Single and Multivariable Calculus

Early Transcendentals

This text was initially written by David Guichard. The single variable material in chapters 1–9 is a modification and expansion of notes written by Neal Koblitz at the University of Washington, who generously gave permission to use, modify, and distribute his work. New material has been added, and old material has been modified, so some portions now bear little resemblance to the original.

The book includes some exercises and examples from *Elementary Calculus: An Approach Using Infinitesimals*, by H. Jerome Keisler, available at http://www.math.wisc.edu/~keisler/calc.html under a Creative Commons license. In addition, the chapter on differential equations (in the multivariable version) and the section on numerical integration are largely derived from the corresponding portions of Keisler's book. Albert Schueller, Barry Balof, and Mike Wills have contributed additional material.

This copy of the text was compiled from source at 7:57 on 12/16/2016.

I will be glad to receive corrections and suggestions for improvement at guichard@whitman.edu.

For Kathleen,
without whose encouragement
this book would not have
been written.

Contents

3

Rules for Finding Derivatives 57

4

Transcendental Functions 73

5

Curve Sketching 107

6

Applications of the Derivative 119

7

Integration 149

8

Techniques of Integration 165

9

10

11

Sequences and Series 257

12

Three Dimensions 299

13

Vector Functions 329

14

15

16

Introduction

The emphasis in this course is on problems—doing calculations and story problems. To master problem solving one needs a tremendous amount of practice doing problems. The more problems you do the better you will be at doing them, as patterns will start to emerge in both the problems and in successful approaches to them. You will learn fastest and best if you devote some time to doing problems every day.

Typically the most difficult problems are story problems, since they require some effort before you can begin calculating. Here are some pointers for doing story problems:

1. Carefully read each problem twice before writing anything.

2. Assign letters to quantities that are described only in words; draw a diagram if appropriate.

3. Decide which letters are constants and which are variables. A letter stands for a constant if its value remains the same throughout the problem.

4. Using mathematical notation, write down what you know and then write down what you want to find.

5. Decide what category of problem it is (this might be obvious if the problem comes at the end of a particular chapter, but will not necessarily be so obvious if it comes on an exam covering several chapters).

6. Double check each step as you go along; don't wait until the end to check your work.

7. Use common sense; if an answer is out of the range of practical possibilities, then check your work to see where you went wrong.

Suggestions for Using This Text

1. Read the example problems carefully, filling in any steps that are left out (ask someone for help if you can't follow the solution to a worked example).

2. Later use the worked examples to study by covering the solutions, and seeing if you can solve the problems on your own.

3. Most exercises have answers in Appendix A; the availability of an answer is marked by "⇒" at the end of the exercise. In the pdf version of the full text, clicking on the arrow will take you to the answer. The answers should be used only as a final check on your work, not as a crutch. Keep in mind that sometimes an answer could be expressed in various ways that are algebraically equivalent, so don't assume that your answer is wrong just because it doesn't have exactly the same form as the answer in the back.

4. A few figures in the pdf and print versions of the book are marked with "(AP)" at the end of the caption. Clicking on this should open a related interactive applet or Sage worksheet in your web browser. Occasionally another link will do the same thing, like this example. (Note to users of a printed text: the words "this example" in the pdf file are blue, and are a link to a Sage worksheet.)

1

Analytic Geometry

Much of the mathematics in this chapter will be review for you. However, the examples will be oriented toward applications and so will take some thought.

In the (x, y) coordinate system we normally write the x-axis horizontally, with positive numbers to the right of the origin, and the y-axis vertically, with positive numbers above the origin. That is, unless stated otherwise, we take "rightward" to be the positive x-direction and "upward" to be the positive y-direction. In a purely mathematical situation, we normally choose the same scale for the x- and y-axes. For example, the line joining the origin to the point (a, a) makes an angle of $45°$ with the x-axis (and also with the y-axis).

In applications, often letters other than x and y are used, and often different scales are chosen in the horizontal and vertical directions. For example, suppose you drop something from a window, and you want to study how its height above the ground changes from second to second. It is natural to let the letter t denote the time (the number of seconds since the object was released) and to let the letter h denote the height. For each t (say, at one-second intervals) you have a corresponding height h. This information can be tabulated, and then plotted on the (t, h) coordinate plane, as shown in figure 1.0.1.

We use the word "quadrant" for each of the four regions into which the plane is divided by the axes: the first quadrant is where points have both coordinates positive, or the "northeast" portion of the plot, and the second, third, and fourth quadrants are counted off counterclockwise, so the second quadrant is the northwest, the third is the southwest, and the fourth is the southeast.

Suppose we have two points A and B in the (x, y)-plane. We often want to know the change in x-coordinate (also called the "horizontal distance") in going from A to B. This

seconds	0	1	2	3	4
meters	80	75.1	60.4	35.9	1.6

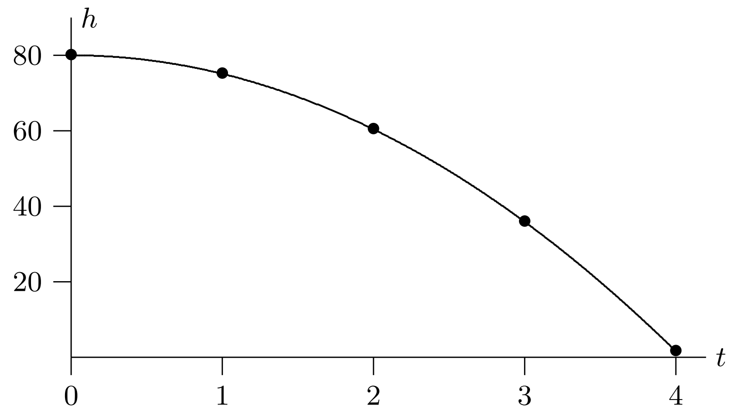


Figure 1.0.1 A data plot, height versus time.

is often written Δx, where the meaning of Δ (a capital delta in the Greek alphabet) is "change in". (Thus, Δx can be read as "change in x" although it usually is read as "delta x". The point is that Δx denotes a single number, and should not be interpreted as "delta times x".) For example, if $A = (2, 1)$ and $B = (3, 3)$, $\Delta x = 3 - 2 = 1$. Similarly, the "change in y" is written Δy. In our example, $\Delta y = 3 - 1 = 2$, the difference between the y-coordinates of the two points. It is the vertical distance you have to move in going from A to B. The general formulas for the change in x and the change in y between a point (x_1, y_1) and a point (x_2, y_2) are:

$$\Delta x = x_2 - x_1, \qquad \Delta y = y_2 - y_1.$$

Note that either or both of these might be negative.

1.1 Lines

If we have two points $A(x_1, y_1)$ and $B(x_2, y_2)$, then we can draw one and only one line through both points. By the *slope* of this line we mean the ratio of Δy to Δx. The slope is often denoted m: $m = \Delta y/\Delta x = (y_2 - y_1)/(x_2 - x_1)$. For example, the line joining the points $(1, -2)$ and $(3, 5)$ has slope $(5 + 2)/(3 - 1) = 7/2$.

EXAMPLE 1.1.1 According to the 1990 U.S. federal income tax schedules, a head of household paid 15% on taxable income up to \$26050. If taxable income was between \$26050 and \$134930, then, in addition, 28% was to be paid on the amount between \$26050 and \$67200, and 33% paid on the amount over \$67200 (if any). Interpret the tax bracket information (15%, 28%, or 33%) using mathematical terminology, and graph the tax on the y-axis against the taxable income on the x-axis.

The percentages, when converted to decimal values 0.15, 0.28, and 0.33, are the *slopes* of the straight lines which form the graph of the tax for the corresponding tax brackets. The tax graph is what's called a *polygonal line*, i.e., it's made up of several straight line segments of different slopes. The first line starts at the point (0,0) and heads upward with slope 0.15 (i.e., it goes upward 15 for every increase of 100 in the x-direction), until it reaches the point above $x = 26050$. Then the graph "bends upward," i.e., the slope changes to 0.28. As the horizontal coordinate goes from $x = 26050$ to $x = 67200$, the line goes upward 28 for each 100 in the x-direction. At $x = 67200$ the line turns upward again and continues with slope 0.33. See figure 1.1.1. □

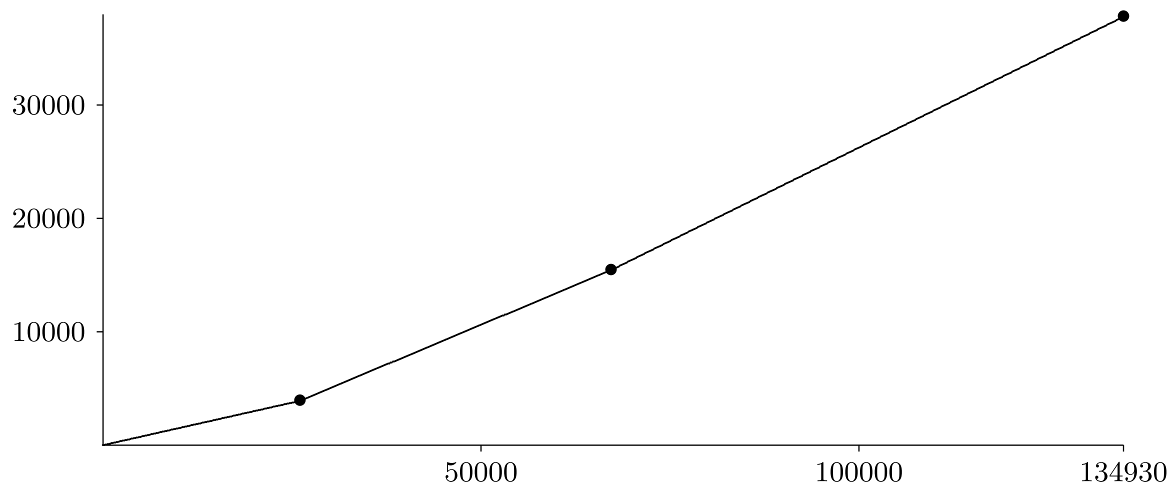


Figure 1.1.1 Tax vs. income.

The most familiar form of the equation of a straight line is: $y = mx + b$. Here m is the slope of the line: if you increase x by 1, the equation tells you that you have to increase y by m. If you increase x by Δx, then y increases by $\Delta y = m\Delta x$. The number b is called the **y-intercept**, because it is where the line crosses the y-axis. If you know two points on a line, the formula $m = (y_2 - y_1)/(x_2 - x_1)$ gives you the slope. Once you know a point and the slope, then the y-intercept can be found by substituting the coordinates of either point in the equation: $y_1 = mx_1 + b$, i.e., $b = y_1 - mx_1$. Alternatively, one can use the "point-slope" form of the equation of a straight line: start with $(y - y_1)/(x - x_1) = m$ and then multiply to get $(y - y_1) = m(x - x_1)$, the point-slope form. Of course, this may be further manipulated to get $y = mx - mx_1 + y_1$, which is essentially the "$mx + b$" form.

It is possible to find the equation of a line between two points directly from the relation $(y - y_1)/(x - x_1) = (y_2 - y_1)/(x_2 - x_1)$, which says "the slope measured between the point (x_1, y_1) and the point (x_2, y_2) is the same as the slope measured between the point (x_1, y_1)

and any other point (x, y) on the line." For example, if we want to find the equation of the line joining our earlier points $A(2, 1)$ and $B(3, 3)$, we can use this formula:

$$\frac{y-1}{x-2} = \frac{3-1}{3-2} = 2, \qquad \text{so that} \qquad y - 1 = 2(x-2), \qquad \text{i.e.,} \qquad y = 2x - 3.$$

Of course, this is really just the point-slope formula, except that we are not computing m in a separate step.

