{f(x_{k-1}) + f(x_k)}{2} \cdot \sqrt{(\Delta x_k)^2 + (\Delta y_k)^2} \\ &= \pi(f(x_{k-1}) + f(x_k))\sqrt{(\Delta x_k)^2 + (\Delta y_k)^2}.\end{aligned}$$

The area of the original surface, being the sum of the areas of the bands swept out by arcs like arc PQ, is approximated by the frustum area sum

$$\sum_{k=1}^{n} \pi(f(x_{k-1}) + f(x_k))\sqrt{(\Delta x_k)^2 + (\Delta y_k)^2}. \tag{1}$$

We expect the approximation to improve as the partition of $[a, b]$ becomes finer. Moreover, if the function f is differentiable, then by the Mean Value Theorem, there is a point $(c_k, f(c_k))$ on the curve between P and Q where the tangent is parallel to the segment PQ (Figure 6.33). At this point,

$$f'(c_k) = \frac{\Delta y_k}{\Delta x_k},$$

$$\Delta y_k = f'(c_k)\,\Delta x_k.$$

With this substitution for Δy_k, the sums in Equation (1) take the form

$$\begin{aligned}&\sum_{k=1}^{n} \pi(f(x_{k-1}) + f(x_k))\sqrt{(\Delta x_k)^2 + (f'(c_k)\,\Delta x_k)^2} \\ &= \sum_{k=1}^{n} \pi(f(x_{k-1}) + f(x_k))\sqrt{1 + (f'(c_k))^2}\,\Delta x_k.\end{aligned} \tag{2}$$

These sums are not the Riemann sums of any function because the points x_{k-1}, x_k, and c_k are not the same. However, it can be proved that as the norm of the partition of $[a, b]$ goes to zero, the sums in Equation (2) converge to the integral

$$\int_a^b 2\pi f(x)\sqrt{1 + (f'(x))^2}\,dx.$$

We therefore define this integral to be the area of the surface swept out by the graph of f from a to b.

DEFINITION If the function $f(x) \geq 0$ is continuously differentiable on $[a, b]$, the **area of the surface** generated by revolving the graph of $y = f(x)$ about the x-axis is

$$S = \int_a^b 2\pi y\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,dx = \int_a^b 2\pi f(x)\sqrt{1 + (f'(x))^2}\,dx. \tag{3}$$

The square root in Equation (3) is the same one that appears in the formula for the arc length differential of the generating curve in Equation (6) of Section 6.3.

EXAMPLE 1 Find the area of the surface generated by revolving the curve $y = 2\sqrt{x}$, $1 \leq x \leq 2$, about the x-axis (Figure 6.34).

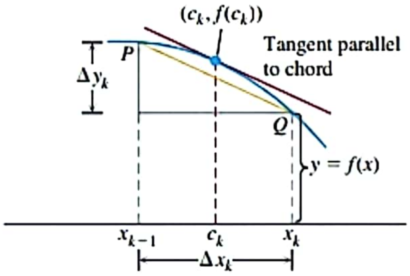


FIGURE 6.33 If f is smooth, the Mean Value Theorem guarantees the existence of a point c_k where the tangent is parallel to segment PQ.

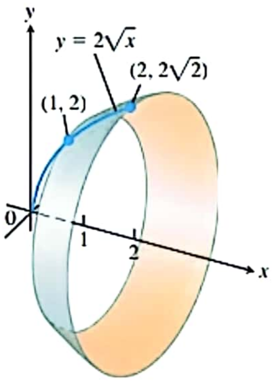


FIGURE 6.34 In Example 1 we calculate the area of this surface.

Solution We evaluate the formula

$$S = \int_a^b 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \qquad \text{Eq. (3)}$$

with

$$a = 1, \qquad b = 2, \qquad y = 2\sqrt{x}, \qquad \frac{dy}{dx} = \frac{1}{\sqrt{x}}.$$

First, we perform some algebraic manipulation on the radical in the integrand to transform it into an expression that is easier to integrate.

$$\begin{aligned}\sqrt{1 + \left(\frac{dy}{dx}\right)^2} &= \sqrt{1 + \left(\frac{1}{\sqrt{x}}\right)^2}\\ &= \sqrt{1 + \frac{1}{x}} = \sqrt{\frac{x+1}{x}} = \frac{\sqrt{x+1}}{\sqrt{x}}\end{aligned}$$

With these substitutions, we have

$$\begin{aligned}S &= \int_1^2 2\pi \cdot 2\sqrt{x}\frac{\sqrt{x+1}}{\sqrt{x}}\, dx = 4\pi \int_1^2 \sqrt{x+1}\, dx\\ &= 4\pi \cdot \frac{2}{3}(x+1)^{3/2}\Big]_1^2 = \frac{8\pi}{3}\left(3\sqrt{3} - 2\sqrt{2}\right).\end{aligned}$$

Revolution About the y-Axis

For revolution about the y-axis, we interchange x and y in Equation (3).

> **Surface Area for Revolution About the y-Axis**
>
> If $x = g(y) \ge 0$ is continuously differentiable on $[c, d]$, the area of the surface generated by revolving the graph of $x = g(y)$ about the y-axis is
>
> $$S = \int_c^d 2\pi x \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy = \int_c^d 2\pi g(y)\sqrt{1 + (g'(y))^2}\, dy. \qquad (4)$$

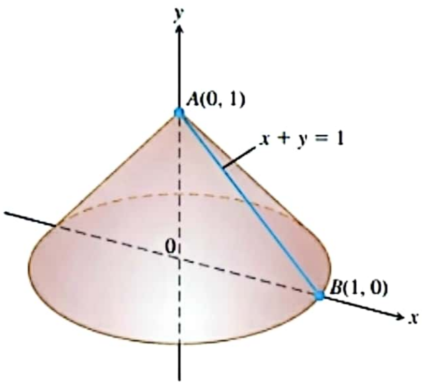


FIGURE 6.35 Revolving line segment AB about the y-axis generates a cone whose lateral surface area we can now calculate in two different ways (Example 2).

EXAMPLE 2 The line segment $x = 1 - y$, $0 \le y \le 1$, is revolved about the y-axis to generate the cone in Figure 6.35. Find its lateral surface area (which excludes the base area).

Solution Here we have a calculation we can check with a formula from geometry:

$$\text{Lateral surface area} = \frac{\text{base circumference}}{2} \times \text{slant height} = \pi\sqrt{2}.$$

To see how Equation (4) gives the same result, we take

$$c = 0, \qquad d = 1, \qquad x = 1 - y, \qquad \frac{dx}{dy} = -1,$$

$$\sqrt{1 + \left(\frac{dx}{dy}\right)^2} = \sqrt{1 + (-1)^2} = \sqrt{2}$$

7 Integrals and Transcendental Functions

OVERVIEW Our treatment of the logarithmic and exponential functions has been rather informal until now, appealing to intuition and graphs to describe what they mean and to explain some of their characteristics. In this chapter, we give a rigorous analytic approach to the definitions and properties of these functions, and we study a wide range of applied problems in which they play a role. We also introduce the hyperbolic functions and their inverses, with their applications to integration and hanging cables. Like the trigonometric functions, all of these functions belong to the class of transcendental functions.

7.1 The Logarithm Defined as an Integral

In Chapter 1, we introduced the natural logarithm function $\ln x$ as the inverse of the exponential function e^x. The function e^x was chosen as that function in the family of general exponential functions $a^x, a > 0$, whose graph has slope 1 as it crosses the y-axis. The function a^x was presented intuitively, however, based on its graph at rational values of x.

In this section we recreate the theory of logarithmic and exponential functions from an entirely different point of view. Here we define these functions analytically and recover their behaviors. To begin, we use the Fundamental Theorem of Calculus to define the natural logarithm function $\ln x$ as an integral. We quickly develop its properties, including the algebraic, geometric, and analytic properties seen before. Next we introduce the function e^x as the inverse function of $\ln x$, and establish its previously seen properties. Defining $\ln x$ as an integral and e^x as its inverse is an indirect approach. While it may at first seem strange, it gives an elegant and powerful way to obtain and validate the key properties of logarithmic and exponential functions.

Definition of the Natural Logarithm Function

The natural logarithm of a positive number x, written as $\ln x$, is the value of an integral. The integral is suggested from our earlier results in Chapter 5.

DEFINITION The **natural logarithm** is the function given by

$$\ln x = \int_1^x \frac{1}{t}\,dt, \qquad x > 0.$$

From the Fundamental Theorem of Calculus, $\ln x$ is a continuous function. Geometrically, if $x > 1$, then $\ln x$ is the area under the curve $y = 1/t$ from $t = 1$ to $t = x$ (Figure 7.1). For $0 < x < 1$, $\ln x$ gives the negative of the area under the curve from x to 1, and the function is not defined for $x \leq 0$. From the Zero Width Interval Rule for definite integrals, we also have

$$\ln 1 = \int_1^1 \frac{1}{t}\,dt = 0.$$

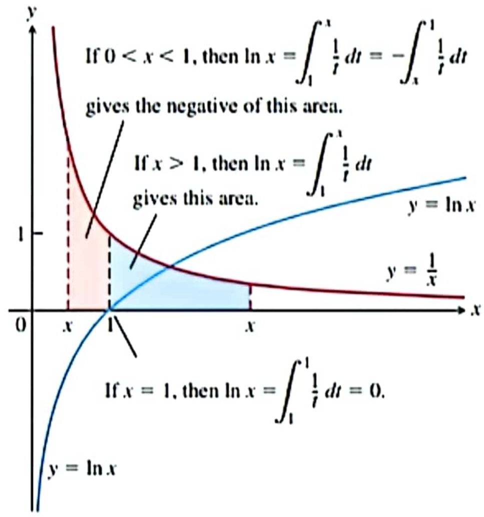


FIGURE 7.1 The graph of $y = \ln x$ and its relation to the function $y = 1/x, x > 0$. The graph of the logarithm rises above the x-axis as x moves from 1 to the right, and it falls below the axis as x moves from 1 to the left.

Notice that we show the graph of $y = 1/x$ in Figure 7.1 but use $y = 1/t$ in the integral. Using x for everything would have us writing

$$\ln x = \int_1^x \frac{1}{x}\,dx,$$

with x meaning two different things. So we change the variable of integration to t.

By using rectangles to obtain finite approximations of the area under the graph of $y = 1/t$ and over the interval between $t = 1$ and $t = x$, as in Section 5.1, we can approximate the values of the function $\ln x$. Several values are given in Table 7.1. There is an important number between $x = 2$ and $x = 3$ whose natural logarithm equals 1. This number, which we now define, exists because $\ln x$ is a continuous function and therefore satisfies the Intermediate Value Theorem on $[2, 3]$.

TABLE 7.1 Typical 2-place values of $\ln x$

x	$\ln x$
0	undefined
0.05	−3.00
0.5	−0.69
1	0
2	0.69
3	1.10
4	1.39
10	2.30

DEFINITION The **number e** is that number in the domain of the natural logarithm satisfying

$$\ln(e) = \int_1^e \frac{1}{t}\,dt = 1.$$

Interpreted geometrically, the number e corresponds to the point on the x-axis for which the area under the graph of $y = 1/t$ and above the interval $[1, e]$ equals the area of the unit square. That is, the area of the region shaded blue in Figure 7.1 is 1 sq unit when $x = e$. We will see further on that this is the same number $e \approx 2.718281828$ we have encountered before.