The slope m of a line in the form $y = mx + b$ tells us the direction in which the line is pointing. If m is positive, the line goes into the 1st quadrant as you go from left to right. If m is large and positive, it has a steep incline, while if m is small and positive, then the line has a small angle of inclination. If m is negative, the line goes into the 4th quadrant as you go from left to right. If m is a large negative number (large in absolute value), then the line points steeply downward; while if m is negative but near zero, then it points only a little downward. These four possibilities are illustrated in figure 1.1.2.

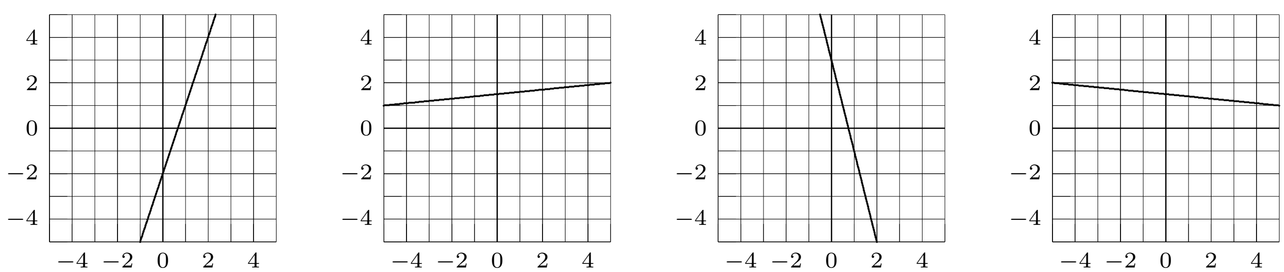


Figure 1.1.2 Lines with slopes 3, 0.1, −4, and −0.1.

If $m = 0$, then the line is horizontal: its equation is simply $y = b$.

There is one type of line that cannot be written in the form $y = mx + b$, namely, vertical lines. A vertical line has an equation of the form $x = a$. Sometimes one says that a vertical line has an "infinite" slope.

Sometimes it is useful to find the x-intercept of a line $y = mx + b$. This is the x-value when $y = 0$. Setting $mx + b$ equal to 0 and solving for x gives: $x = -b/m$. For example, the line $y = 2x - 3$ through the points $A(2, 1)$ and $B(3, 3)$ has x-intercept $3/2$.

EXAMPLE 1.1.2 Suppose that you are driving to Seattle at constant speed, and notice that after you have been traveling for 1 hour (i.e., $t = 1$), you pass a sign saying it is 110 miles to Seattle, and after driving another half-hour you pass a sign saying it is 85 miles to Seattle. Using the horizontal axis for the time t and the vertical axis for the distance y from Seattle, graph and find the equation $y = mt + b$ for your distance from Seattle. Find the slope, y-intercept, and t-intercept, and describe the practical meaning of each.

The graph of y versus t is a straight line because you are traveling at constant speed. The line passes through the two points $(1, 110)$ and $(1.5, 85)$, so its slope is $m = (85 -$

$110)/(1.5-1) = -50$. The meaning of the slope is that you are traveling at 50 mph; m is negative because you are traveling *toward* Seattle, i.e., your distance y is *decreasing*. The word "velocity" is often used for $m = -50$, when we want to indicate direction, while the word "speed" refers to the magnitude (absolute value) of velocity, which is 50 mph. To find the equation of the line, we use the point-slope formula:

$$\frac{y-110}{t-1} = -50, \qquad \text{so that} \qquad y = -50(t-1) + 110 = -50t + 160.$$

The meaning of the y-intercept 160 is that when $t = 0$ (when you started the trip) you were 160 miles from Seattle. To find the t-intercept, set $0 = -50t+160$, so that $t = 160/50 = 3.2$. The meaning of the t-intercept is the duration of your trip, from the start until you arrive in Seattle. After traveling 3 hours and 12 minutes, your distance y from Seattle will be 0. □

Exercises 1.1.

1. Find the equation of the line through $(1, 1)$ and $(-5, -3)$ in the form $y = mx + b$. ⇒

2. Find the equation of the line through $(-1, 2)$ with slope -2 in the form $y = mx + b$. ⇒

3. Find the equation of the line through $(-1, 1)$ and $(5, -3)$ in the form $y = mx + b$. ⇒

4. Change the equation $y - 2x = 2$ to the form $y = mx + b$, graph the line, and find the y-intercept and x-intercept. ⇒

5. Change the equation $x+y = 6$ to the form $y = mx+b$, graph the line, and find the y-intercept and x-intercept. ⇒

6. Change the equation $x = 2y - 1$ to the form $y = mx + b$, graph the line, and find the y-intercept and x-intercept. ⇒

7. Change the equation $3 = 2y$ to the form $y = mx + b$, graph the line, and find the y-intercept and x-intercept. ⇒

8. Change the equation $2x + 3y + 6 = 0$ to the form $y = mx + b$, graph the line, and find the y-intercept and x-intercept. ⇒

9. Determine whether the lines $3x + 6y = 7$ and $2x + 4y = 5$ are parallel. ⇒

10. Suppose a triangle in the x, y–plane has vertices $(-1, 0)$, $(1, 0)$ and $(0, 2)$. Find the equations of the three lines that lie along the sides of the triangle in $y = mx + b$ form. ⇒

11. Suppose that you are driving to Seattle at constant speed. After you have been traveling for an hour you pass a sign saying it is 130 miles to Seattle, and after driving another 20 minutes you pass a sign saying it is 105 miles to Seattle. Using the horizontal axis for the time t and the vertical axis for the distance y from your starting point, graph and find the equation $y = mt + b$ for your distance from your starting point. How long does the trip to Seattle take? ⇒

12. Let x stand for temperature in degrees Celsius (centigrade), and let y stand for temperature in degrees Fahrenheit. A temperature of $0°$C corresponds to $32°$F, and a temperature of $100°$C corresponds to $212°$F. Find the equation of the line that relates temperature Fahrenheit y to temperature Celsius x in the form $y = mx + b$. Graph the line, and find the point at which this line intersects $y = x$. What is the practical meaning of this point? ⇒

13. A car rental firm has the following charges for a certain type of car: \$25 per day with 100 free miles included, \$0.15 per mile for more than 100 miles. Suppose you want to rent a car for one day, and you know you'll use it for more than 100 miles. What is the equation relating the cost y to the number of miles x that you drive the car? ⇒

14. A photocopy store advertises the following prices: 5¢ per copy for the first 20 copies, 4¢ per copy for the 21st through 100th copy, and 3¢ per copy after the 100th copy. Let x be the number of copies, and let y be the total cost of photocopying. (a) Graph the cost as x goes from 0 to 200 copies. (b) Find the equation in the form $y = mx + b$ that tells you the cost of making x copies when x is more than 100. ⇒

15. In the Kingdom of Xyg the tax system works as follows. Someone who earns less than 100 gold coins per month pays no tax. Someone who earns between 100 and 1000 gold coins pays tax equal to 10% of the amount over 100 gold coins that he or she earns. Someone who earns over 1000 gold coins must hand over to the King all of the money earned over 1000 in addition to the tax on the first 1000. (a) Draw a graph of the tax paid y versus the money earned x, and give formulas for y in terms of x in each of the regions $0 \le x \le 100$, $100 \le x \le 1000$, and $x \ge 1000$. (b) Suppose that the King of Xyg decides to use the second of these line segments (for $100 \le x \le 1000$) for $x \le 100$ as well. Explain in practical terms what the King is doing, and what the meaning is of the y-intercept. ⇒

16. The tax for a single taxpayer is described in the figure 1.1.3. Use this information to graph tax versus taxable income (i.e., x is the amount on Form 1040, line 37, and y is the amount on Form 1040, line 38). Find the slope and y-intercept of each line that makes up the polygonal graph, up to $x = 97620$. ⇒

1990 Tax Rate Schedules

Schedule X—Use if your filing status is **Single**

If the amount on Form 1040 line 37 is over:	But not over:	Enter on Form 1040 line 38	of the amount over:
\$0	\$19,450	**15%**	**\$0**
19,450	47,050	**\$2,917.50+28%**	**19,450**
47,050	97,620	**\$10,645.50+33%**	**47,050**
97,620	....	Use **Worksheet** below to figure your tax	

Schedule Z—Use if your filing status is **Head of household**

If the amount on Form 1040 line 37 is over:	But not over:	Enter on Form 1040 line 38	of the amount over:
\$0	\$26,050	**15%**	**\$0**
26,050	67,200	**\$3,907.50+28%**	**26,050**
67,200	134,930	**\$15,429.50+33%**	**67,200**
134,930	....	Use **Worksheet** below to figure your tax	

Figure 1.1.3 Tax Schedule.

17. Market research tells you that if you set the price of an item at \$1.50, you will be able to sell 5000 items; and for every 10 cents you lower the price below \$1.50 you will be able to sell another 1000 items. Let x be the number of items you can sell, and let P be the price of an item. (a) Express P linearly in terms of x, in other words, express P in the form $P = mx + b$. (b) Express x linearly in terms of P. ⇒

18. An instructor gives a 100-point final exam, and decides that a score 90 or above will be a grade of 4.0, a score of 40 or below will be a grade of 0.0, and between 40 and 90 the grading will be linear. Let x be the exam score, and let y be the corresponding grade. Find a formula of the form $y = mx + b$ which applies to scores x between 40 and 90. ⇒

1.2 Distance Between Two Points; Circles

Given two points (x_1, y_1) and (x_2, y_2), recall that their horizontal distance from one another is $\Delta x = x_2 - x_1$ and their vertical distance from one another is $\Delta y = y_2 - y_1$. (Actually, the word "distance" normally denotes "positive distance". Δx and Δy are *signed* distances, but this is clear from context.) The actual (positive) distance from one point to the other is the length of the hypotenuse of a right triangle with legs $|\Delta x|$ and $|\Delta y|$, as shown in figure 1.2.1. The Pythagorean theorem then says that the distance between the two points is the square root of the sum of the squares of the horizontal and vertical sides:

$$\text{distance} = \sqrt{(\Delta x)^2 + (\Delta y)^2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

For example, the distance between points $A(2, 1)$ and $B(3, 3)$ is $\sqrt{(3-2)^2 + (3-1)^2} = \sqrt{5}$.

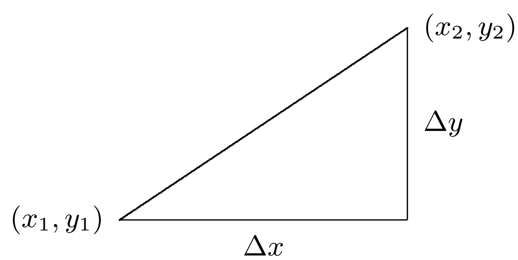


Figure 1.2.1 Distance between two points, Δx and Δy positive.

As a special case of the distance formula, suppose we want to know the distance of a point (x, y) to the origin. According to the distance formula, this is $\sqrt{(x-0)^2 + (y-0)^2} = \sqrt{x^2 + y^2}$.

A point (x, y) is at a distance r from the origin if and only if $\sqrt{x^2 + y^2} = r$, or, if we square both sides: $x^2 + y^2 = r^2$. This is the equation of the circle of radius r centered at the origin. The special case $r = 1$ is called the unit circle; its equation is $x^2 + y^2 = 1$.