The Derivative of $y = \ln x$

By the first part of the Fundamental Theorem of Calculus (Section 5.4),

$$\frac{d}{dx}\ln x = \frac{d}{dx}\int_1^x \frac{1}{t}\,dt = \frac{1}{x}.$$

For every positive value of x, we have

$$\frac{d}{dx}\ln x = \frac{1}{x}. \tag{1}$$

Therefore, the function $y = \ln x$ is a solution to the initial value problem $dy/dx = 1/x$, $x > 0$, with $y(1) = 0$. Notice that the derivative is always positive.

If u is a differentiable function of x whose values are positive, so that $\ln u$ is defined, then applying the Chain Rule we obtain

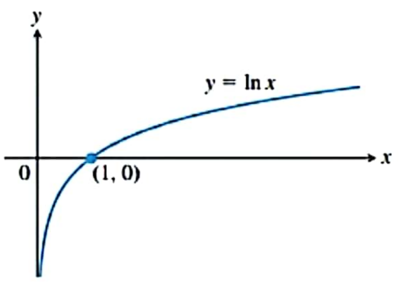
$$\boxed{\frac{d}{dx}\ln u = \frac{1}{u}\frac{du}{dx}, \qquad u > 0.} \tag{2}$$

The derivative of $\ln|x|$ can be found just as in Example 3(c) of Section 3.8, giving

$$\boxed{\frac{d}{dx}\ln|x| = \frac{1}{x}, \qquad x \neq 0.} \tag{3}$$

Moreover, if b is any constant with $bx > 0$, Equation (2) gives

$$\frac{d}{dx}\ln bx = \frac{1}{bx}\cdot\frac{d}{dx}(bx) = \frac{1}{bx}(b) = \frac{1}{x}.$$



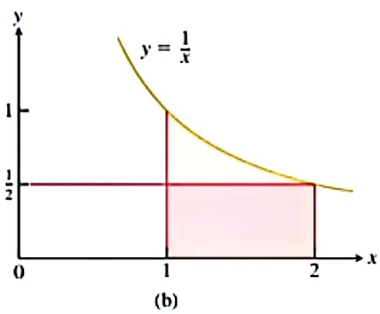


FIGURE 7.2 (a) The graph of the natural logarithm. (b) The rectangle of height $y = 1/2$ fits beneath the graph of $y = 1/x$ for the interval $1 \leq x \leq 2$.

The Graph and Range of $\ln x$

The derivative $d(\ln x)/dx = 1/x$ is positive for $x > 0$, so $\ln x$ is an increasing function of x. The second derivative, $-1/x^2$, is negative, so the graph of $\ln x$ is concave down. (See Figure 7.2a.)

The function $\ln x$ has the following familiar algebraic properties, which we stated in Section 1.6. In Section 4.2 we showed these properties are a consequence of Corollary 2 of the Mean Value Theorem, and those derivations still apply.

1. $\ln bx = \ln b + \ln x$	2. $\ln \frac{b}{x} = \ln b - \ln x$
3. $\ln \frac{1}{x} = -\ln x$	4. $\ln x^r = r\ln x$, r rational

We can estimate the value of $\ln 2$ by considering the area under the graph of $y = 1/x$ and above the interval $[1, 2]$. In Figure 7.2(b) a rectangle of height $1/2$ over the interval

7.5 Indeterminate Forms and L'Hôpital's Rule

HISTORICAL BIOGRAPHY
Guillaume François Antoine de l'Hôpital
(1661–1704)
Johann Bernoulli
(1667–1748)

John (Johann) Bernoulli discovered a rule using derivatives to calculate limits of fractions whose numerators and denominators both approach zero or $+\infty$. The rule is known today as **l'Hôpital's Rule**, after Guillaume de l'Hôpital. He was a French nobleman who wrote the first introductory differential calculus text, where the rule first appeared in print. Limits involving transcendental functions often require some use of the rule for their calculation.

Indeterminate Form 0/0

If we want to know how the function

$$F(x) = \frac{x - \sin x}{x^3}$$

behaves *near* $x = 0$ (where it is undefined), we can examine the limit of $F(x)$ as $x \to 0$. We cannot apply the Quotient Rule for limits (Theorem 1 of Chapter 2) because the limit of the denominator is 0. Moreover, in this case, *both* the numerator and denominator approach 0, and $0/0$ is undefined. Such limits may or may not exist in general, but the limit does exist for the function $F(x)$ under discussion by applying l'Hôpital's Rule, as we will see in Example 1d.

If the continuous functions $f(x)$ and $g(x)$ are both zero at $x = a$, then

$$\lim_{x \to a} \frac{f(x)}{g(x)}$$

cannot be found by substituting $x = a$. The substitution produces $0/0$, a meaningless expression, which we cannot evaluate. We use $0/0$ as a notation for an expression known as an **indeterminate form**. Other meaningless expressions often occur, such as ∞/∞, $\infty \cdot 0$, $\infty - \infty$, 0^0, and 1^∞, which cannot be evaluated in a consistent way; these are called indeterminate forms as well. Sometimes, but not always, limits that lead to indeterminate forms may be found by cancelation, rearrangement of terms, or other algebraic manipulations. This was our experience in Chapter 2. It took considerable analysis in Section 2.4 to find $\lim_{x \to 0} (\sin x)/x$. But we have had success with the limit

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a},$$

from which we calculate derivatives and which produces the indeterminant form $0/0$ when we attempt to substitute $x = a$. L'Hôpital's Rule enables us to draw on our success with derivatives to evaluate limits that otherwise lead to indeterminate forms.

THEOREM 5—L'Hôpital's Rule Suppose that $f(a) = g(a) = 0$, that f and g are differentiable on an open interval I containing a, and that $g'(x) \neq 0$ on I if $x \neq a$. Then

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)},$$

assuming that the limit on the right side of this equation exists.

We give a proof of Theorem 5 at the end of this section.

Caution

To apply l'Hôpital's Rule to f/g, divide the derivative of f by the derivative of g. Do not fall into the trap of taking the derivative of f/g. The quotient to use is f'/g', not $(f/g)'$.

EXAMPLE 1 The following limits involve $0/0$ indeterminate forms, so we apply l'Hôpital's Rule. In some cases, it must be applied repeatedly.

(a) $$\lim_{x\to 0}\frac{3x-\sin x}{x} = \lim_{x\to 0}\frac{3-\cos x}{1} = \left.\frac{3-\cos x}{1}\right|_{x=0} = 2$$

(b) $$\lim_{x\to 0}\frac{\sqrt{1+x}-1}{x} = \lim_{x\to 0}\frac{\frac{1}{2\sqrt{1+x}}}{1} = \frac{1}{2}$$

(c) $$\lim_{x\to 0}\frac{\sqrt{1+x}-1-x/2}{x^2} \qquad \frac{0}{0}\text{; apply l'Hôpital's Rule.}$$

$$= \lim_{x\to 0}\frac{(1/2)(1+x)^{-1/2}-1/2}{2x} \qquad \text{Still } \frac{0}{0}\text{; apply l'Hôpital's Rule again.}$$

$$= \lim_{x\to 0}\frac{-(1/4)(1+x)^{-3/2}}{2} = -\frac{1}{8} \qquad \text{Not } \frac{0}{0}\text{; limit is found.}$$

(d) $$\lim_{x\to 0}\frac{x-\sin x}{x^3} \qquad \frac{0}{0}\text{; apply l'Hôpital's Rule.}$$

$$= \lim_{x\to 0}\frac{1-\cos x}{3x^2} \qquad \text{Still } \frac{0}{0}\text{; apply l'Hôpital's Rule again.}$$

$$= \lim_{x\to 0}\frac{\sin x}{6x} \qquad \text{Still } \frac{0}{0}\text{; apply l'Hôpital's Rule again.}$$

$$= \lim_{x\to 0}\frac{\cos x}{6} = \frac{1}{6} \qquad \text{Not } \frac{0}{0}\text{; limit is found.}$$

■

Here is a summary of the procedure we followed in Example 1.

Using L'Hôpital's Rule

To find

$$\lim_{x\to a}\frac{f(x)}{g(x)}$$

by l'Hôpital's Rule, we continue to differentiate f and g, so long as we still get the form $0/0$ at $x = a$. But as soon as one or the other of these derivatives is different from zero at $x = a$ we stop differentiating. L'Hôpital's Rule does not apply when either the numerator or denominator has a finite nonzero limit.

EXAMPLE 2 Be careful to apply l'Hôpital's Rule correctly:

$$\lim_{x\to 0}\frac{1-\cos x}{x+x^2} \qquad \frac{0}{0}$$

$$= \lim_{x\to 0}\frac{\sin x}{1+2x} \qquad \text{Not } \frac{0}{0}$$

It is tempting to try to apply l'Hôpital's Rule again, which would result in

$$\lim_{x\to 0}\frac{\cos x}{2} = \frac{1}{2},$$

but this is not the correct limit. L'Hôpital's Rule can be applied only to limits that give indeterminate forms, and $\lim_{x\to 0}(\sin x)/(1+2x)$ does not give an indeterminate form. Instead, this limit is $0/1 = 0$, and the correct answer for the original limit is 0. ■

L'Hôpital's Rule applies to one-sided limits as well.

EXAMPLE 3 In this example the one-sided limits are different.

Recall that ∞ and $+\infty$ mean the same thing.

(a) $\lim\limits_{x\to 0^+} \dfrac{\sin x}{x^2}$ $\qquad \dfrac{0}{0}$

$= \lim\limits_{x\to 0^+} \dfrac{\cos x}{2x} = \infty$ $\qquad$ Positive for $x > 0$

(b) $\lim\limits_{x\to 0^-} \dfrac{\sin x}{x^2}$ $\qquad \dfrac{0}{0}$

$= \lim\limits_{x\to 0^-} \dfrac{\cos x}{2x} = -\infty$ $\qquad$ Negative for $x < 0$ ■

Indeterminate Forms ∞/∞, $\infty \cdot 0$, $\infty - \infty$

Sometimes when we try to evaluate a limit as $x \to a$ by substituting $x = a$ we get an indeterminant form like ∞/∞, $\infty \cdot 0$, or $\infty - \infty$, instead of $0/0$. We first consider the form ∞/∞.

More advanced treatments of calculus prove that l'Hôpital's Rule applies to the indeterminate form ∞/∞, as well as to $0/0$. If $f(x) \to \pm\infty$ and $g(x) \to \pm\infty$ as $x \to a$, then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

provided the limit on the right exists. In the notation $x \to a$, a may be either finite or infinite. Moreover, $x \to a$ may be replaced by the one-sided limits $x \to a^+$ or $x \to a^-$.

EXAMPLE 4 Find the limits of these ∞/∞ forms:

(a) $\lim\limits_{x\to\pi/2} \dfrac{\sec x}{1 + \tan x}$ $\qquad$ (b) $\lim\limits_{x\to\infty} \dfrac{\ln x}{2\sqrt{x}}$ $\qquad$ (c) $\lim\limits_{x\to\infty} \dfrac{e^x}{x^2}$.