Similarly, if $C(h, k)$ is any fixed point, then a point (x, y) is at a distance r from the point C if and only if $\sqrt{(x-h)^2 + (y-k)^2} = r$, i.e., if and only if

$$(x-h)^2 + (y-k)^2 = r^2.$$

This is the equation of the circle of radius r centered at the point (h, k). For example, the circle of radius 5 centered at the point $(0, -6)$ has equation $(x-0)^2 + (y - -6)^2 = 25$, or $x^2 + (y+6)^2 = 25$. If we expand this we get $x^2 + y^2 + 12y + 36 = 25$ or $x^2 + y^2 + 12y + 11 = 0$, but the original form is usually more useful.

EXAMPLE 1.2.1 Graph the circle $x^2 - 2x + y^2 + 4y - 11 = 0$. With a little thought we convert this to $(x-1)^2 + (y+2)^2 - 16 = 0$ or $(x-1)^2 + (y+2)^2 = 16$. Now we see that this is the circle with radius 4 and center $(1, -2)$, which is easy to graph. □

Exercises 1.2.

1. Find the equation of the circle of radius 3 centered at:

a) $(0, 0)$
b) $(5, 6)$
c) $(-5, -6)$
d) $(0, 3)$
e) $(0, -3)$
f) $(3, 0)$

⇒

2. For each pair of points $A(x_1, y_1)$ and $B(x_2, y_2)$ find (i) Δx and Δy in going from A to B, (ii) the slope of the line joining A and B, (iii) the equation of the line joining A and B in the form $y = mx + b$, (iv) the distance from A to B, and (v) an equation of the circle with center at A that goes through B.

a) $A(2, 0)$, $B(4, 3)$
b) $A(1, -1)$, $B(0, 2)$
c) $A(0, 0)$, $B(-2, -2)$
d) $A(-2, 3)$, $B(4, 3)$
e) $A(-3, -2)$, $B(0, 0)$
f) $A(0.01, -0.01)$, $B(-0.01, 0.05)$

⇒

3. Graph the circle $x^2 + y^2 + 10y = 0$.

4. Graph the circle $x^2 - 10x + y^2 = 24$.

5. Graph the circle $x^2 - 6x + y^2 - 8y = 0$.

6. Find the standard equation of the circle passing through $(-2, 1)$ and tangent to the line $3x - 2y = 6$ at the point $(4, 3)$. Sketch. (Hint: The line through the center of the circle and the point of tangency is perpendicular to the tangent line.) ⇒

1.3 FUNCTIONS

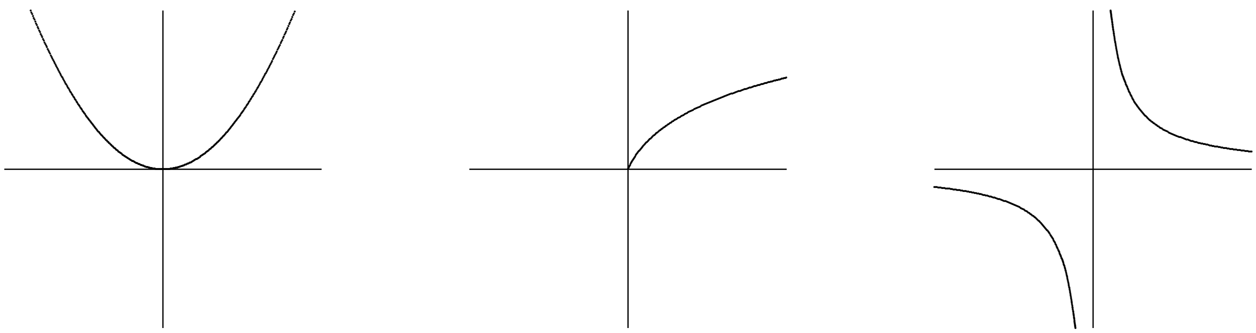
A **function** $y = f(x)$ is a rule for determining y when we're given a value of x. For example, the rule $y = f(x) = 2x + 1$ is a function. Any line $y = mx + b$ is called a **linear**function. The graph of a function looks like a curve above (or below) the x-axis, where for any value of x the rule $y = f(x)$ tells us how far to go above (or below) the x-axis to reach the curve.

Functions can be defined in various ways: by an algebraic formula or several algebraic formulas, by a graph, or by an experimentally determined table of values. (In the latter case, the table gives a bunch of points in the plane, which we might then interpolate with a smooth curve, if that makes sense.)

Given a value of x, a function must give at most one value of y. Thus, vertical lines are not functions. For example, the line $x = 1$ has infinitely many values of y if $x = 1$. It is also true that if x is any number not 1 there is no y which corresponds to x, but that is not a problem—only multiple y values is a problem.

In addition to lines, another familiar example of a function is the parabola $y = f(x) = x^2$. We can draw the graph of this function by taking various values of x (say, at regular intervals) and plotting the points $(x, f(x)) = (x, x^2)$. Then connect the points with a smooth curve. (See figure 1.3.1.)

The two examples $y = f(x) = 2x + 1$ and $y = f(x) = x^2$ are both functions which can be evaluated at *any* value of x from negative infinity to positive infinity. For many functions, however, it only makes sense to take x in some interval or outside of some "forbidden" region. The interval of x-values at which we're allowed to evaluate the function is called the **domain** of the function.



$y = f(x) = x^2$ $\qquad$ $y = f(x) = \sqrt{x}$ $\qquad$ $y = f(x) = 1/x$

Figure 1.3.1 Some graphs.

For example, the square-root function $y = f(x) = \sqrt{x}$ is the rule which says, given an x-value, take the nonnegative number whose square is x. This rule only makes sense if x is positive or zero. We say that the domain of this function is $x \geq 0$, or more formally $\{x \in \mathbb{R} \mid x \geq 0\}$. Alternately, we can use interval notation, and write that the domain is $[0, \infty)$. (In interval notation, square brackets mean that the endpoint is included, and a parenthesis means that the endpoint is not included.) The fact that the domain of $y = \sqrt{x}$ is $[0, \infty)$ means that in the graph of this function ((see figure 1.3.1) we have points (x, y) only above x-values on the right side of the x-axis.

Another example of a function whose domain is not the entire x-axis is: $y = f(x) = 1/x$, the reciprocal function. We cannot substitute $x = 0$ in this formula. The function makes sense, however, for any nonzero x, so we take the domain to be: $\{x \in \mathbb{R} \mid x \neq 0\}$. The graph of this function does not have any point (x, y) with $x = 0$. As x gets close to 0 from either side, the graph goes off toward infinity. We call the vertical line $x = 0$ an **asymptote**.

To summarize, two reasons why certain x-values are excluded from the domain of a function are that (i) we cannot divide by zero, and (ii) we cannot take the square root of a negative number. We will encounter some other ways in which functions might be undefined later.

Another reason why the domain of a function might be restricted is that in a given situation the x-values outside of some range might have no practical meaning. For example, if y is the area of a square of side x, then we can write $y = f(x) = x^2$. In a purely mathematical context the domain of the function $y = x^2$ is all of $\mathbb{R}$. But in the story-problem context of finding areas of squares, we restrict the domain to positive values of x, because a square with negative or zero side makes no sense.

In a problem in pure mathematics, we usually take the domain to be all values of x at which the formulas can be evaluated. But in a story problem there might be further restrictions on the domain because only certain values of x are of interest or make practical sense.

In a story problem, often letters different from x and y are used. For example, the volume V of a sphere is a function of the radius r, given by the formula $V = f(r) = {}^4\!/_3\pi r^3$. Also, letters different from f may be used. For example, if y is the velocity of something at time t, we may write $y = v(t)$ with the letter v (instead of f) standing for the velocity function (and t playing the role of x).

The letter playing the role of x is called the **independent variable**, and the letter playing the role of y is called the **dependent variable** (because its value "depends on" the value of the independent variable). In story problems, when one has to translate from English into mathematics, a crucial step is to determine what letters stand for variables. If only words and no letters are given, then we have to decide which letters to use. Some letters are traditional. For example, almost always, t stands for time.

EXAMPLE 1.3.1 An open-top box is made from an $a \times b$ rectangular piece of cardboard by cutting out a square of side x from each of the four corners, and then folding the sides up and sealing them with duct tape. Find a formula for the volume V of the box as a function of x, and find the domain of this function.

The box we get will have height x and rectangular base of dimensions $a-2x$ by $b-2x$. Thus,

$$V = f(x) = x(a-2x)(b-2x).$$

Here a and b are constants, and V is the variable that depends on x, i.e., V is playing the role of y.

This formula makes mathematical sense for any x, but in the story problem the domain is much less. In the first place, x must be positive. In the second place, it must be less than half the length of either of the sides of the cardboard. Thus, the domain is

$$\{x \in \mathbb{R} \mid 0 < x < \frac{1}{2}(\text{minimum of } a \text{ and } b)\}.$$

In interval notation we write: the domain is the interval $(0, \min(a,b)/2)$. (You might think about whether we could allow 0 or $\min(a,b)/2$ to be in the domain. They make a certain physical sense, though we normally would not call the result a box. If we were to allow these values, what would the corresponding volumes be? Does that make sense?) □

EXAMPLE 1.3.2 Circle of radius r centered at the origin The equation for this circle is usually given in the form $x^2 + y^2 = r^2$. To write the equation in the form $y = f(x)$ we solve for y, obtaining $y = \pm\sqrt{r^2 - x^2}$. But *this is not a function*, because when we substitute a value in $(-r, r)$ for x there are two corresponding values of y. To get a function, we must choose one of the two signs in front of the square root. If we choose the positive sign, for example, we get the upper semicircle $y = f(x) = \sqrt{r^2 - x^2}$ (see figure 1.3.2). The domain of this function is the interval $[-r, r]$, i.e., x must be between $-r$ and r (including the endpoints). If x is outside of that interval, then $r^2 - x^2$ is negative, and we cannot take the square root. In terms of the graph, this just means that there are no points on the curve whose x-coordinate is greater than r or less than $-r$. □

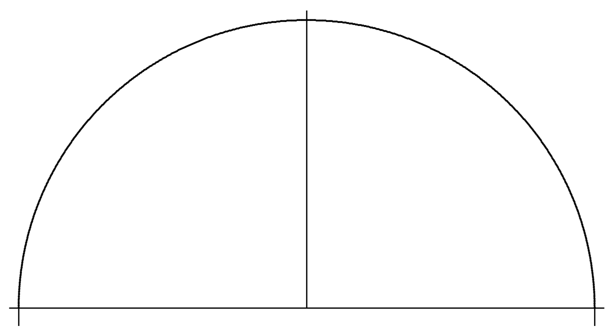


Figure 1.3.2 Upper semicircle $y = \sqrt{r^2 - x^2}$

EXAMPLE 1.3.3 Find the domain of

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

To answer this question, we must rule out the x-values that make $4x - x^2$ negative (because we cannot take the square root of a negative number) and also the x-values that make $4x - x^2$ zero (because if $4x - x^2 = 0$, then when we take the square root we get 0, and we cannot divide by 0). In other words, the domain consists of all x for which $4x - x^2$ is strictly positive. We give two different methods to find out when $4x - x^2 > 0$.

First method. Factor $4x - x^2$ as $x(4 - x)$. The product of two numbers is positive when either both are positive or both are negative, i.e., if either $x > 0$ and $4 - x > 0$, or else $x < 0$ and $4 - x < 0$. The latter alternative is impossible, since if x is negative, then $4 - x$ is greater than 4, and so cannot be negative. As for the first alternative, the condition $4 - x > 0$ can be rewritten (adding x to both sides) as $4 > x$, so we need: $x > 0$ and $4 > x$ (this is sometimes combined in the form $4 > x > 0$, or, equivalently, $0 < x < 4$). In interval notation, this says that the domain is the interval $(0, 4)$.

Second method. Write $4x - x^2$ as $-(x^2 - 4x)$, and then complete the square, obtaining $-\left((x-2)^2 - 4\right) = 4 - (x-2)^2$. For this to be positive we need $(x-2)^2 < 4$, which means that $x - 2$ must be less than 2 and greater than -2: $-2 < x - 2 < 2$. Adding 2 to everything gives $0 < x < 4$. Both of these methods are equally correct; you may use either in a problem of this type. □

A function does not always have to be given by a single formula, as we have already seen (in the income tax problem, for example). Suppose that $y = v(t)$ is the velocity function for a car which starts out from rest (zero velocity) at time $t = 0$; then increases its speed steadily to 20 m/sec, taking 10 seconds to do this; then travels at constant speed 20 m/sec for 15 seconds; and finally applies the brakes to decrease speed steadily to 0, taking 5 seconds to do this. The formula for $y = v(t)$ is different in each of the three time intervals: first $y = 2x$, then $y = 20$, then $y = -4x + 120$. The graph of this function is shown in figure 1.3.3.

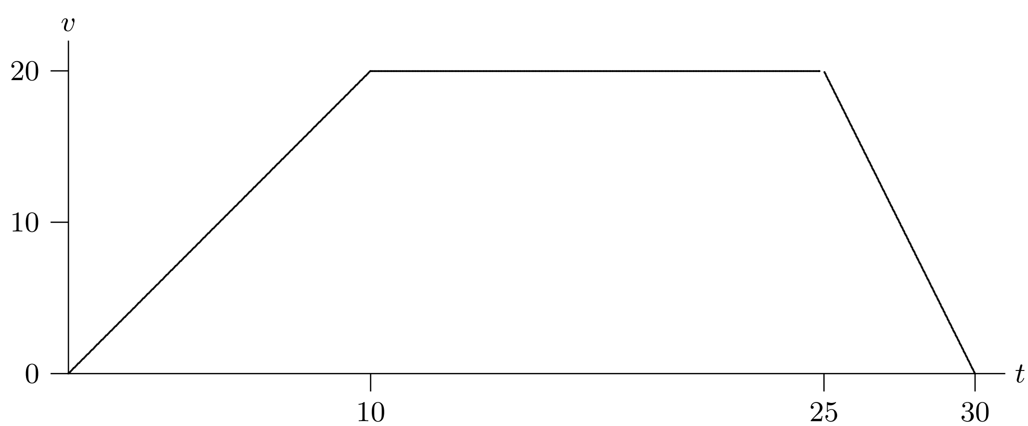


Figure 1.3.3 A velocity function.

Not all functions are given by formulas at all. A function can be given by an experimentally determined table of values, or by a description other than a formula. For example, the population y of the U.S. is a function of the time t: we can write $y = f(t)$. This is a perfectly good function—we could graph it (up to the present) if we had data for various t—but we can't find an algebraic formula for it.

Exercises 1.3.

Find the domain of each of the following functions:

1. $y = f(x) = \sqrt{2x-3}$ ⇒

2. $y = f(x) = 1/(x+1)$ ⇒

3. $y = f(x) = 1/(x^2-1)$ ⇒

4. $y = f(x) = \sqrt{-1/x}$ ⇒

5. $y = f(x) = \sqrt[3]{x}$ ⇒

6. $y = f(x) = \sqrt[4]{x}$ ⇒

7. $y = f(x) = \sqrt{r^2-(x-h)^2}$, where r is a positive constant. ⇒

8. $y = f(x) = \sqrt{1-(1/x)}$ ⇒

9. $y = f(x) = 1/\sqrt{1-(3x)^2}$ ⇒

10. $y = f(x) = \sqrt{x} + 1/(x-1)$ ⇒

11. $y = f(x) = 1/(\sqrt{x}-1)$ ⇒

12. Find the domain of $h(x) = \begin{cases} (x^2-9)/(x-3) & x \neq 3 \\ 6 & \text{if } x = 3. \end{cases}$ ⇒

13. Suppose $f(x) = 3x - 9$ and $g(x) = \sqrt{x}$. What is the domain of the composition $(g \circ f)(x)$? (Recall that **composition** is defined as $(g \circ f)(x) = g(f(x))$.) What is the domain of $(f \circ g)(x)$? ⇒

14. A farmer wants to build a fence along a river. He has 500 feet of fencing and wants to enclose a rectangular pen on three sides (with the river providing the fourth side). If x is the length of the side perpendicular to the river, determine the area of the pen as a function of x. What is the domain of this function? ⇒

15. A can in the shape of a cylinder is to be made with a total of 100 square centimeters of material in the side, top, and bottom; the manufacturer wants the can to hold the maximum possible volume. Write the volume as a function of the radius r of the can; find the domain of the function. ⇒

16. A can in the shape of a cylinder is to be made to hold a volume of one liter (1000 cubic centimeters). The manufacturer wants to use the least possible material for the can. Write the surface area of the can (total of the top, bottom, and side) as a function of the radius r of the can; find the domain of the function. ⇒

1.4 Shifts and Dilations

Many functions in applications are built up from simple functions by inserting constants in various places. It is important to understand the effect such constants have on the appearance of the graph.

Horizontal shifts. *If we replace x by $x-C$ everywhere it occurs in the formula for $f(x)$, then the graph shifts over C to the right.* (If C is negative, then this means that the graph shifts over $|C|$ to the left.) For example, the graph of $y = (x-2)^2$ is the x^2-parabola shifted over to have its vertex at the point 2 on the x-axis. The graph of $y = (x+1)^2$ is the same parabola shifted over to the left so as to have its vertex at -1 on the x-axis. Note well: when replacing x by $x - C$ we must pay attention to meaning, not merely appearance. Starting with $y = x^2$ and literally replacing x by $x - 2$ gives $y = x - 2^2$. This is $y = x - 4$, a line with slope 1, not a shifted parabola.

Vertical shifts. *If we replace y by $y - D$, then the graph moves up D units.* (If D is negative, then this means that the graph moves down $|D|$ units.) If the formula is written in the form $y = f(x)$ and if y is replaced by $y - D$ to get $y - D = f(x)$, we can equivalently move D to the other side of the equation and write $y = f(x) + D$. Thus, this principle can be stated: *to get the graph of $y = f(x) + D$, take the graph of $y = f(x)$ and move it D units up.* For example, the function $y = x^2 - 4x = (x - 2)^2 - 4$ can be obtained from $y = (x - 2)^2$ (see the last paragraph) by moving the graph 4 units down. The result is the x^2-parabola shifted 2 units to the right and 4 units down so as to have its vertex at the point $(2, -4)$.

Warning. Do not confuse $f(x) + D$ and $f(x + D)$. For example, if $f(x)$ is the function x^2, then $f(x) + 2$ is the function $x^2 + 2$, while $f(x + 2)$ is the function $(x + 2)^2 = x^2 + 4x + 4$.

EXAMPLE 1.4.1 Circles An important example of the above two principles starts with the circle $x^2 + y^2 = r^2$. This is the circle of radius r centered at the origin. (As we saw, this is not a single function $y = f(x)$, but rather two functions $y = \pm\sqrt{r^2 - x^2}$ put together; in any case, the two shifting principles apply to equations like this one that are not in the form $y = f(x)$.) If we replace x by $x - C$ and replace y by $y - D$—getting the equation $(x - C)^2 + (y - D)^2 = r^2$—the effect on the circle is to move it C to the right and D up, thereby obtaining the circle of radius r centered at the point (C, D). This tells us how to write the equation of any circle, not necessarily centered at the origin. □

We will later want to use two more principles concerning the effects of constants on the appearance of the graph of a function.

Horizontal dilation. *If x is replaced by x/A in a formula and $A > 1$, then the effect on the graph is to expand it by a factor of A in the x-direction (away from the y-axis).* If A is between 0 and 1 then the effect on the graph is to contract by a factor of $1/A$ (towards the y-axis). We use the word "dilate" to mean expand or contract.

For example, replacing x by $x/0.5 = x/(1/2) = 2x$ has the effect of contracting toward the y-axis by a factor of 2. If A is negative, we dilate by a factor of $|A|$ and then flip about the y-axis. Thus, replacing x by $-x$ has the effect of taking the mirror image of the graph with respect to the y-axis. For example, the function $y = \sqrt{-x}$, which has domain $\{x \in \mathbb{R} \mid x \leq 0\}$, is obtained by taking the graph of $\sqrt{x}$ and flipping it around the y-axis into the second quadrant.

Vertical dilation. *If y is replaced by y/B in a formula and $B > 0$, then the effect on the graph is to dilate it by a factor of B in the vertical direction.* As before, this is an expansion or contraction depending on whether B is larger or smaller than one. Note that if we have a function $y = f(x)$, replacing y by y/B is equivalent to multiplying the function on the right by B: $y = Bf(x)$. The effect on the graph is to expand the picture away from the x-axis by a factor of B if $B > 1$, to contract it toward the x-axis by a factor of $1/B$ if $0 < B < 1$, and to dilate by $|B|$ and then flip about the x-axis if B is negative.

EXAMPLE 1.4.2 Ellipses A basic example of the two expansion principles is given by an **ellipse of semimajor axis a and semiminor axis b**. We get such an ellipse by starting with the unit circle—the circle of radius 1 centered at the origin, the equation of which is $x^2 + y^2 = 1$—and dilating by a factor of a horizontally and by a factor of b vertically. To get the equation of the resulting ellipse, which crosses the x-axis at $\pm a$ and crosses the y-axis at $\pm b$, we replace x by x/a and y by y/b in the equation for the unit circle. This gives

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1 \qquad \text{or} \qquad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

□

Finally, if we want to analyze a function that involves both shifts and dilations, it is usually simplest to work with the dilations first, and then the shifts. For instance, if we want to dilate a function by a factor of A in the x-direction and then shift C to the right, we do this by replacing x first by x/A and then by $(x - C)$ in the formula. As an example, suppose that, after dilating our unit circle by a in the x-direction and by b in the y-direction to get the ellipse in the last paragraph, we then wanted to shift it a distance h to the right and a distance k upward, so as to be centered at the point (h, k). The new ellipse would have equation

$$\left(\frac{x-h}{a}\right)^2 + \left(\frac{y-k}{b}\right)^2 = 1.$$

Note well that this is different than first doing shifts by h and k and then dilations by a and b:

$$\left(\frac{x}{a} - h\right)^2 + \left(\frac{y}{b} - k\right)^2 = 1.$$

See figure 1.4.1.