Solution

(a) The numerator and denominator are discontinuous at $x = \pi/2$, so we investigate the one-sided limits there. To apply l'Hôpital's Rule, we can choose I to be any open interval with $x = \pi/2$ as an endpoint.

$\lim\limits_{x\to(\pi/2)^-} \dfrac{\sec x}{1 + \tan x}$ $\qquad \dfrac{\infty}{\infty}$ from the left so we apply l'Hôpital's Rule.

$= \lim\limits_{x\to(\pi/2)^-} \dfrac{\sec x \tan x}{\sec^2 x} = \lim\limits_{x\to(\pi/2)^-} \sin x = 1$

The right-hand limit is 1 also, with $(-\infty)/(-\infty)$ as the indeterminate form. Therefore, the two-sided limit is equal to 1.

(b) $\lim\limits_{x\to\infty} \dfrac{\ln x}{2\sqrt{x}} = \lim\limits_{x\to\infty} \dfrac{1/x}{1/\sqrt{x}} = \lim\limits_{x\to\infty} \dfrac{1}{\sqrt{x}} = 0$ $\qquad \dfrac{1/x}{1/\sqrt{x}} = \dfrac{\sqrt{x}}{x} = \dfrac{1}{\sqrt{x}}$

(c) $\lim\limits_{x\to\infty} \dfrac{e^x}{x^2} = \lim\limits_{x\to\infty} \dfrac{e^x}{2x} = \lim\limits_{x\to\infty} \dfrac{e^x}{2} = \infty$ ■

Next we turn our attention to the indeterminate forms $\infty \cdot 0$ and $\infty - \infty$. Sometimes these forms can be handled by using algebra to convert them to a $0/0$ or ∞/∞ form. Here again we do not mean to suggest that $\infty \cdot 0$ or $\infty - \infty$ is a number. They are only notations for functional behaviors when considering limits. Here are examples of how we might work with these indeterminate forms.

EXAMPLE 5 Find the limits of these $\infty \cdot 0$ forms:

(a) $\lim_{x\to\infty}\left(x \sin \frac{1}{x}\right)$ (b) $\lim_{x\to 0^+} \sqrt{x} \ln x$

Solution

(a) $$\lim_{x\to\infty}\left(x \sin \frac{1}{x}\right) = \lim_{h\to 0^+}\left(\frac{1}{h}\sin h\right) = \lim_{h\to 0^+}\frac{\sin h}{h} = 1$$ $\infty \cdot 0$; let $h = 1/x$.

(b) $$\lim_{x\to 0^+}\sqrt{x}\ln x = \lim_{x\to 0^+}\frac{\ln x}{1/\sqrt{x}}$$ $\infty \cdot 0$ converted to ∞/∞

$$= \lim_{x\to 0^+}\frac{1/x}{-1/2x^{3/2}}$$ l'Hôpital's Rule applied

$$= \lim_{x\to 0^+}(-2\sqrt{x}) = 0$$ ■

EXAMPLE 6 Find the limit of this $\infty - \infty$ form:

$$\lim_{x\to 0}\left(\frac{1}{\sin x} - \frac{1}{x}\right).$$

Solution If $x \to 0^+$, then $\sin x \to 0^+$ and

$$\frac{1}{\sin x} - \frac{1}{x} \to \infty - \infty.$$

Similarly, if $x \to 0^-$, then $\sin x \to 0^-$ and

$$\frac{1}{\sin x} - \frac{1}{x} \to -\infty - (-\infty) = -\infty + \infty.$$

Neither form reveals what happens in the limit. To find out, we first combine the fractions:

$$\frac{1}{\sin x} - \frac{1}{x} = \frac{x - \sin x}{x \sin x}.$$ Common denominator is $x \sin x$.

Then we apply l'Hôpital's Rule to the result:

$$\lim_{x\to 0}\left(\frac{1}{\sin x} - \frac{1}{x}\right) = \lim_{x\to 0}\frac{x - \sin x}{x \sin x}$$ $\frac{0}{0}$

$$= \lim_{x\to 0}\frac{1 - \cos x}{\sin x + x\cos x}$$ Still $\frac{0}{0}$

$$= \lim_{x\to 0}\frac{\sin x}{2\cos x - x\sin x} = \frac{0}{2} = 0.$$ ■

Indeterminate Powers

Limits that lead to the indeterminate forms 1^∞, 0^0, and ∞^0 can sometimes be handled by first taking the logarithm of the function. We use l'Hôpital's Rule to find the limit of the logarithm expression and then exponentiate the result to find the original function limit. This procedure is justified by the continuity of the exponential function and Theorem 10 in Section 2.5, and it is formulated as follows. (The formula is also valid for one-sided limits.)

> If $\lim_{x\to a} \ln f(x) = L$, then
>
> $$\lim_{x\to a} f(x) = \lim_{x\to a} e^{\ln f(x)} = e^L.$$
>
> Here a may be either finite or infinite.

EXAMPLE 7 Apply l'Hôpital's Rule to show that $\lim_{x\to 0^+}(1+x)^{1/x} = e$.

Solution The limit leads to the indeterminate form 1^∞. We let $f(x) = (1+x)^{1/x}$ and find $\lim_{x\to 0^+} \ln f(x)$. Since

$$\ln f(x) = \ln(1+x)^{1/x} = \frac{1}{x}\ln(1+x),$$

l'Hôpital's Rule now applies to give

$$\begin{aligned}\lim_{x\to 0^+} \ln f(x) &= \lim_{x\to 0^+} \frac{\ln(1+x)}{x} && \frac{0}{0}\\ &= \lim_{x\to 0^+} \frac{\frac{1}{1+x}}{1} && \text{l'Hôpital's Rule applied}\\ &= \frac{1}{1} = 1.\end{aligned}$$

Therefore, $\lim_{x\to 0^+}(1+x)^{1/x} = \lim_{x\to 0^+} f(x) = \lim_{x\to 0^+} e^{\ln f(x)} = e^1 = e.$ ∎

EXAMPLE 8 Find $\lim_{x\to\infty} x^{1/x}$.

Solution The limit leads to the indeterminate form ∞^0. We let $f(x) = x^{1/x}$ and find $\lim_{x\to\infty} \ln f(x)$. Since

$$\ln f(x) = \ln x^{1/x} = \frac{\ln x}{x},$$

l'Hôpital's Rule gives

$$\begin{aligned}\lim_{x\to\infty} \ln f(x) &= \lim_{x\to\infty} \frac{\ln x}{x} && \frac{\infty}{\infty}\\ &= \lim_{x\to\infty} \frac{1/x}{1} && \text{l'Hôpital's Rule applied}\\ &= \frac{0}{1} = 0.\end{aligned}$$

Therefore $\lim_{x\to\infty} x^{1/x} = \lim_{x\to\infty} f(x) = \lim_{x\to\infty} e^{\ln f(x)} = e^0 = 1.$ ∎

Proof of L'Hôpital's Rule

Before we prove l'Hôpital's Rule, we consider a special case to provide some geometric insight for its reasonableness. Consider the two functions $f(x)$ and $g(x)$ having *continuous* derivatives and satisfying $f(a) = g(a) = 0$, $g'(a) \neq 0$. The graphs of $f(x)$ and $g(x)$, together with their linearizations $y = f'(a)(x-a)$ and $y = g'(a)(x-a)$, are shown in Figure 7.19. We know that near $x = a$, the linearizations provide good approximations to the functions. In fact,

$$f(x) = f'(a)(x-a) + \epsilon_1(x-a) \quad \text{and} \quad g(x) = g'(a)(x-a) + \epsilon_2(x-a)$$

where $\epsilon_1 \to 0$ and $\epsilon_2 \to 0$ as $x \to a$. So, as Figure 7.19 suggests,

$$\begin{aligned}\lim_{x\to a} \frac{f(x)}{g(x)} &= \lim_{x\to a} \frac{f'(a)(x-a) + \epsilon_1(x-a)}{g'(a)(x-a) + \epsilon_2(x-a)}\\ &= \lim_{x\to a} \frac{f'(a) + \epsilon_1}{g'(a) + \epsilon_2} = \frac{f'(a)}{g'(a)} && g'(a) \neq 0\\ &= \lim_{x\to a} \frac{f'(x)}{g'(x)}, && \text{Continuous derivatives}\end{aligned}$$

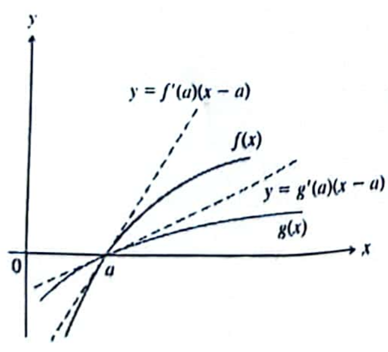


FIGURE 7.19 The two functions in l'Hôpital's Rule, graphed with their linear approximations at $x = a$.

HISTORICAL BIOGRAPHY
Augustin-Louis Cauchy
(1789–1857)

as asserted by l'Hôpital's Rule. We now proceed to a proof of the rule based on the more general assumptions stated in Theorem 5, which do not require that $g'(a) \neq 0$ and that the two functions have *continuous* derivatives.

The proof of l'Hôpital's Rule is based on Cauchy's Mean Value Theorem, an extension of the Mean Value Theorem that involves two functions instead of one. We prove Cauchy's Theorem first and then show how it leads to l'Hôpital's Rule.

When $g(x) = x$, Theorem 6 is the Mean Value Theorem in Chapter 4.

THEOREM 6—Cauchy's Mean Value Theorem Suppose functions f and g are continuous on $[a, b]$ and differentiable throughout (a, b) and also suppose $g'(x) \neq 0$ throughout (a, b). Then there exists a number c in (a, b) at which

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Proof We apply the Mean Value Theorem of Section 4.2 twice. First we use it to show that $g(a) \neq g(b)$. For if $g(b)$ did equal $g(a)$, then the Mean Value Theorem would give

$$g'(c) = \frac{g(b) - g(a)}{b - a} = 0$$

for some c between a and b, which cannot happen because $g'(x) \neq 0$ in (a, b).

We next apply the Mean Value Theorem to the function

$$F(x) = f(x) - f(a) - \frac{f(b) - f(a)}{g(b) - g(a)}[g(x) - g(a)].$$

This function is continuous and differentiable where f and g are, and $F(b) = F(a) = 0$. Therefore, there is a number c between a and b for which $F'(c) = 0$. When expressed in terms of f and g, this equation becomes

$$F'(c) = f'(c) - \frac{f(b) - f(a)}{g(b) - g(a)}[g'(c)] = 0$$

so that

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}. \qquad \blacksquare$$

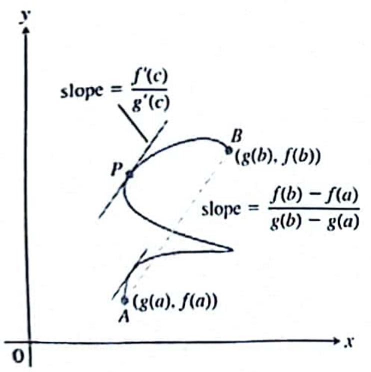


FIGURE 7.20 There is at least one point P on the curve C for which the slope of the tangent to the curve at P is the same as the slope of the secant line joining the points $A(g(a), f(a))$ and $B(g(b), f(b))$.