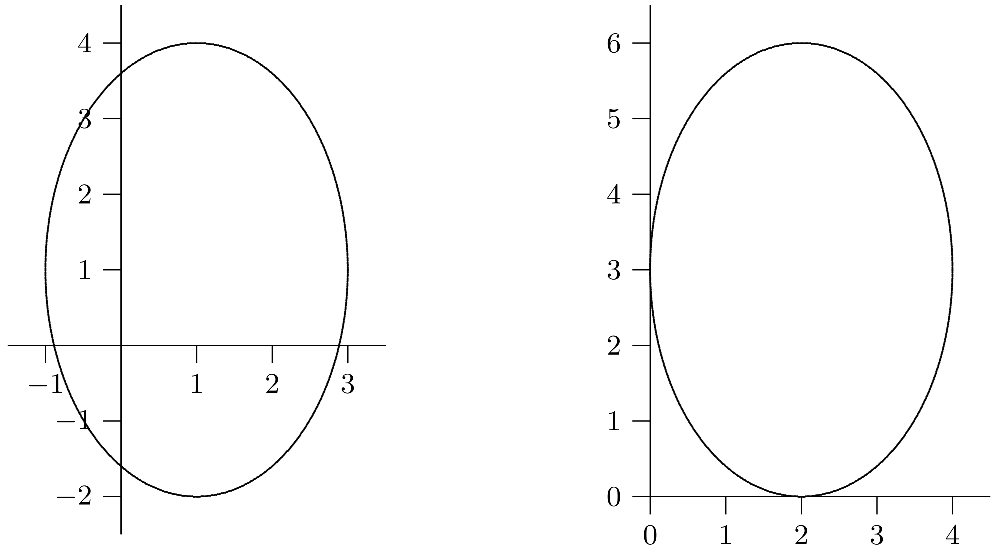


Figure 1.4.1 Ellipses: $\left(\frac{x-1}{2}\right)^2 + \left(\frac{y-1}{3}\right)^2 = 1$ on the left, $\left(\frac{x}{2}-1\right)^2 + \left(\frac{y}{3}-1\right)^2 = 1$ on the right.

Exercises 1.4.

Starting with the graph of $y = \sqrt{x}$, the graph of $y = 1/x$, and the graph of $y = \sqrt{1-x^2}$ (the upper unit semicircle), sketch the graph of each of the following functions:

1. $f(x) = \sqrt{x-2}$

2. $f(x) = -1 - 1/(x+2)$

3. $f(x) = 4 + \sqrt{x+2}$

4. $y = f(x) = x/(1-x)$

5. $y = f(x) = -\sqrt{-x}$

6. $f(x) = 2 + \sqrt{1-(x-1)^2}$

7. $f(x) = -4 + \sqrt{-(x-2)}$

8. $f(x) = 2\sqrt{1-(x/3)^2}$

9. $f(x) = 1/(x+1)$

10. $f(x) = 4 + 2\sqrt{1-(x-5)^2/9}$

11. $f(x) = 1 + 1/(x-1)$

12. $f(x) = \sqrt{100-25(x-1)^2} + 2$

The graph of $f(x)$ is shown below. Sketch the graphs of the following functions.

13. $y = f(x-1)$

14. $y = 1 + f(x+2)$

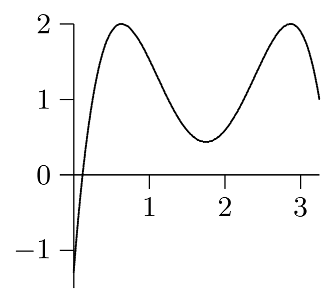
15. $y = 1 + 2f(x)$

16. $y = 2f(3x)$

17. $y = 2f(3(x-2)) + 1$

18. $y = (1/2)f(3x-3)$

19. $y = f(1 + x/3) + 2$



2

Instantaneous Rate of Change: The Derivative

2.1 The slope of a function

Suppose that y is a function of x, say $y = f(x)$. It is often necessary to know how sensitive the value of y is to small changes in x.

EXAMPLE 2.1.1 Take, for example, $y = f(x) = \sqrt{625 - x^2}$ (the upper semicircle of radius 25 centered at the origin). When $x = 7$, we find that $y = \sqrt{625 - 49} = 24$. Suppose we want to know how much y changes when x increases a little, say to 7.1 or 7.01.

In the case of a straight line $y = mx+b$, the slope $m = \Delta y/\Delta x$ measures the change in y per unit change in x. This can be interpreted as a measure of "sensitivity"; for example, if $y = 100x + 5$, a small change in x corresponds to a change one hundred times as large in y, so y is quite sensitive to changes in x.

Let us look at the same ratio $\Delta y/\Delta x$ for our function $y = f(x) = \sqrt{625 - x^2}$ when x changes from 7 to 7.1. Here $\Delta x = 7.1 - 7 = 0.1$ is the change in x, and

$$\begin{aligned}\Delta y &= f(x + \Delta x) - f(x) = f(7.1) - f(7)\\ &= \sqrt{625 - 7.1^2} - \sqrt{625 - 7^2} \approx 23.9706 - 24 = -0.0294.\end{aligned}$$

Thus, $\Delta y/\Delta x \approx -0.0294/0.1 = -0.294$. This means that y changes by less than one third the change in x, so apparently y is not very sensitive to changes in x at $x = 7$. We say "apparently" here because we don't really know what happens between 7 and 7.1. Perhaps y changes dramatically as x runs through the values from 7 to 7.1, but at 7.1 y just happens to be close to its value at 7. This is not in fact the case for this particular function, but we don't yet know why. □

One way to interpret the above calculation is by reference to a line. We have computed the slope of the line through $(7, 24)$ and $(7.1, 23.9706)$, called a **chord** of the circle. In general, if we draw the chord from the point $(7, 24)$ to a nearby point on the semicircle $(7 + \Delta x,\ f(7 + \Delta x))$, the slope of this chord is the so-called **difference quotient**

$$\text{slope of chord} = \frac{f(7 + \Delta x) - f(7)}{\Delta x} = \frac{\sqrt{625 - (7 + \Delta x)^2} - 24}{\Delta x}.$$

For example, if x changes only from 7 to 7.01, then the difference quotient (slope of the chord) is approximately equal to $(23.997081 - 24)/0.01 = -0.2919$. This is slightly less steep than the chord from $(7, 24)$ to $(7.1, 23.9706)$.

As the second value $7 + \Delta x$ moves in towards 7, the chord joining $(7, f(7))$ to $(7 + \Delta x, f(7 + \Delta x))$ shifts slightly. As indicated in figure 2.1.1, as Δx gets smaller and smaller, the chord joining $(7, 24)$ to $(7 + \Delta x, f(7 + \Delta x))$ gets closer and closer to the **tangent line** to the circle at the point $(7, 24)$. (Recall that the tangent line is the line that just grazes the circle at that point, i.e., it doesn't meet the circle at any second point.) Thus, as Δx gets smaller and smaller, the slope $\Delta y/\Delta x$ of the chord gets closer and closer to the slope of the tangent line. This is actually quite difficult to see when Δx is small, because of the scale of the graph. The values of Δx used for the figure are 1, 5, 10 and 15, not really very small values. The tangent line is the one that is uppermost at the right hand endpoint.

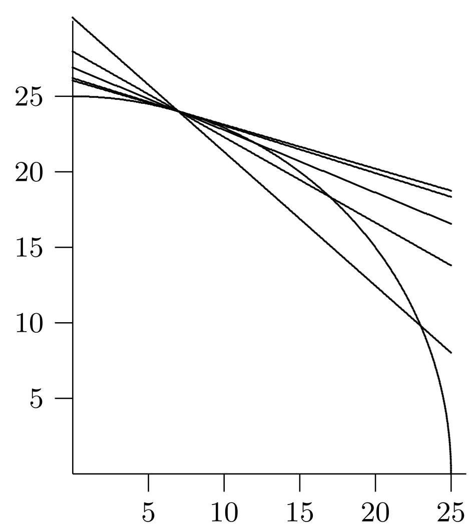


Figure 2.1.1 Chords approximating the tangent line. (AP)

So far we have found the slopes of two chords that should be close to the slope of the tangent line, but what is the slope of the tangent line exactly? Since the tangent line touches the circle at just one point, we will never be able to calculate its slope directly, using two "known" points on the line. What we need is a way to capture what happens to the slopes of the chords as they get "closer and closer" to the tangent line.

Instead of looking at more particular values of Δx, let's see what happens if we do some algebra with the difference quotient using just Δx. The slope of a chord from $(7,24)$ to a nearby point is given by

$$\begin{aligned}
\frac{\sqrt{625-(7+\Delta x)^2}-24}{\Delta x} &= \frac{\sqrt{625-(7+\Delta x)^2}-24}{\Delta x}\frac{\sqrt{625-(7+\Delta x)^2}+24}{\sqrt{625-(7+\Delta x)^2}+24} \\
&= \frac{625-(7+\Delta x)^2-24^2}{\Delta x(\sqrt{625-(7+\Delta x)^2}+24)} \\
&= \frac{49-49-14\Delta x-\Delta x^2}{\Delta x(\sqrt{625-(7+\Delta x)^2}+24)} \\
&= \frac{\Delta x(-14-\Delta x)}{\Delta x(\sqrt{625-(7+\Delta x)^2}+24)} \\
&= \frac{-14-\Delta x}{\sqrt{625-(7+\Delta x)^2}+24}
\end{aligned}$$

Now, can we tell by looking at this last formula what happens when Δx gets very close to zero? The numerator clearly gets very close to -14 while the denominator gets very close to $\sqrt{625-7^2}+24=48$. Is the fraction therefore very close to $-14/48=-7/24\cong-0.29167$? It certainly seems reasonable, and in fact it is true: as Δx gets closer and closer to zero, the difference quotient does in fact get closer and closer to $-7/24$, and so the slope of the tangent line is exactly $-7/24$.

What about the slope of the tangent line at $x=12$? Well, 12 can't be all that different from 7; we just have to redo the calculation with 12 instead of 7. This won't be hard, but it will be a bit tedious. What if we try to do all the algebra without using a specific value for x? Let's copy from above, replacing 7 by x. We'll have to do a bit more than that—for example, the "24" in the calculation came from $\sqrt{625-7^2}$, so we'll need to fix that too.

$$\begin{aligned}
&\frac{\sqrt{625-(x+\Delta x)^2}-\sqrt{625-x^2}}{\Delta x} = \\
&= \frac{\sqrt{625-(x+\Delta x)^2}-\sqrt{625-x^2}}{\Delta x}\frac{\sqrt{625-(x+\Delta x)^2}+\sqrt{625-x^2}}{\sqrt{625-(x+\Delta x)^2}+\sqrt{625-x^2}} \\
&= \frac{625-(x+\Delta x)^2-625+x^2}{\Delta x(\sqrt{625-(x+\Delta x)^2}+\sqrt{625-x^2})} \\
&= \frac{625-x^2-2x\Delta x-\Delta x^2-625+x^2}{\Delta x(\sqrt{625-(x+\Delta x)^2}+\sqrt{625-x^2})} \\
&= \frac{\Delta x(-2x-\Delta x)}{\Delta x(\sqrt{625-(x+\Delta x)^2}+\sqrt{625-x^2})} \\
&= \frac{-2x-\Delta x}{\sqrt{625-(x+\Delta x)^2}+\sqrt{625-x^2}}
\end{aligned}$$

Now what happens when Δx is very close to zero? Again it seems apparent that the quotient will be very close to

$$\frac{-2x}{\sqrt{625-x^2}+\sqrt{625-x^2}} = \frac{-2x}{2\sqrt{625-x^2}} = \frac{-x}{\sqrt{625-x^2}}.$$

Replacing x by 7 gives $-7/24$, as before, and now we can easily do the computation for 12 or any other value of x between -25 and 25.

So now we have a single, simple formula, $-x/\sqrt{625-x^2}$, that tells us the slope of the tangent line for any value of x. This slope, in turn, tells us how sensitive the value of y is to changes in the value of x.