Cauchy's Mean Value Theorem has a geometric interpretation for a general winding curve C in the plane joining the two points $A = (g(a), f(a))$ and $B = (g(b), f(b))$. In Chapter 11 you will learn how the curve C can be formulated to show that there is at least one point P on the curve for which the tangent to the curve at P is parallel to the secant line joining the points A and B. The slope of that tangent line turns out to be the quotient f'/g' evaluated at the number c in the interval (a, b), which is the left-hand side of the equation in Theorem 6. Because the slope of the secant line joining A and B is

$$\frac{f(b) - f(a)}{g(b) - g(a)},$$

the equation in Cauchy's Mean Value Theorem says that the slope of the tangent line equals the slope of the secant line. This geometric interpretation is shown in Figure 7.20. Notice from the figure that it is possible for more than one point on the curve C to have a tangent line that is parallel to the secant line joining A and B.

Proof of l'Hôpital's Rule We first establish the limit equation for the case $x \to a^+$. The method needs almost no change to apply to $x \to a^-$, and the combination of these two cases establishes the result.

7.6 Inverse Trigonometric Functions

Inverse trigonometric functions arise when we want to calculate angles from side measurements in triangles. They also provide useful antiderivatives and appear frequently in the solutions of differential equations. This section shows how these functions are defined, graphed, and evaluated, how their derivatives are computed, and why they appear as important antiderivatives.

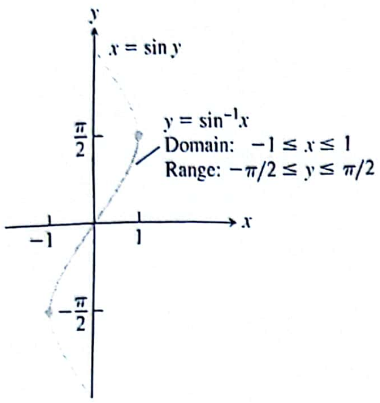
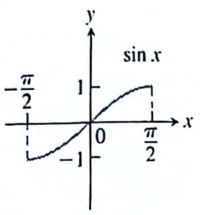


FIGURE 7.21 The graph of $y = \sin^{-1} x$.

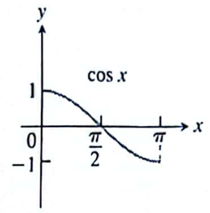
Defining the Inverses

The six basic trigonometric functions are not one-to-one (their values repeat periodically). However, we can restrict their domains to intervals on which they are one-to-one. The sine function increases from -1 at $x = -\pi/2$ to $+1$ at $x = \pi/2$. By restricting its domain to the interval $[-\pi/2, \pi/2]$ we make it one-to-one, so that it has an inverse $\sin^{-1} x$ (Figure 7.21). Similar domain restrictions can be applied to all six trigonometric functions.

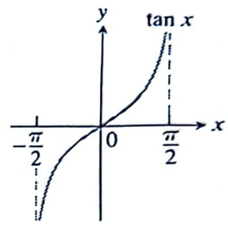
Domain restrictions that make the trigonometric functions one-to-one



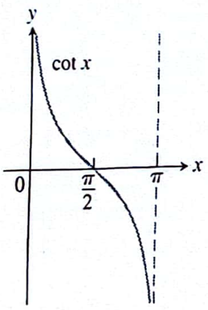
$y = \sin x$
Domain: $[-\pi/2, \pi/2]$
Range: $[-1, 1]$



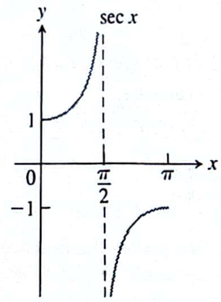
$y = \cos x$
Domain: $[0, \pi]$
Range: $[-1, 1]$



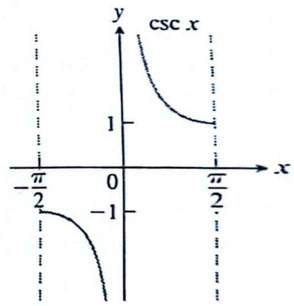
$y = \tan x$
Domain: $(-\pi/2, \pi/2)$
Range: $(-\infty, \infty)$



$y = \cot x$
Domain: $(0, \pi)$
Range: $(-\infty, \infty)$



$y = \sec x$
Domain: $[0, \pi/2) \cup (\pi/2, \pi]$
Range: $(-\infty, -1] \cup [1, \infty)$



$y = \csc x$
Domain: $[-\pi/2, 0) \cup (0, \pi/2]$
Range: $(-\infty, -1] \cup [1, \infty)$

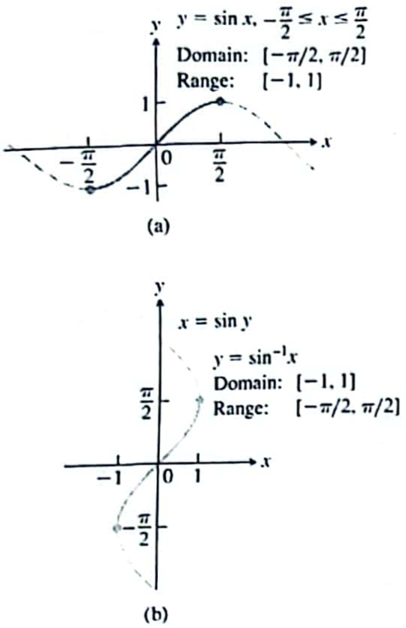


FIGURE 7.22 The graphs of (a) $y = \sin x, -\pi/2 \le x \le \pi/2$, and (b) its inverse, $y = \sin^{-1} x$. The graph of $\sin^{-1} x$, obtained by reflection across the line $y = x$, is a portion of the curve $x = \sin y$.

Since these restricted functions are now one-to-one, they have inverses, which we denote by

$$
\begin{array}{lll}
y = \sin^{-1} x & \text{or} & y = \arcsin x \\
y = \cos^{-1} x & \text{or} & y = \arccos x \\
y = \tan^{-1} x & \text{or} & y = \arctan x \\
y = \cot^{-1} x & \text{or} & y = \operatorname{arccot} x \\
y = \sec^{-1} x & \text{or} & y = \operatorname{arcsec} x \\
y = \csc^{-1} x & \text{or} & y = \operatorname{arccsc} x
\end{array}
$$

These equations are read "y equals the arcsine of x" or "y equals arcsin x" and so on.

Caution The -1 in the expressions for the inverse means "inverse." It does *not* mean reciprocal. For example, the *reciprocal* of $\sin x$ is $(\sin x)^{-1} = 1/\sin x = \csc x$.

The graphs of the six inverse trigonometric functions are obtained by reflecting the graphs of the restricted trigonometric functions through the line $y = x$. Figure 7.22b shows the graph of $y = \sin^{-1} x$ and Figure 7.23 shows the graphs of all six functions. We now take a closer look at these functions.

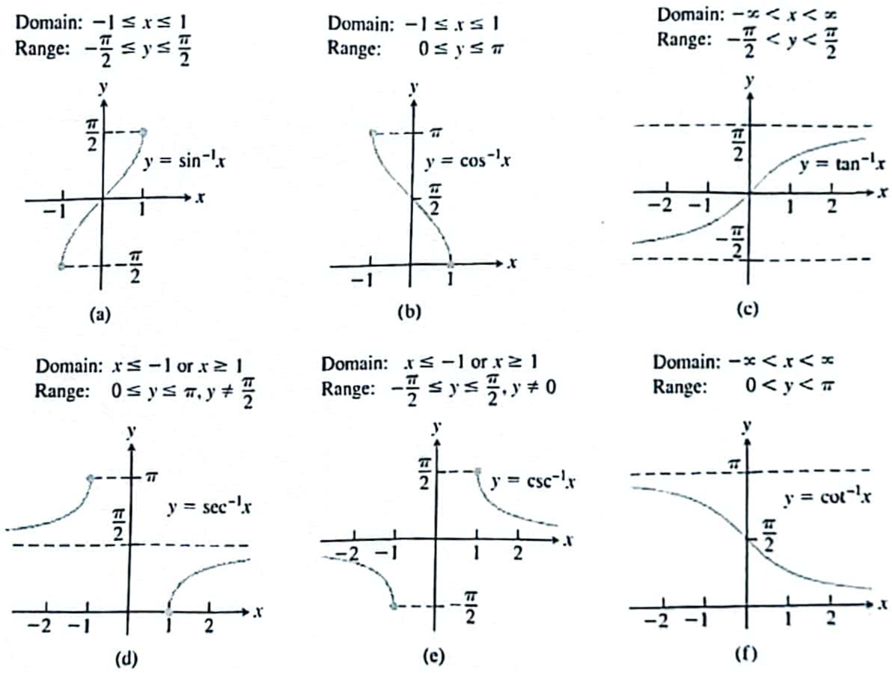
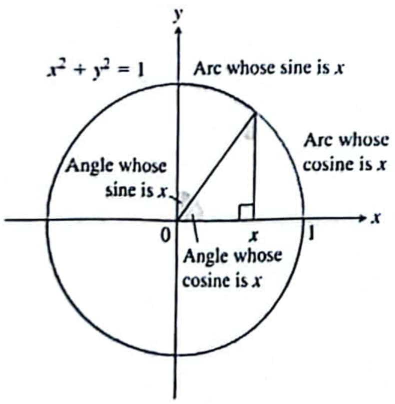


FIGURE 7.23 Graphs of the six basic inverse trigonometric functions.

The "Arc" in Arcsine and Arccosine
For a unit circle and radian angles, the arc length equation $s = r\theta$ becomes $s = \theta$, so central angles and the arcs they subtend have the same measure. If $x = \sin y$, then, in addition to being the angle whose sine is x, y is also the length of arc on the unit circle that subtends an angle whose sine is x. So we call y "the arc whose sine is x."



The Arcsine and Arccosine Functions

We define the arcsine and arccosine as functions whose values are angles (measured in radians) that belong to restricted domains of the sine and cosine functions.

DEFINITION

$y = \sin^{-1} x$ is the number in $[-\pi/2, \pi/2]$ for which $\sin y = x$.
$y = \cos^{-1} x$ is the number in $[0, \pi]$ for which $\cos y = x$.

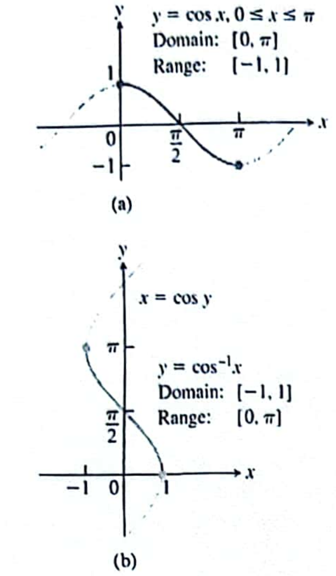


FIGURE 7.24 The graphs of (a) $y = \cos x$, $0 \le x \le \pi$, and (b) its inverse, $y = \cos^{-1}x$. The graph of $\cos^{-1}x$, obtained by reflection across the line $y = x$, is a portion of the curve $x = \cos y$.