What do we call such a formula? That is, a formula with one variable, so that substituting an "input" value for the variable produces a new "output" value? This is a function. Starting with one function, $\sqrt{625-x^2}$, we have derived, by means of some slightly nasty algebra, a new function, $-x/\sqrt{625-x^2}$, that gives us important information about the original function. This new function in fact is called the **derivative** of the original function. If the original is referred to as f or y then the derivative is often written f' or y' and pronounced "f prime" or "y prime", so in this case we might write $f'(x) = -x/\sqrt{625-x^2}$. At a particular point, say $x = 7$, we say that $f'(7) = -7/24$ or "f prime of 7 is $-7/24$" or "the derivative of f at 7 is $-7/24$."

To summarize, we compute the derivative of $f(x)$ by forming the difference quotient

$$\frac{f(x+\Delta x)-f(x)}{\Delta x}, \tag{2.1.1}$$

which is the slope of a line, then we figure out what happens when Δx gets very close to 0.

We should note that in the particular case of a circle, there's a simple way to find the derivative. Since the tangent to a circle at a point is perpendicular to the radius drawn to the point of contact, its slope is the negative reciprocal of the slope of the radius. The radius joining $(0,0)$ to $(7,24)$ has slope $24/7$. Hence, the tangent line has slope $-7/24$. In general, a radius to the point $(x,\sqrt{625-x^2})$ has slope $\sqrt{625-x^2}/x$, so the slope of the tangent line is $-x/\sqrt{625-x^2}$, as before. It is **NOT** always true that a tangent line is perpendicular to a line from the origin—don't use this shortcut in any other circumstance.

As above, and as you might expect, for different values of x we generally get different values of the derivative $f'(x)$. Could it be that the derivative always has the same value? This would mean that the slope of f, or the slope of its tangent line, is the same everywhere. One curve that always has the same slope is a line; it seems odd to talk about the tangent line to a line, but if it makes sense at all the tangent line must be the line itself. It is not hard to see that the derivative of $f(x) = mx + b$ is $f'(x) = m$; see exercise 6.

Exercises 2.1.

1. Draw the graph of the function $y = f(x) = \sqrt{169-x^2}$ between $x = 0$ and $x = 13$. Find the slope $\Delta y/\Delta x$ of the chord between the points of the circle lying over (a) $x = 12$ and $x = 13$, (b) $x = 12$ and $x = 12.1$, (c) $x = 12$ and $x = 12.01$, (d) $x = 12$ and $x = 12.001$. Now use the geometry of tangent lines on a circle to find (e) the exact value of the derivative $f'(12)$. Your answers to (a)–(d) should be getting closer and closer to your answer to (e). ⇒

2. Use geometry to find the derivative $f'(x)$ of the function $f(x) = \sqrt{625-x^2}$ in the text for each of the following x: (a) 20, (b) 24, (c) -7, (d) -15. Draw a graph of the upper semicircle, and draw the tangent line at each of these four points. ⇒

3. Draw the graph of the function $y = f(x) = 1/x$ between $x = 1/2$ and $x = 4$. Find the slope of the chord between (a) $x = 3$ and $x = 3.1$, (b) $x = 3$ and $x = 3.01$, (c) $x = 3$ and $x = 3.001$. Now use algebra to find a simple formula for the slope of the chord between $(3, f(3))$ and $(3+\Delta x, f(3+\Delta x))$. Determine what happens when Δx approaches 0. In your graph of $y = 1/x$, draw the straight line through the point $(3, 1/3)$ whose slope is this limiting value of the difference quotient as Δx approaches 0. ⇒

4. Find an algebraic expression for the difference quotient $\big(f(1+\Delta x)-f(1)\big)/\Delta x$ when $f(x) = x^2 - (1/x)$. Simplify the expression as much as possible. Then determine what happens as Δx approaches 0. That value is $f'(1)$. ⇒

5. Draw the graph of $y = f(x) = x^3$ between $x = 0$ and $x = 1.5$. Find the slope of the chord between (a) $x = 1$ and $x = 1.1$, (b) $x = 1$ and $x = 1.001$, (c) $x = 1$ and $x = 1.00001$. Then use algebra to find a simple formula for the slope of the chord between 1 and $1+\Delta x$. (Use the expansion $(A+B)^3 = A^3 + 3A^2B + 3AB^2 + B^3$.) Determine what happens as Δx approaches 0, and in your graph of $y = x^3$ draw the straight line through the point $(1,1)$ whose slope is equal to the value you just found. ⇒

6. Find an algebraic expression for the difference quotient $(f(x+\Delta x)-f(x))/\Delta x$ when $f(x) = mx + b$. Simplify the expression as much as possible. Then determine what happens as Δx approaches 0. That value is $f'(x)$. ⇒

7. Sketch the unit circle. Discuss the behavior of the slope of the tangent line at various angles around the circle. Which trigonometric function gives the slope of the tangent line at an angle θ? Why? Hint: think in terms of ratios of sides of triangles.

8. Sketch the parabola $y = x^2$. For what values of x on the parabola is the slope of the tangent line positive? Negative? What do you notice about the graph at the point(s) where the sign of the slope changes from positive to negative and vice versa?

2.2 An example

We started the last section by saying, "It is often necessary to know how sensitive the value of y is to small changes in x." We have seen one purely mathematical example of this: finding the "steepness" of a curve at a point is precisely this problem. Here is a more applied example.

With careful measurement it might be possible to discover that a dropped ball has height $h(t) = h_0 - kt^2$, t seconds after it is released. (Here h_0 is the initial height of the ball, when $t = 0$, and k is some number determined by the experiment.) A natural question is then, "How fast is the ball going at time t?" We can certainly get a pretty good idea with a little simple arithmetic. To make the calculation more concrete, let's say $h_0 = 100$ meters and $k = 4.9$ and suppose we're interested in the speed at $t = 2$. We know that when $t = 2$ the height is $100 - 4 \cdot 4.9 = 80.4$. A second later, at $t = 3$, the height is $100 - 9 \cdot 4.9 = 55.9$, so in that second the ball has traveled $80.4 - 55.9 = 24.5$ meters. This means that the *average* speed during that time was 24.5 meters per second. So we might guess that 24.5 meters per second is not a terrible estimate of the speed at $t = 2$. But certainly we can do better. At $t = 2.5$ the height is $100 - 4.9(2.5)^2 = 69.375$. During the half second from $t = 2$ to $t = 2.5$ the ball dropped $80.4 - 69.375 = 11.025$ meters, at an average speed of $11.025/(1/2) = 22.05$ meters per second; this should be a better estimate of the speed at $t = 2$. So it's clear now how to get better and better approximations: compute average speeds over shorter and shorter time intervals. Between $t = 2$ and $t = 2.01$, for example, the ball drops 0.19649 meters in one hundredth of a second, at an average speed of 19.649 meters per second.

We can't do this forever, and we still might reasonably ask what the actual speed precisely at $t = 2$ is. If Δt is some tiny amount of time, what we want to know is what happens to the average speed $(h(2) - h(2+\Delta t))/\Delta t$ as Δt gets smaller and smaller. Doing a bit of algebra:

$$\begin{aligned}
\frac{h(2)-h(2+\Delta t)}{\Delta t} &= \frac{80.4-(100-4.9(2+\Delta t)^2)}{\Delta t}\\
&= \frac{80.4-100+19.6+19.6\Delta t+4.9\Delta t^2}{\Delta t}\\
&= \frac{19.6\Delta t+4.9\Delta t^2}{\Delta t}\\
&= 19.6+4.9\Delta t
\end{aligned}$$

When Δt is very small, this is very close to 19.6, and indeed it seems clear that as Δt goes to zero, the average speed goes to 19.6, so the exact speed at $t=2$ is 19.6 meters per second. This calculation should look very familiar. In the language of the previous section, we might have started with $f(x)=100-4.9x^2$ and asked for the slope of the tangent line at $x=2$. We would have answered that question by computing

$$\frac{f(2+\Delta x)-f(2)}{\Delta x} = \frac{-19.6\Delta x-4.9\Delta x^2}{\Delta x} = -19.6-4.9\Delta x$$

The algebra is the same, except that following the pattern of the previous section the subtraction would be reversed, and we would say that the slope of the tangent line is -19.6. Indeed, in hindsight, perhaps we should have subtracted the other way even for the dropping ball. At $t=2$ the height is 80.4; one second later the height is 55.9. The usual way to compute a "distance traveled" is to subtract the earlier position from the later one, or $55.9-80.4=-24.5$. This tells us that the distance traveled is 24.5 meters, and the negative sign tells us that the height went down during the second. If we continue the original calculation we then get -19.6 meters per second as the exact speed at $t=2$. If we interpret the negative sign as meaning that the motion is downward, which seems reasonable, then in fact this is the same answer as before, but with even more information, since the numerical answer contains the direction of motion as well as the speed. Thus, the speed of the ball is the value of the derivative of a certain function, namely, of the function that gives the position of the ball. (More properly, this is the *velocity* of the ball; velocity is signed speed, that is, speed with a direction indicated by the sign.)

The upshot is that this problem, finding the speed of the ball, is *exactly* the same problem mathematically as finding the slope of a curve. This may already be enough evidence to convince you that whenever some quantity is changing (the height of a curve or the height of a ball or the size of the economy or the distance of a space probe from earth or the population of the world) the rate at which the quantity is changing can, in principle, be computed in exactly the same way, by finding a derivative.

Exercises 2.2.

1. An object is traveling in a straight line so that its position (that is, distance from some fixed point) is given by this table:

time (seconds)	0	1	2	3
distance (meters)	0	10	25	60

Find the average speed of the object during the following time intervals: $[0,1]$, $[0,2]$, $[0,3]$, $[1,2]$, $[1,3]$, $[2,3]$. If you had to guess the speed at $t = 2$ just on the basis of these, what would you guess? ⇒

2. Let $y = f(t) = t^2$, where t is the time in seconds and y is the distance in meters that an object falls on a certain airless planet. Draw a graph of this function between $t = 0$ and $t = 3$. Make a table of the average speed of the falling object between (a) 2 sec and 3 sec, (b) 2 sec and 2.1 sec, (c) 2 sec and 2.01 sec, (d) 2 sec and 2.001 sec. Then use algebra to find a simple formula for the average speed between time 2 and time $2 + \Delta t$. (If you substitute $\Delta t = 1$, 0.1, 0.01, 0.001 in this formula you should again get the answers to parts (a)–(d).) Next, in your formula for average speed (which should be in simplified form) determine what happens as Δt approaches zero. This is the instantaneous speed. Finally, in your graph of $y = t^2$ draw the straight line through the point $(2,4)$ whose slope is the instantaneous velocity you just computed; it should of course be the tangent line. ⇒

3. If an object is dropped from an 80-meter high window, its height y above the ground at time t seconds is given by the formula $y = f(t) = 80-4.9t^2$. (Here we are neglecting air resistance; the graph of this function was shown in figure 1.0.1.) Find the average velocity of the falling object between (a) 1 sec and 1.1 sec, (b) 1 sec and 1.01 sec, (c) 1 sec and 1.001 sec. Now use algebra to find a simple formula for the average velocity of the falling object between 1 sec and $1 + \Delta t$ sec. Determine what happens to this average velocity as Δt approaches 0. That is the instantaneous velocity at time $t = 1$ second (it will be negative, because the object is falling). ⇒

2.3 Limits

In the previous two sections we computed some quantities of interest (slope, velocity) by seeing that some expression "goes to" or "approaches" or "gets really close to" a particular value. In the examples we saw, this idea may have been clear enough, but it is too fuzzy to rely on in more difficult circumstances. In this section we will see how to make the idea more precise.