The graph of $y = \sin^{-1}x$ (Figure 7.22b) is symmetric about the origin (it lies along the graph of $x = \sin y$). The arcsine is therefore an odd function:

$$\sin^{-1}(-x) = -\sin^{-1}x. \tag{1}$$

The graph of $y = \cos^{-1}x$ (Figure 7.24b) has no such symmetry.

EXAMPLE 1 Evaluate (a) $\sin^{-1}\left(\frac{\sqrt{3}}{2}\right)$ and (b) $\cos^{-1}\left(-\frac{1}{2}\right)$.

Solution

(a) We see that

$$\sin^{-1}\left(\frac{\sqrt{3}}{2}\right) = \frac{\pi}{3}$$

because $\sin(\pi/3) = \sqrt{3}/2$ and $\pi/3$ belongs to the range $[-\pi/2, \pi/2]$ of the arcsine function. See Figure 7.25a.

(b) We have

$$\cos^{-1}\left(-\frac{1}{2}\right) = \frac{2\pi}{3}$$

because $\cos(2\pi/3) = -1/2$ and $2\pi/3$ belongs to the range $[0, \pi]$ of the arccosine function. See Figure 7.25b. ■

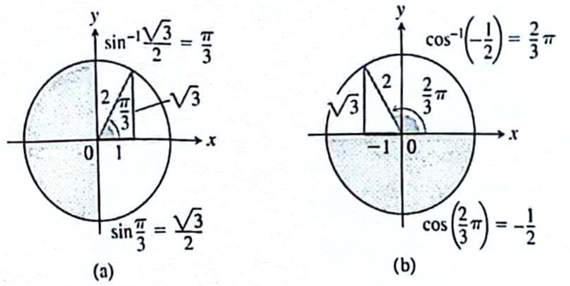


FIGURE 7.25 Values of the arcsine and arccosine functions (Example 1).

Using the same procedure illustrated in Example 1, we can create the following table of common values for the arcsine and arccosine functions.

x	$\sin^{-1}x$	$\cos^{-1}x$
$\sqrt{3}/2$	$\pi/3$	$\pi/6$
$\sqrt{2}/2$	$\pi/4$	$\pi/4$
$1/2$	$\pi/6$	$\pi/3$
$-1/2$	$-\pi/6$	$2\pi/3$
$-\sqrt{2}/2$	$-\pi/4$	$3\pi/4$
$-\sqrt{3}/2$	$-\pi/3$	$5\pi/6$

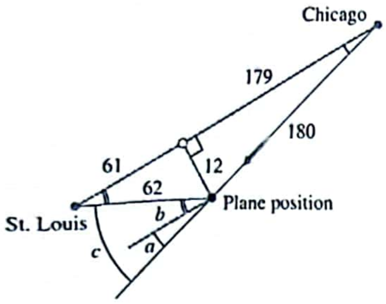


FIGURE 7.26 Diagram for drift correction (Example 2), with distances rounded to the nearest mile (drawing not to scale).

EXAMPLE 2 During a 240 mi airplane flight from Chicago to St. Louis, after flying 180 mi the navigator determines that the plane is 12 mi off course, as shown in Figure 7.26. Find the angle a for a course parallel to the original correct course, the angle b, and the drift correction angle $c = a + b$.

Solution From the Pythagorean theorem and given information, we compute an approximate hypothetical flight distance of 179 mi, had the plane been flying along the original correct course (see Figure 7.26). Knowing the flight distance from Chicago to St. Louis, we next calculate the remaining leg of the original course to be 61 mi. Applying the Pythagorean theorem again then gives an approximate distance of 62 mi from the position of the plane to St. Louis. Finally, from Figure 7.26, we see that $180 \sin a = 12$ and $62 \sin b = 12$, so

$$a = \sin^{-1}\frac{12}{180} \approx 0.067 \text{ radian} \approx 3.8^\circ$$

$$b = \sin^{-1}\frac{12}{62} \approx 0.195 \text{ radian} \approx 11.2^\circ$$

$$c = a + b \approx 15^\circ. \quad \blacksquare$$

Identities Involving Arcsine and Arccosine

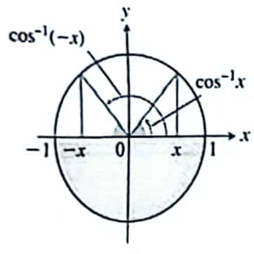


FIGURE 7.27 $\cos^{-1}x$ and $\cos^{-1}(-x)$ are supplementary angles (so their sum is π).

As we can see from Figure 7.27, the arccosine of x satisfies the identity

$$\cos^{-1}x + \cos^{-1}(-x) = \pi, \tag{2}$$

or

$$\cos^{-1}(-x) = \pi - \cos^{-1}x. \tag{3}$$

Also, we can see from the triangle in Figure 7.28 that for $x > 0$,

$$\sin^{-1}x + \cos^{-1}x = \pi/2. \tag{4}$$

Equation (4) holds for the other values of x in $[-1, 1]$ as well, but we cannot conclude this from the triangle in Figure 7.28. It is, however, a consequence of Equations (1) and (3) (Exercise 113).

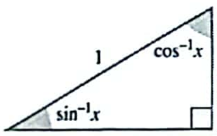


FIGURE 7.28 $\sin^{-1}x$ and $\cos^{-1}x$ are complementary angles (so their sum is $\pi/2$).

Inverses of tan x, cot x, sec x, and csc x

The arctangent of x is a radian angle whose tangent is x. The arccotangent of x is an angle whose cotangent is x, and so forth. The angles belong to the restricted domains of the tangent, cotangent, secant, and cosecant functions.

DEFINITIONS

$y = \tan^{-1}x$ is the number in $(-\pi/2, \pi/2)$ for which $\tan y = x$.

$y = \cot^{-1}x$ is the number in $(0, \pi)$ for which $\cot y = x$.

$y = \sec^{-1}x$ is the number in $[0, \pi/2) \cup (\pi/2, \pi]$ for which $\sec y = x$.

$y = \csc^{-1}x$ is the number in $[-\pi/2, 0) \cup (0, \pi/2]$ for which $\csc y = x$.

We use open or half-open intervals to avoid values for which the tangent, cotangent, secant, and cosecant functions are undefined. (See Figure 7.23.)

The graph of $y = \tan^{-1}x$ is symmetric about the origin because it is a branch of the graph $x = \tan y$ that is symmetric about the origin (Figure 7.23c). Algebraically this means that

$$\tan^{-1}(-x) = -\tan^{-1}x;$$

the arctangent is an odd function. The graph of $y = \cot^{-1}x$ has no such symmetry (Figure 7.23f). Notice from Figure 7.23c that the graph of the arctangent function has two horizontal asymptotes: one at $y = \pi/2$ and the other at $y = -\pi/2$.

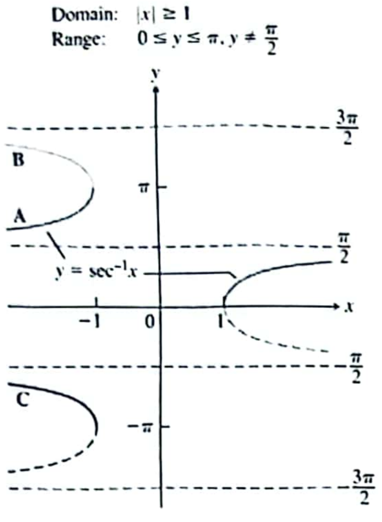


FIGURE 7.29 There are several logical choices for the left-hand branch of $y = \sec^{-1}x$. With choice A, $\sec^{-1}x = \cos^{-1}(1/x)$, a useful identity employed by many calculators.

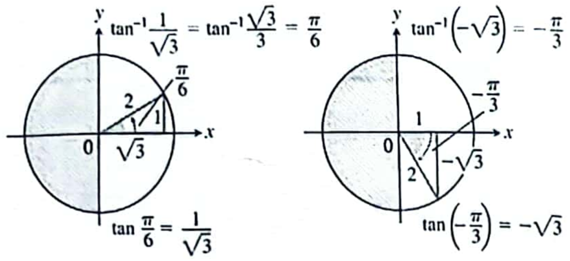
The inverses of the restricted forms of $\sec x$ and $\csc x$ are chosen to be the functions graphed in Figures 7.23d and 7.23e.

Caution There is no general agreement about how to define $\sec^{-1}x$ for negative values of x. We chose angles in the second quadrant between $\pi/2$ and π. This choice makes $\sec^{-1}x = \cos^{-1}(1/x)$. It also makes $\sec^{-1}x$ an increasing function on each interval of its domain. Some texts choose $\sec^{-1}x$ to lie in $[-\pi, -\pi/2)$ for $x < 0$ and some texts choose it to lie in $[\pi, 3\pi/2)$ (Figure 7.29). These choices simplify the formula for the derivative (our formula needs absolute value signs) but fail to satisfy the computational equation $\sec^{-1}x = \cos^{-1}(1/x)$. From this, we can derive the identity

$$\sec^{-1}x = \cos^{-1}\left(\frac{1}{x}\right) = \frac{\pi}{2} - \sin^{-1}\left(\frac{1}{x}\right) \tag{5}$$

by applying Equation (4).

EXAMPLE 3 The accompanying figures show two values of $\tan^{-1}x$.



x	$\tan^{-1}x$
$\sqrt{3}$	$\pi/3$
1	$\pi/4$
$\sqrt{3}/3$	$\pi/6$
$-\sqrt{3}/3$	$-\pi/6$
-1	$-\pi/4$
$-\sqrt{3}$	$-\pi/3$

The angles come from the first and fourth quadrants because the range of $\tan^{-1}x$ is $(-\pi/2, \pi/2)$. ■

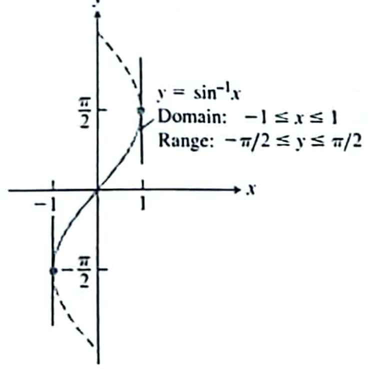


FIGURE 7.30 The graph of $y = \sin^{-1}x$ has vertical tangents at $x = -1$ and $x = 1$.

The Derivative of $y = \sin^{-1}u$

We know that the function $x = \sin y$ is differentiable in the interval $-\pi/2 < y < \pi/2$ and that its derivative, the cosine, is positive there. Theorem 1 in Section 7.1 therefore assures us that the inverse function $y = \sin^{-1}x$ is differentiable throughout the interval $-1 < x < 1$. We cannot expect it to be differentiable at $x = 1$ or $x = -1$ because the tangents to the graph are vertical at these points (see Figure 7.30).

We find the derivative of $y = \sin^{-1} x$ by applying Theorem 1 with $f(x) = \sin x$ and $f^{-1}(x) = \sin^{-1} x$:

$$\begin{aligned}
(f^{-1})'(x) &= \frac{1}{f'(f^{-1}(x))} && \text{Theorem 1} \\
&= \frac{1}{\cos(\sin^{-1} x)} && f'(u) = \cos u \\
&= \frac{1}{\sqrt{1 - \sin^2(\sin^{-1} x)}} && \cos u = \sqrt{1 - \sin^2 u} \\
&= \frac{1}{\sqrt{1 - x^2}}. && \sin(\sin^{-1} x) = x
\end{aligned}$$

If u is a differentiable function of x with $|u| < 1$, we apply the Chain Rule to get the general formula

$$\frac{d}{dx}(\sin^{-1} u) = \frac{1}{\sqrt{1 - u^2}} \frac{du}{dx}, \qquad |u| < 1.$$

EXAMPLE 4 Using the Chain Rule, we calculate the derivative

$$\frac{d}{dx}(\sin^{-1} x^2) = \frac{1}{\sqrt{1 - (x^2)^2}} \cdot \frac{d}{dx}(x^2) = \frac{2x}{\sqrt{1 - x^4}}.$$ ■

The Derivative of $y = \tan^{-1} u$

We find the derivative of $y = \tan^{-1} x$ by applying Theorem 1 with $f(x) = \tan x$ and $f^{-1}(x) = \tan^{-1} x$. Theorem 1 can be applied because the derivative of $\tan x$ is positive for $-\pi/2 < x < \pi/2$:

$$\begin{aligned}
(f^{-1})'(x) &= \frac{1}{f'(f^{-1}(x))} && \text{Theorem 1} \\
&= \frac{1}{\sec^2(\tan^{-1} x)} && f'(u) = \sec^2 u \\
&= \frac{1}{1 + \tan^2(\tan^{-1} x)} && \sec^2 u = 1 + \tan^2 u \\
&= \frac{1}{1 + x^2}. && \tan(\tan^{-1} x) = x
\end{aligned}$$

The derivative is defined for all real numbers. If u is a differentiable function of x, we get the Chain Rule form:

$$\frac{d}{dx}(\tan^{-1} u) = \frac{1}{1 + u^2} \frac{du}{dx}.$$

The Derivative of $y = \sec^{-1} u$

Since the derivative of $\sec x$ is positive for $0 < x < \pi/2$ and $\pi/2 < x < \pi$, Theorem 1 says that the inverse function $y = \sec^{-1} x$ is differentiable. Instead of applying the formula in Theorem 1 directly, we find the derivative of $y = \sec^{-1} x$, $|x| > 1$, using implicit differentiation and the Chain Rule as follows:

$$
\begin{aligned}
y &= \sec^{-1} x \\
\sec y &= x && \text{Inverse function relationship} \\
\frac{d}{dx}(\sec y) &= \frac{d}{dx}x && \text{Differentiate both sides.} \\
\sec y \tan y \frac{dy}{dx} &= 1 && \text{Chain Rule} \\
\frac{dy}{dx} &= \frac{1}{\sec y \tan y}. && \text{Since } |x| > 1,\ y \text{ lies in } (0, \pi/2) \cup (\pi/2, \pi) \text{ and } \sec y \tan y \neq 0.
\end{aligned}
$$

To express the result in terms of x, we use the relationships

$$\sec y = x \qquad \text{and} \qquad \tan y = \pm\sqrt{\sec^2 y - 1} = \pm\sqrt{x^2 - 1}$$

to get

$$\frac{dy}{dx} = \pm\frac{1}{x\sqrt{x^2 - 1}}.$$

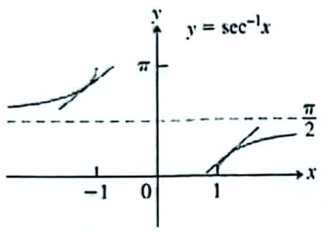


FIGURE 7.31 The slope of the curve $y = \sec^{-1} x$ is positive for both $x < -1$ and $x > 1$.

Can we do anything about the $\pm$ sign? A glance at Figure 7.31 shows that the slope of the graph $y = \sec^{-1} x$ is always positive. Thus,

$$\frac{d}{dx}\sec^{-1} x = \begin{cases} +\dfrac{1}{x\sqrt{x^2 - 1}} & \text{if } x > 1 \\ -\dfrac{1}{x\sqrt{x^2 - 1}} & \text{if } x < -1. \end{cases}$$

With the absolute value symbol, we can write a single expression that eliminates the "$\pm$" ambiguity:

$$\frac{d}{dx}\sec^{-1} x = \frac{1}{|x|\sqrt{x^2 - 1}}.$$

If u is a differentiable function of x with $|u| > 1$, we have the formula

$$\boxed{\frac{d}{dx}(\sec^{-1} u) = \frac{1}{|u|\sqrt{u^2 - 1}}\frac{du}{dx}, \qquad |u| > 1.}$$

EXAMPLE 5 Using the Chain Rule and derivative of the arcsecant function, we find

$$
\begin{aligned}
\frac{d}{dx}\sec^{-1}(5x^4) &= \frac{1}{|5x^4|\sqrt{(5x^4)^2 - 1}}\frac{d}{dx}(5x^4) \\
&= \frac{1}{5x^4\sqrt{25x^8 - 1}}(20x^3) && 5x^4 > 1 > 0 \\
&= \frac{4}{x\sqrt{25x^8 - 1}}.
\end{aligned}
$$

∎

Derivatives of the Other Three Inverse Trigonometric Functions

We could use the same techniques to find the derivatives of the other three inverse trigonometric functions—arccosine, arccotangent, and arccosecant—but there is an easier way, thanks to the following identities.

Inverse Function–Inverse Cofunction Identities

$$\cos^{-1}x = \pi/2 - \sin^{-1}x$$
$$\cot^{-1}x = \pi/2 - \tan^{-1}x$$
$$\csc^{-1}x = \pi/2 - \sec^{-1}x$$

We saw the first of these identities in Equation (4). The others are derived in a similar way. It follows easily that the derivatives of the inverse cofunctions are the negatives of the derivatives of the corresponding inverse functions. For example, the derivative of $\cos^{-1}x$ is calculated as follows:

$$\begin{aligned} \frac{d}{dx}(\cos^{-1}x) &= \frac{d}{dx}\left(\frac{\pi}{2} - \sin^{-1}x\right) && \text{Identity} \\ &= -\frac{d}{dx}(\sin^{-1}x) \\ &= -\frac{1}{\sqrt{1-x^2}}. && \text{Derivative of arcsine} \end{aligned}$$

The derivatives of the inverse trigonometric functions are summarized in Table 7.3.

TABLE 7.3 Derivatives of the inverse trigonometric functions

1. $\dfrac{d(\sin^{-1}u)}{dx} = \dfrac{1}{\sqrt{1-u^2}}\dfrac{du}{dx}, \quad |u| < 1$

2. $\dfrac{d(\cos^{-1}u)}{dx} = -\dfrac{1}{\sqrt{1-u^2}}\dfrac{du}{dx}, \quad |u| < 1$

3. $\dfrac{d(\tan^{-1}u)}{dx} = \dfrac{1}{1+u^2}\dfrac{du}{dx}$

4. $\dfrac{d(\cot^{-1}u)}{dx} = -\dfrac{1}{1+u^2}\dfrac{du}{dx}$

5. $\dfrac{d(\sec^{-1}u)}{dx} = \dfrac{1}{|u|\sqrt{u^2-1}}\dfrac{du}{dx}, \quad |u| > 1$

6. $\dfrac{d(\csc^{-1}u)}{dx} = -\dfrac{1}{|u|\sqrt{u^2-1}}\dfrac{du}{dx}, \quad |u| > 1$

Integration Formulas

The derivative formulas in Table 7.3 yield three useful integration formulas in Table 7.4. The formulas are readily verified by differentiating the functions on the right-hand sides.

TABLE 7.4 Integrals evaluated with inverse trigonometric functions

The following formulas hold for any constant $a > 0$.

1. $\displaystyle\int \frac{du}{\sqrt{a^2 - u^2}} = \sin^{-1}\left(\frac{u}{a}\right) + C$ (Valid for $u^2 < a^2$)

2. $\displaystyle\int \frac{du}{a^2 + u^2} = \frac{1}{a}\tan^{-1}\left(\frac{u}{a}\right) + C$ (Valid for all u)

3. $\displaystyle\int \frac{du}{u\sqrt{u^2 - a^2}} = \frac{1}{a}\sec^{-1}\left|\frac{u}{a}\right| + C$ (Valid for $|u| > a > 0$)

The derivative formulas in Table 7.3 have $a = 1$, but in most integrations $a \neq 1$, and the formulas in Table 7.4 are more useful.

EXAMPLE 6 These examples illustrate how we use Table 7.4.

(a)
$$\int_{\sqrt{2}/2}^{\sqrt{3}/2} \frac{dx}{\sqrt{1 - x^2}} = \sin^{-1} x \Big]_{\sqrt{2}/2}^{\sqrt{3}/2} \qquad a = 1,\ u = x \text{ in Table 7.4, Formula 1}$$
$$= \sin^{-1}\left(\frac{\sqrt{3}}{2}\right) - \sin^{-1}\left(\frac{\sqrt{2}}{2}\right) = \frac{\pi}{3} - \frac{\pi}{4} = \frac{\pi}{12}$$

(b)
$$\int \frac{dx}{\sqrt{3 - 4x^2}} = \frac{1}{2}\int \frac{du}{\sqrt{a^2 - u^2}} \qquad a = \sqrt{3},\ u = 2x, \text{ and } du/2 = dx$$
$$= \frac{1}{2}\sin^{-1}\left(\frac{u}{a}\right) + C \qquad \text{Table 7.4, Formula 1}$$
$$= \frac{1}{2}\sin^{-1}\left(\frac{2x}{\sqrt{3}}\right) + C$$

(c)
$$\int \frac{dx}{\sqrt{e^{2x} - 6}} = \int \frac{du/u}{\sqrt{u^2 - a^2}} \qquad u = e^x,\ du = e^x\,dx,\ dx = du/e^x = du/u,\ a = \sqrt{6}$$
$$= \int \frac{du}{u\sqrt{u^2 - a^2}}$$
$$= \frac{1}{a}\sec^{-1}\left|\frac{u}{a}\right| + C \qquad \text{Table 7.4, Formula 3}$$
$$= \frac{1}{\sqrt{6}}\sec^{-1}\left(\frac{e^x}{\sqrt{6}}\right) + C$$

■

EXAMPLE 7 Evaluate

(a) $\displaystyle\int \frac{dx}{\sqrt{4x - x^2}}$ (b) $\displaystyle\int \frac{dx}{4x^2 + 4x + 2}$

Solution

(a) The expression $\sqrt{4x - x^2}$ does not match any of the formulas in Table 7.4, so we first rewrite $4x - x^2$ by completing the square:

$$4x - x^2 = -(x^2 - 4x) = -(x^2 - 4x + 4) + 4 = 4 - (x - 2)^2.$$

Then we substitute $a = 2$, $u = x - 2$, and $du = dx$ to get

$$\int \frac{dx}{\sqrt{4x - x^2}} = \int \frac{dx}{\sqrt{4 - (x-2)^2}}$$

$$= \int \frac{du}{\sqrt{a^2 - u^2}} \qquad a = 2,\ u = x - 2, \text{ and } du = dx$$

$$= \sin^{-1}\left(\frac{u}{a}\right) + C \qquad \text{Table 7.4, Formula 1}$$

$$= \sin^{-1}\left(\frac{x-2}{2}\right) + C$$

(b) We complete the square on the binomial $4x^2 + 4x$:

$$4x^2 + 4x + 2 = 4(x^2 + x) + 2 = 4\left(x^2 + x + \frac{1}{4}\right) + 2 - \frac{4}{4}$$

$$= 4\left(x + \frac{1}{2}\right)^2 + 1 = (2x + 1)^2 + 1.$$

Then,

$$\int \frac{dx}{4x^2 + 4x + 2} = \int \frac{dx}{(2x+1)^2 + 1} = \frac{1}{2}\int \frac{du}{u^2 + a^2} \qquad a = 1,\ u = 2x + 1, \text{ and } du/2 = dx$$

$$= \frac{1}{2} \cdot \frac{1}{a}\tan^{-1}\left(\frac{u}{a}\right) + C \qquad \text{Table 7.4, Formula 2}$$

$$= \frac{1}{2}\tan^{-1}(2x + 1) + C \qquad a = 1,\ u = 2x + 1$$ ■

7.7 Hyperbolic Functions

The hyperbolic functions are formed by taking combinations of the two exponential functions e^x and e^{-x}. The hyperbolic functions simplify many mathematical expressions and occur frequently in mathematical and engineering applications. In this section we give a brief introduction to these functions, their graphs, their derivatives, their integrals, and their inverse functions.