There is an important feature of the examples we have seen. Consider again the formula

$$\frac{-19.6\Delta x - 4.9\Delta x^2}{\Delta x}.$$

We wanted to know what happens to this fraction as "Δx goes to zero." Because we were able to simplify the fraction, it was easy to see the answer, but it was not quite as simple as "substituting zero for Δx," as that would give

$$\frac{-19.6 \cdot 0 - 4.9 \cdot 0}{0},$$

which is meaningless. The quantity we are really interested in does not make sense "at zero," and this is why the answer to the original problem (finding a velocity or a slope) was not immediately obvious. In other words, we are generally going to want to figure out what a quantity "approaches" in situations where we can't merely plug in a value. If you would like to think about a hard example (which we will analyze later) consider what happens to $(\sin x)/x$ as x approaches zero.

EXAMPLE 2.3.1 Does $\sqrt{x}$ approach 1.41 as x approaches 2? In this case it is possible to compute the actual value $\sqrt{2}$ to a high precision to answer the question. But since in general we won't be able to do that, let's not. We might start by computing $\sqrt{x}$ for values of x close to 2, as we did in the previous sections. Here are some values: $\sqrt{2.05} = 1.431782106$, $\sqrt{2.04} = 1.428285686$, $\sqrt{2.03} = 1.424780685$, $\sqrt{2.02} = 1.421267040$, $\sqrt{2.01} = 1.417744688$, $\sqrt{2.005} = 1.415980226$, $\sqrt{2.004} = 1.415627070$, $\sqrt{2.003} = 1.415273825$, $\sqrt{2.002} = 1.414920492$, $\sqrt{2.001} = 1.414567072$. So it looks at least possible that indeed these values "approach" 1.41—already $\sqrt{2.001}$ is quite close. If we continue this process, however, at some point we will appear to "stall." In fact, $\sqrt{2} = 1.414213562\ldots$, so we will never even get as far as 1.4142, no matter how long we continue the sequence. □

So in a fuzzy, everyday sort of sense, it is true that $\sqrt{x}$ "gets close to" 1.41, but it does not "approach" 1.41 in the sense we want. To compute an exact slope or an exact velocity, what we want to know is that a given quantity becomes "arbitrarily close" to a fixed value, meaning that the first quantity can be made "as close as we like" to the fixed value. Consider again the quantities

$$\frac{-19.6\Delta x - 4.9\Delta x^2}{\Delta x} = -19.6 - 4.9\Delta x.$$

These two quantities are equal as long as Δx is not zero; if Δx is zero, the left hand quantity is meaningless, while the right hand one is -19.6. Can we say more than we did before about why the right hand side "approaches" -19.6, in the desired sense? Can we really make it "as close as we want" to -19.6? Let's try a test case. Can we make $-19.6 - 4.9\Delta x$ within one millionth (0.000001) of -19.6? The values within a millionth of -19.6 are those in the interval $(-19.600001, -19.599999)$. As Δx approaches zero, does $-19.6 - 4.9\Delta x$ eventually reside inside this interval? If Δx is positive, this would require that $-19.6 - 4.9\Delta x > -19.600001$. This is something we can manipulate with a little algebra:

$$\begin{aligned} -19.6 - 4.9\Delta x &> -19.600001 \\ -4.9\Delta x &> -0.000001 \\ \Delta x &< -0.000001/-4.9 \\ \Delta x &< 0.0000002040816327\ldots \end{aligned}$$

Thus, we can say with certainty that if Δx is positive and less than 0.0000002, then $\Delta x < 0.0000002040816327\ldots$ and so $-19.6 - 4.9\Delta x > -19.600001$. We could do a similar calculation if Δx is negative.

So now we know that we can make $-19.6 - 4.9\Delta x$ within one millionth of -19.6. But can we make it "as close as we want"? In this case, it is quite simple to see that the answer is yes, by modifying the calculation we've just done. It may be helpful to think of this as a game. I claim that I can make $-19.6 - 4.9\Delta x$ as close as you desire to -19.6 by making Δx "close enough" to zero. So the game is: you give me a number, like 10^{-6}, and I have to come up with a number representing how close Δx must be to zero to guarantee that $-19.6 - 4.9\Delta x$ is at least as close to -19.6 as you have requested.

Now if we actually play this game, I could redo the calculation above for each new number you provide. What I'd like to do is somehow see that I will always succeed, and even more, I'd like to have a simple strategy so that I don't have to do all that algebra every time. A strategy in this case would be a formula that gives me a correct answer no matter what you specify. So suppose the number you give me is ϵ. How close does Δx have to be to zero to guarantee that $-19.6 - 4.9\Delta x$ is in $(-19.6 - \epsilon, -19.6 + \epsilon)$? If Δx is positive, we need:

$$\begin{aligned} -19.6 - 4.9\Delta x &> -19.6 - \epsilon \\ -4.9\Delta x &> -\epsilon \\ \Delta x &< -\epsilon/-4.9 \\ \Delta x &< \epsilon/4.9 \end{aligned}$$

So if I pick any number δ that is less than $\epsilon/4.9$, the algebra tells me that whenever $\Delta x < \delta$ then $\Delta x < \epsilon/4.9$ and so $-19.6 - 4.9\Delta x$ is within ϵ of -19.6. (This is exactly what I did in the example: I picked $\delta = 0.0000002 < 0.0000002040816327\ldots$.) A similar calculation again works for negative Δx. The important fact is that this is now a completely general result—it shows that I can always win, no matter what "move" you make.

Now we can codify this by giving a precise definition to replace the fuzzy, "gets closer and closer" language we have used so far. Henceforward, we will say something like "the limit of $(-19.6\Delta x - 4.9\Delta x^2)/\Delta x$ as Δx goes to zero is -19.6," and abbreviate this mouthful as

$$\lim_{\Delta x\to 0} \frac{-19.6\Delta x - 4.9\Delta x^2}{\Delta x} = -19.6.$$

Here is the actual, official definition of "limit".

DEFINITION 2.3.2 Limit Suppose f is a function. We say that $\lim_{x\to a} f(x) = L$ if for every $\epsilon > 0$ there is a $\delta > 0$ so that whenever $0 < |x - a| < \delta$, $|f(x) - L| < \epsilon$. □

The ϵ and δ here play exactly the role they did in the preceding discussion. The definition says, in a very precise way, that $f(x)$ can be made as close as desired to L (that's the $|f(x) - L| < \epsilon$ part) by making x close enough to a (the $0 < |x - a| < \delta$ part). Note that we specifically make no mention of what must happen if $x = a$, that is, if $|x - a| = 0$. This is because in the cases we are most interested in, substituting a for x doesn't even make sense.

Make sure you are not confused by the names of important quantities. The generic definition talks about $f(x)$, but the function and the variable might have other names. In the discussion above, the function we analyzed was

$$\frac{-19.6\Delta x - 4.9\Delta x^2}{\Delta x}.$$

and the variable of the limit was not x but Δx. The x was the variable of the original function; when we were trying to compute a slope or a velocity, x was essentially a fixed quantity, telling us at what point we wanted the slope. (In the velocity problem, it was literally a fixed quantity, as we focused on the time 2.) The quantity a of the definition in all the examples was zero: we were always interested in what happened as Δx became very close to zero.

Armed with a precise definition, we can now prove that certain quantities behave in a particular way. The bad news is that even proofs for simple quantities can be quite tedious and complicated; the good news is that we rarely need to do such proofs, because most expressions act the way you would expect, and this can be proved once and for all.

EXAMPLE 2.3.3 Let's show carefully that $\lim_{x\to 2} x + 4 = 6$. This is not something we "need" to prove, since it is "obviously" true. But if we couldn't prove it using our official definition there would be something very wrong with the definition.

As is often the case in mathematical proofs, it helps to work backwards. We want to end up showing that under certain circumstances $x + 4$ is close to 6; precisely, we want to show that $|x + 4 - 6| < \epsilon$, or $|x - 2| < \epsilon$. Under what circumstances? We want this to be true whenever $0 < |x - 2| < \delta$. So the question becomes: can we choose a value for δ that guarantees that $0 < |x - 2| < \delta$ implies $|x - 2| < \epsilon$? Of course: no matter what ϵ is, $\delta = \epsilon$ works. □

So it turns out to be very easy to prove something "obvious," which is nice. It doesn't take long before things get trickier, however.

EXAMPLE 2.3.4 It seems clear that $\lim_{x\to 2} x^2 = 4$. Let's try to prove it. We will want to be able to show that $|x^2 - 4| < \epsilon$ whenever $0 < |x - 2| < \delta$, by choosing δ carefully. Is there any connection between $|x-2|$ and $|x^2-4|$? Yes, and it's not hard to spot, but it is not so simple as the previous example. We can write $|x^2-4| = |(x+2)(x-2)|$. Now when $|x-2|$ is small, part of $|(x+2)(x-2)|$ is small, namely $(x-2)$. What about $(x+2)$? If x is close to 2, $(x+2)$ certainly can't be too big, but we need to somehow be precise about it. Let's recall the "game" version of what is going on here. You get to pick an ϵ and I have to pick a δ that makes things work out. Presumably it is the really tiny values of ϵ I need to worry about, but I have to be prepared for anything, even an apparently "bad" move like $\epsilon = 1000$. I expect that ϵ is going to be small, and that the corresponding δ will be small, certainly less than 1. If $\delta \le 1$ then $|x+2| < 5$ when $|x-2| < \delta$ (because if x is within 1 of 2, then x is between 1 and 3 and $x+2$ is between 3 and 5). So then I'd be trying to show that $|(x+2)(x-2)| < 5|x-2| < \epsilon$. So now how can I pick δ so that $|x-2| < \delta$ implies $5|x-2| < \epsilon$? This is easy: use $\delta = \epsilon/5$, so $5|x-2| < 5(\epsilon/5) = \epsilon$. But what if the ϵ you choose is not small? If you choose $\epsilon = 1000$, should I pick $\delta = 200$? No, to keep things "sane" I will never pick a δ bigger than 1. Here's the final "game strategy:" When you pick a value for ϵ I will pick $\delta = \epsilon/5$ or $\delta = 1$, whichever is smaller. Now when $|x-2| < \delta$, I know both that $|x+2| < 5$ and that $|x-2| < \epsilon/5$. Thus $|(x+2)(x-2)| < 5(\epsilon/5) = \epsilon$.

This has been a long discussion, but most of it was explanation and scratch work. If this were written down as a proof, it would be quite short, like this:

Proof that $\lim_{x\to 2} x^2 = 4$. Given any ϵ, pick $\delta = \epsilon/5$ or $\delta = 1$, whichever is smaller. Then when $|x-2| < \delta$, $|x+2| < 5$ and $|x-2| < \epsilon/5$. Hence $|x^2-4| = |(x+2)(x-2)| < 5(\epsilon/5) = \epsilon$. □

It probably seems obvious that $\lim_{x\to 2} x^2 = 4$, and it is worth examining more closely why it seems obvious. If we write $x^2 = x \cdot x$, and ask what happens when x approaches 2, we might say something like, "Well, the first x approaches 2, and the second x approaches 2, so the product must approach $2 \cdot 2$." In fact this is pretty much right on the money, except for that word "must." Is it really true that if x approaches a and y approaches b then xy approaches ab? It is, but it is not really obvious, since x and y might be quite complicated. The good news is that we can see that this is true once and for all, and then we don't have to worry about it ever again. When we say that x might be "complicated" we really mean that in practice it might be a function. Here is then what we want to know:

THEOREM 2.3.5 Suppose $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$. Then $\lim_{x\to a} f(x)g(x) = LM$.