Definitions and Identities

The hyperbolic sine and hyperbolic cosine functions are defined by the equations

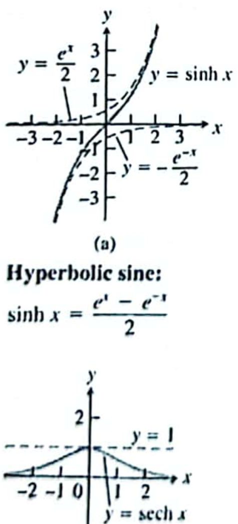
$$\sinh x = \frac{e^x - e^{-x}}{2} \quad \text{and} \quad \cosh x = \frac{e^x + e^{-x}}{2}.$$

We pronounce $\sinh x$ as "cinch x," rhyming with "pinch x," and $\cosh x$ as "kosh x," rhyming with "gosh x." From this basic pair, we define the hyperbolic tangent, cotangent, secant, and cosecant functions. The defining equations and graphs of these functions are shown in Table 7.5. We will see that the hyperbolic functions bear many similarities to the trigonometric functions after which they are named.

Hyperbolic functions satisfy the identities in Table 7.6. Except for differences in sign, these resemble identities we know for the trigonometric functions. The identities are proved directly from the definitions, as we show here for the second one:

$$\begin{aligned} 2\sinh x \cosh x &= 2\left(\frac{e^x - e^{-x}}{2}\right)\left(\frac{e^x + e^{-x}}{2}\right) \\ &= \frac{e^{2x} - e^{-2x}}{2} \\ &= \sinh 2x. \end{aligned}$$

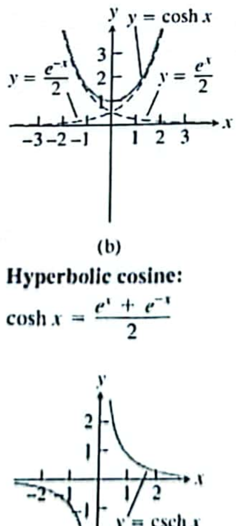
TABLE 7.5 The six basic hyperbolic functions



(a)

Hyperbolic sine:

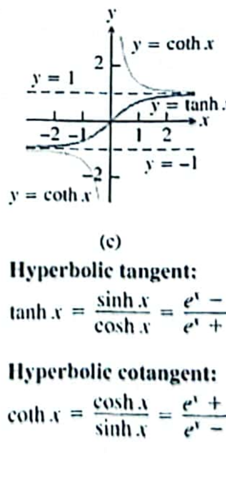
$$\sinh x = \frac{e^x - e^{-x}}{2}$$



(b)

Hyperbolic cosine:

$$\cosh x = \frac{e^x + e^{-x}}{2}$$



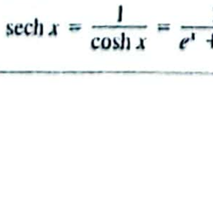
(c)

Hyperbolic tangent:

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Hyperbolic cotangent:

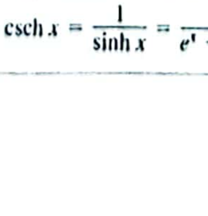
$$\coth x = \frac{\cosh x}{\sinh x} = \frac{e^x + e^{-x}}{e^x - e^{-x}}$$



(d)

Hyperbolic secant:

$$\operatorname{sech} x = \frac{1}{\cosh x} = \frac{2}{e^x + e^{-x}}$$



(e)

Hyperbolic cosecant:

$$\operatorname{csch} x = \frac{1}{\sinh x} = \frac{2}{e^x - e^{-x}}$$

TABLE 7.6 Identities for hyperbolic functions

$\cosh^2 x - \sinh^2 x = 1$
$\sinh 2x = 2\sinh x \cosh x$
$\cosh 2x = \cosh^2 x + \sinh^2 x$
$\cosh^2 x = \frac{\cosh 2x + 1}{2}$
$\sinh^2 x = \frac{\cosh 2x - 1}{2}$
$\tanh^2 x = 1 - \operatorname{sech}^2 x$
$\coth^2 x = 1 + \operatorname{csch}^2 x$

The other identities are obtained similarly, by substituting in the definitions of the hyperbolic functions and using algebra. Like many standard functions, hyperbolic functions and their inverses are easily evaluated with calculators, which often have special keys for that purpose.

For any real number u, we know the point with coordinates $(\cos u, \sin u)$ lies on the unit circle $x^2 + y^2 = 1$. So the trigonometric functions are sometimes called the *circular* functions. Because of the first identity

$$\cosh^2 u - \sinh^2 u = 1,$$

with u substituted for x in Table 7.6, the point having coordinates $(\cosh u, \sinh u)$ lies on the right-hand branch of the hyperbola $x^2 - y^2 = 1$. This is where the *hyperbolic* functions get their names (see Exercise 86).

Hyperbolic functions are useful in finding integrals, which we will see in Chapter 8. They play an important role in science and engineering as well. The hyperbolic cosine describes the shape of a hanging cable or wire that is strung between two points at the same height and hanging freely (see Exercise 83). The shape of the St. Louis Arch is an inverted hyperbolic cosine. The hyperbolic tangent occurs in the formula for the velocity of an ocean wave moving over water having a constant depth, and the inverse hyperbolic tangent describes how relative velocities sum according to Einstein's Law in the Special Theory of Relativity.

TABLE 7.7 Derivatives of hyperbolic functions

$$\frac{d}{dx}(\sinh u) = \cosh u \frac{du}{dx}$$

$$\frac{d}{dx}(\cosh u) = \sinh u \frac{du}{dx}$$

$$\frac{d}{dx}(\tanh u) = \operatorname{sech}^2 u \frac{du}{dx}$$

$$\frac{d}{dx}(\coth u) = -\operatorname{csch}^2 u \frac{du}{dx}$$

$$\frac{d}{dx}(\operatorname{sech} u) = -\operatorname{sech} u \tanh u \frac{du}{dx}$$

$$\frac{d}{dx}(\operatorname{csch} u) = -\operatorname{csch} u \coth u \frac{du}{dx}$$

Derivatives and Integrals of Hyperbolic Functions

The six hyperbolic functions, being rational combinations of the differentiable functions e^x and e^{-x}, have derivatives at every point at which they are defined (Table 7.7). Again, there are similarities with trigonometric functions.

The derivative formulas are derived from the derivative of e^u:

$$\begin{aligned}
\frac{d}{dx}(\sinh u) &= \frac{d}{dx}\left(\frac{e^u - e^{-u}}{2}\right) && \text{Definition of } \sinh u \\
&= \frac{e^u\, du/dx + e^{-u}\, du/dx}{2} && \text{Derivative of } e^u \\
&= \cosh u \frac{du}{dx}. && \text{Definition of } \cosh u
\end{aligned}$$

This gives the first derivative formula. From the definition, we can calculate the derivative of the hyperbolic cosecant function, as follows:

$$\begin{aligned}
\frac{d}{dx}(\operatorname{csch} u) &= \frac{d}{dx}\left(\frac{1}{\sinh u}\right) && \text{Definition of } \operatorname{csch} u \\
&= -\frac{\cosh u}{\sinh^2 u}\frac{du}{dx} && \text{Quotient Rule for derivatives} \\
&= -\frac{1}{\sinh u}\frac{\cosh u}{\sinh u}\frac{du}{dx} && \text{Rearrange terms.} \\
&= -\operatorname{csch} u \coth u \frac{du}{dx} && \text{Definitions of } \operatorname{csch} u \text{ and } \coth u
\end{aligned}$$

TABLE 7.8 Integral formulas for hyperbolic functions

$$\int \sinh u\, du = \cosh u + C$$

$$\int \cosh u\, du = \sinh u + C$$

$$\int \operatorname{sech}^2 u\, du = \tanh u + C$$

$$\int \operatorname{csch}^2 u\, du = -\coth u + C$$

$$\int \operatorname{sech} u \tanh u\, du = -\operatorname{sech} u + C$$

$$\int \operatorname{csch} u \coth u\, du = -\operatorname{csch} u + C$$

The other formulas in Table 7.7 are obtained similarly.

The derivative formulas lead to the integral formulas in Table 7.8.

EXAMPLE 1 We illustrate the derivative and integral formulas.

(a) $$\frac{d}{dt}\left(\tanh\sqrt{1+t^2}\right) = \operatorname{sech}^2\sqrt{1+t^2}\cdot\frac{d}{dt}\left(\sqrt{1+t^2}\right)$$
$$= \frac{t}{\sqrt{1+t^2}}\operatorname{sech}^2\sqrt{1+t^2}$$

(b) $$\int \coth 5x\,dx = \int\frac{\cosh 5x}{\sinh 5x}dx = \frac{1}{5}\int\frac{du}{u} \qquad u = \sinh 5x,\; du = 5\cosh 5x\,dx$$
$$= \frac{1}{5}\ln|u| + C = \frac{1}{5}\ln|\sinh 5x| + C$$

(c) $$\int_0^1 \sinh^2 x\,dx = \int_0^1 \frac{\cosh 2x - 1}{2}dx \qquad \text{Table 7.6}$$
$$= \frac{1}{2}\int_0^1 (\cosh 2x - 1)\,dx = \frac{1}{2}\left[\frac{\sinh 2x}{2} - x\right]_0^1$$
$$= \frac{\sinh 2}{4} - \frac{1}{2} \approx 0.40672 \qquad \text{Evaluate with a calculator.}$$

(d) $$\int_0^{\ln 2} 4e^x\sinh x\,dx = \int_0^{\ln 2} 4e^x\frac{e^x - e^{-x}}{2}dx = \int_0^{\ln 2}(2e^{2x} - 2)\,dx$$
$$= \left[e^{2x} - 2x\right]_0^{\ln 2} = (e^{2\ln 2} - 2\ln 2) - (1 - 0)$$
$$= 4 - 2\ln 2 - 1 \approx 1.6137$$

Inverse Hyperbolic Functions

The inverses of the six basic hyperbolic functions are very useful in integration (see Chapter 8). Since $d(\sinh x)/dx = \cosh x > 0$, the hyperbolic sine is an increasing function of x. We denote its inverse by

$$y = \sinh^{-1} x.$$

For every value of x in the interval $-\infty < x < \infty$, the value of $y = \sinh^{-1} x$ is the number whose hyperbolic sine is x. The graphs of $y = \sinh x$ and $y = \sinh^{-1} x$ are shown in Figure 7.32a.

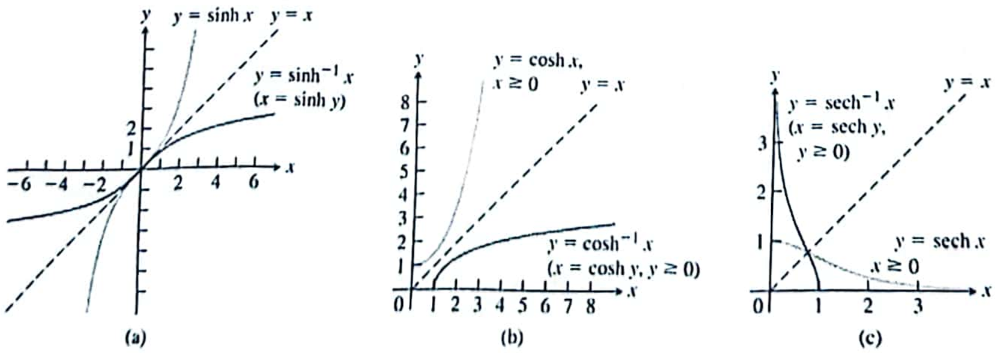


FIGURE 7.32 The graphs of the inverse hyperbolic sine, cosine, and secant of x. Notice the symmetries about the line $y = x$.

The function $y = \cosh x$ is not one-to-one because its graph in Table 7.5 does not pass the horizontal line test. The restricted function $y = \cosh x, x \geq 0$, however, is one-to-one and therefore has an inverse, denoted by

$$y = \cosh^{-1} x.$$

For every value of $x \geq 1$, $y = \cosh^{-1} x$ is the number in the interval $0 \leq y < \infty$ whose hyperbolic cosine is x. The graphs of $y = \cosh x, x \geq 0$, and $y = \cosh^{-1} x$ are shown in Figure 7.32b.

Like $y = \cosh x$, the function $y = \operatorname{sech} x = 1/\cosh x$ fails to be one-to-one, but its restriction to nonnegative values of x does have an inverse, denoted by

$$y = \operatorname{sech}^{-1} x.$$

For every value of x in the interval $(0, 1]$, $y = \operatorname{sech}^{-1} x$ is the nonnegative number whose hyperbolic secant is x. The graphs of $y = \operatorname{sech} x, x \geq 0$, and $y = \operatorname{sech}^{-1} x$ are shown in Figure 7.32c.

The hyperbolic tangent, cotangent, and cosecant are one-to-one on their domains and therefore have inverses, denoted by

$$y = \tanh^{-1} x, \qquad y = \coth^{-1} x, \qquad y = \operatorname{csch}^{-1} x.$$

These functions are graphed in Figure 7.33.

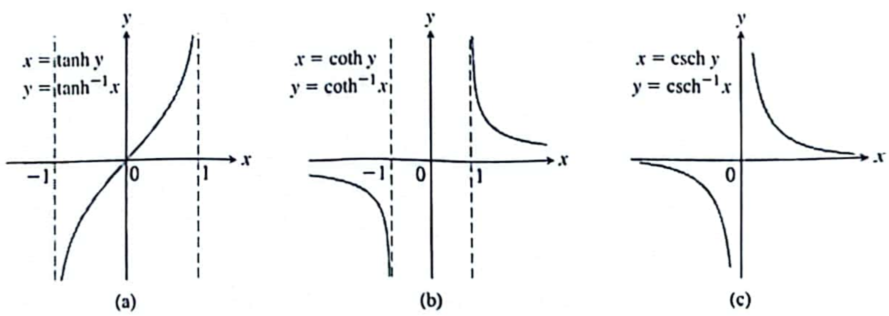


FIGURE 7.33 The graphs of the inverse hyperbolic tangent, cotangent, and cosecant of x.

Useful Identities

TABLE 7.9 Identities for inverse hyperbolic functions
$\operatorname{sech}^{-1} x = \cosh^{-1} \frac{1}{x}$
$\operatorname{csch}^{-1} x = \sinh^{-1} \frac{1}{x}$
$\coth^{-1} x = \tanh^{-1} \frac{1}{x}$

We use the identities in Table 7.9 to calculate the values of $\operatorname{sech}^{-1} x$, $\operatorname{csch}^{-1} x$, and $\coth^{-1} x$ on calculators that give only $\cosh^{-1} x$, $\sinh^{-1} x$, and $\tanh^{-1} x$. These identities are direct consequences of the definitions. For example, if $0 < x \leq 1$, then

$$\operatorname{sech}\left(\cosh^{-1}\left(\frac{1}{x}\right)\right) = \frac{1}{\cosh\left(\cosh^{-1}\left(\frac{1}{x}\right)\right)} = \frac{1}{\left(\frac{1}{x}\right)} = x.$$

We also know that $\operatorname{sech}(\operatorname{sech}^{-1} x) = x$, so because the hyperbolic secant is one-to-one on $(0, 1]$, we have

$$\cosh^{-1}\left(\frac{1}{x}\right) = \operatorname{sech}^{-1} x.$$

Derivatives of Inverse Hyperbolic Functions

An important use of inverse hyperbolic functions lies in antiderivatives that reverse the derivative formulas in Table 7.10.

TABLE 7.10 Derivatives of inverse hyperbolic functions

$\dfrac{d(\sinh^{-1} u)}{dx} = \dfrac{1}{\sqrt{1+u^2}}\dfrac{du}{dx}$	
$\dfrac{d(\cosh^{-1} u)}{dx} = \dfrac{1}{\sqrt{u^2-1}}\dfrac{du}{dx},$	$u > 1$
$\dfrac{d(\tanh^{-1} u)}{dx} = \dfrac{1}{1-u^2}\dfrac{du}{dx},$	$\lvert u \rvert < 1$
$\dfrac{d(\coth^{-1} u)}{dx} = \dfrac{1}{1-u^2}\dfrac{du}{dx},$	$\lvert u \rvert > 1$
$\dfrac{d(\operatorname{sech}^{-1} u)}{dx} = -\dfrac{1}{u\sqrt{1-u^2}}\dfrac{du}{dx},$	$0 < u < 1$
$\dfrac{d(\operatorname{csch}^{-1} u)}{dx} = -\dfrac{1}{\lvert u \rvert\sqrt{1+u^2}}\dfrac{du}{dx},$	$u \neq 0$

The restrictions $|u| < 1$ and $|u| > 1$ on the derivative formulas for $\tanh^{-1} u$ and $\coth^{-1} u$ come from the natural restrictions on the values of these functions. (See Figure 7.33a and b.) The distinction between $|u| < 1$ and $|u| > 1$ becomes important when we convert the derivative formulas into integral formulas.

We illustrate how the derivatives of the inverse hyperbolic functions are found in Example 2, where we calculate $d(\cosh^{-1} u)/dx$. The other derivatives are obtained by similar calculations.

EXAMPLE 2 Show that if u is a differentiable function of x whose values are greater than 1, then

$$\frac{d}{dx}(\cosh^{-1} u) = \frac{1}{\sqrt{u^2-1}}\frac{du}{dx}.$$

Solution First we find the derivative of $y = \cosh^{-1} x$ for $x > 1$ by applying Theorem 1 of Section 7.1 with $f(x) = \cosh x$ and $f^{-1}(x) = \cosh^{-1} x$. Theorem 1 can be applied because the derivative of $\cosh x$ is positive for $0 < x$.

$$\begin{aligned}
(f^{-1})'(x) &= \frac{1}{f'(f^{-1}(x))} && \text{Theorem 1} \\
&= \frac{1}{\sinh(\cosh^{-1} x)} && f'(u) = \sinh u \\
&= \frac{1}{\sqrt{\cosh^2(\cosh^{-1} x) - 1}} && \cosh^2 u - \sinh^2 u = 1, \ \sinh u = \sqrt{\cosh^2 u - 1} \\
&= \frac{1}{\sqrt{x^2-1}} && \cosh(\cosh^{-1} x) = x
\end{aligned}$$

HISTORICAL BIOGRAPHY
Sonya Kovalevsky
(1850–1891)

The Chain Rule gives the final result:

$$\frac{d}{dx}(\cosh^{-1} u) = \frac{1}{\sqrt{u^2 - 1}}\frac{du}{dx}.$$

With appropriate substitutions, the derivative formulas in Table 7.10 lead to the integration formulas in Table 7.11. Each of the formulas in Table 7.11 can be verified by differentiating the expression on the right-hand side.

TABLE 7.11 Integrals leading to inverse hyperbolic functions

1.	$\displaystyle\int \frac{du}{\sqrt{a^2 + u^2}} = \sinh^{-1}\left(\frac{u}{a}\right) + C,$	$a > 0$
2.	$\displaystyle\int \frac{du}{\sqrt{u^2 - a^2}} = \cosh^{-1}\left(\frac{u}{a}\right) + C,$	$u > a > 0$
3.	$\displaystyle\int \frac{du}{a^2 - u^2} = \begin{cases} \frac{1}{a}\tanh^{-1}\left(\frac{u}{a}\right) + C, & u^2 < a^2 \\ \frac{1}{a}\coth^{-1}\left(\frac{u}{a}\right) + C, & u^2 > a^2 \end{cases}$	
4.	$\displaystyle\int \frac{du}{u\sqrt{a^2 - u^2}} = -\frac{1}{a}\operatorname{sech}^{-1}\left(\frac{u}{a}\right) + C,$	$0 < u < a$
5.	$\displaystyle\int \frac{du}{u\sqrt{a^2 + u^2}} = -\frac{1}{a}\operatorname{csch}^{-1}\left\lvert\frac{u}{a}\right\rvert + C,$	$u \neq 0$ and $a > 0$

EXAMPLE 3 Evaluate

$$\int_0^1 \frac{2\,dx}{\sqrt{3 + 4x^2}}.$$

Solution The indefinite integral is

$$\int \frac{2\,dx}{\sqrt{3 + 4x^2}} = \int \frac{du}{\sqrt{a^2 + u^2}} \qquad u = 2x,\ du = 2\,dx,\ a = \sqrt{3}$$

$$= \sinh^{-1}\left(\frac{u}{a}\right) + C \qquad \text{Formula from Table 7.11}$$

$$= \sinh^{-1}\left(\frac{2x}{\sqrt{3}}\right) + C.$$

Therefore,

$$\int_0^1 \frac{2\,dx}{\sqrt{3 + 4x^2}} = \sinh^{-1}\left(\frac{2x}{\sqrt{3}}\right)\Bigg]_0^1 = \sinh^{-1}\left(\frac{2}{\sqrt{3}}\right) - \sinh^{-1}(0)$$

$$= \sinh^{-1}\left(\frac{2}{\sqrt{3}}\right) - 0 \approx 0.98665.$$