Proof. We have to use the official definition of limit to make sense of this. So given any ϵ we need to find a δ so that $0 < |x-a| < \delta$ implies $|f(x)g(x) - LM| < \epsilon$. What do we have to work with? We know that we can make $f(x)$ close to L and $g(x)$ close to M, and we have to somehow connect these facts to make $f(x)g(x)$ close to LM.

We use, as is so often the case, a little algebraic trick:

$$\begin{aligned} |f(x)g(x) - LM| &= |f(x)g(x) - f(x)M + f(x)M - LM| \\ &= |f(x)(g(x) - M) + (f(x) - L)M| \\ &\le |f(x)(g(x) - M)| + |(f(x) - L)M| \\ &= |f(x)||g(x) - M| + |f(x) - L||M|. \end{aligned}$$

This is all straightforward except perhaps for the "$\le$". That is an example of the **triangle inequality** , which says that if a and b are any real numbers then $|a+b| \le |a| + |b|$. If you look at a few examples, using positive and negative numbers in various combinations for a and b, you should quickly understand why this is true; we will not prove it formally.

Since $\lim_{x\to a} f(x) = L$, there is a value δ_1 so that $0 < |x-a| < \delta_1$ implies $|f(x) - L| < |\epsilon/(2M)|$, This means that $0 < |x-a| < \delta_1$ implies $|f(x) - L||M| < \epsilon/2$. You can see where this is going: if we can make $|f(x)||g(x) - M| < \epsilon/2$ also, then we'll be done.

We can make $|g(x) - M|$ smaller than any fixed number by making x close enough to a; unfortunately, $\epsilon/(2f(x))$ is not a fixed number, since x is a variable. Here we need another little trick, just like the one we used in analyzing x^2. We can find a δ_2 so that $|x-a| < \delta_2$ implies that $|f(x) - L| < 1$, meaning that $L - 1 < f(x) < L + 1$. This means that $|f(x)| < N$, where N is either $|L-1|$ or $|L+1|$, depending on whether L is negative or positive. The important point is that N doesn't depend on x. Finally, we know that there is a δ_3 so that $0 < |x-a| < \delta_3$ implies $|g(x) - M| < \epsilon/(2N)$. Now we're ready to put everything together. Let δ be the smallest of δ_1, δ_2, and δ_3. Then $|x-a| < \delta$ implies that $|f(x) - L| < |\epsilon/(2M)|$, $|f(x)| < N$, and $|g(x) - M| < \epsilon/(2N)$. Then

$$\begin{aligned} |f(x)g(x) - LM| &\le |f(x)||g(x) - M| + |f(x) - L||M| \\ &< N\frac{\epsilon}{2N} + \left|\frac{\epsilon}{2M}\right| |M| \\ &= \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

This is just what we needed, so by the official definition, $\lim_{x\to a} f(x)g(x) = LM$. ∎

A handful of such theorems give us the tools to compute many limits without explicitly working with the definition of limit.

THEOREM 2.3.6 Suppose that $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$ and k is some constant. Then

$$\lim_{x\to a} kf(x) = k\lim_{x\to a} f(x) = kL$$

$$\lim_{x\to a} (f(x)+g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x) = L + M$$

$$\lim_{x\to a} (f(x)-g(x)) = \lim_{x\to a} f(x) - \lim_{x\to a} g(x) = L - M$$

$$\lim_{x\to a} (f(x)g(x)) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x) = LM$$

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} = \frac{L}{M}, \text{ if } M \text{ is not } 0$$

■

Roughly speaking, these rules say that to compute the limit of an algebraic expression, it is enough to compute the limits of the "innermost bits" and then combine these limits. This often means that it is possible to simply plug in a value for the variable, since $\lim_{x\to a} x = a$.

EXAMPLE 2.3.7 Compute $\lim_{x\to 1} \frac{x^2-3x+5}{x-2}$. If we apply the theorem in all its gory detail, we get

$$\begin{aligned}
\lim_{x\to 1} \frac{x^2-3x+5}{x-2} &= \frac{\lim_{x\to 1}(x^2-3x+5)}{\lim_{x\to 1}(x-2)} \\
&= \frac{(\lim_{x\to 1} x^2) - (\lim_{x\to 1} 3x) + (\lim_{x\to 1} 5)}{(\lim_{x\to 1} x) - (\lim_{x\to 1} 2)} \\
&= \frac{(\lim_{x\to 1} x)^2 - 3(\lim_{x\to 1} x) + 5}{(\lim_{x\to 1} x) - 2} \\
&= \frac{1^2 - 3\cdot 1 + 5}{1-2} \\
&= \frac{1-3+5}{-1} = -3
\end{aligned}$$

□

It is worth commenting on the trivial limit $\lim_{x\to 1} 5$. From one point of view this might seem meaningless, as the number 5 can't "approach" any value, since it is simply a fixed number. But 5 can, and should, be interpreted here as the function that has value 5 everywhere, $f(x) = 5$, with graph a horizontal line. From this point of view it makes sense to ask what happens to the height of the function as x approaches 1.

Of course, as we've already seen, we're primarily interested in limits that aren't so easy, namely, limits in which a denominator approaches zero. There are a handful of algebraic tricks that work on many of these limits.

EXAMPLE 2.3.8 Compute $\lim_{x\to 1} \frac{x^2+2x-3}{x-1}$. We can't simply plug in $x = 1$ because that makes the denominator zero. However:

$$\begin{aligned} \lim_{x\to 1} \frac{x^2+2x-3}{x-1} &= \lim_{x\to 1} \frac{(x-1)(x+3)}{x-1} \\ &= \lim_{x\to 1} (x+3) = 4 \end{aligned}$$

□

While theorem 2.3.6 is very helpful, we need a bit more to work easily with limits. Since the theorem applies when some limits are already known, we need to know the behavior of some functions that cannot themselves be constructed from the simple arithmetic operations of the theorem, such as $\sqrt{x}$. Also, there is one other extraordinarily useful way to put functions together: composition. If $f(x)$ and $g(x)$ are functions, we can form two functions by composition: $f(g(x))$ and $g(f(x))$. For example, if $f(x) = \sqrt{x}$ and $g(x) = x^2 + 5$, then $f(g(x)) = \sqrt{x^2+5}$ and $g(f(x)) = (\sqrt{x})^2 + 5 = x + 5$. Here is a companion to theorem 2.3.6 for composition:

THEOREM 2.3.9 Suppose that $\lim_{x\to a} g(x) = L$ and $\lim_{x\to L} f(x) = f(L)$. Then

$$\lim_{x\to a} f(g(x)) = f(L).$$

■

Note the special form of the condition on f: it is not enough to know that $\lim_{x\to L} f(x) = M$, though it is a bit tricky to see why. Many of the most familiar functions do have this property, and this theorem can therefore be applied. For example:

THEOREM 2.3.10 Suppose that n is a positive integer. Then

$$\lim_{x\to a} \sqrt[n]{x} = \sqrt[n]{a},$$

provided that a is positive if n is even. ■

This theorem is not too difficult to prove from the definition of limit.

Another of the most common algebraic tricks was used in section 2.1. Here's another example:

EXAMPLE 2.3.11 Compute $\lim_{x\to -1} \frac{\sqrt{x+5}-2}{x+1}$.

$$\begin{aligned}
\lim_{x\to -1} \frac{\sqrt{x+5}-2}{x+1} &= \lim_{x\to -1} \frac{\sqrt{x+5}-2}{x+1}\frac{\sqrt{x+5}+2}{\sqrt{x+5}+2} \\
&= \lim_{x\to -1} \frac{x+5-4}{(x+1)(\sqrt{x+5}+2)} \\
&= \lim_{x\to -1} \frac{x+1}{(x+1)(\sqrt{x+5}+2)} \\
&= \lim_{x\to -1} \frac{1}{\sqrt{x+5}+2} = \frac{1}{4}
\end{aligned}$$

At the very last step we have used theorems 2.3.9 and 2.3.10. □

Occasionally we will need a slightly modified version of the limit definition. Consider the function $f(x) = \sqrt{1-x^2}$, the upper half of the unit circle. What can we say about $\lim_{x\to 1} f(x)$? It is apparent from the graph of this familiar function that as x gets close to 1 from the left, the value of $f(x)$ gets close to zero. It does not even make sense to ask what happens as x approaches 1 from the right, since $f(x)$ is not defined there. The definition of the limit, however, demands that $f(1+\Delta x)$ be close to $f(1)$ whether Δx is positive or negative. Sometimes the limit of a function exists from one side or the other (or both) even though the limit does not exist. Since it is useful to be able to talk about this situation, we introduce the concept of **one sided limit**:

DEFINITION 2.3.12 **One-sided limit** Suppose that $f(x)$ is a function. We say that $\lim_{x\to a^-} f(x) = L$ if for every $\epsilon > 0$ there is a $\delta > 0$ so that whenever $0 < a - x < \delta$, $|f(x) - L| < \epsilon$. We say that $\lim_{x\to a^+} f(x) = L$ if for every $\epsilon > 0$ there is a $\delta > 0$ so that whenever $0 < x - a < \delta$, $|f(x) - L| < \epsilon$. □

Usually $\lim_{x\to a^-} f(x)$ is read "the limit of $f(x)$ from the left" and $\lim_{x\to a^+} f(x)$ is read "the limit of $f(x)$ from the right".

EXAMPLE 2.3.13 Discuss $\lim_{x\to 0} \frac{x}{|x|}$, $\lim_{x\to 0^-} \frac{x}{|x|}$, and $\lim_{x\to 0^+} \frac{x}{|x|}$.

The function $f(x) = x/|x|$ is undefined at 0; when $x > 0$, $|x| = x$ and so $f(x) = 1$; when $x < 0$, $|x| = -x$ and $f(x) = -1$. Thus $\lim_{x\to 0^-} \frac{x}{|x|} = \lim_{x\to 0^-} -1 = -1$ while $\lim_{x\to 0^+} \frac{x}{|x|} =$

$\lim_{x \to 0^+} 1 = 1$. The limit of $f(x)$ must be equal to both the left and right limits; since they are different, the limit $\lim_{x \to 0} \frac{x}{|x|}$ does not exist. □

Exercises 2.3.

Compute the limits. If a limit does not exist, explain why.

1. $\lim_{x \to 3} \frac{x^2 + x - 12}{x - 3}$ ⇒

2. $\lim_{x \to 1} \frac{x^2 + x - 12}{x - 3}$ ⇒

3. $\lim_{x \to -4} \frac{x^2 + x - 12}{x - 3}$ ⇒

4. $\lim_{x \to 2} \frac{x^2 + x - 12}{x - 2}$ ⇒

5. $\lim_{x \to 1} \frac{\sqrt{x + 8} - 3}{x - 1}$ ⇒

6. $\lim_{x \to 0^+} \sqrt{\frac{1}{x} + 2} - \sqrt{\frac{1}{x}}$. ⇒

7. $\lim_{x \to 2} 3$ ⇒

8. $\lim_{x \to 4} 3x^3 - 5x$ ⇒

9. $\lim_{x \to 0} \frac{4x - 5x^2}{x - 1}$ ⇒

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

11. $\lim_{x \to 0^+} \frac{\sqrt{2 - x^2}}{x}$ ⇒

12. $\lim_{x \to 0^+} \frac{\sqrt{2 - x^2}}{x + 1}$ ⇒

13. $\lim_{x \to a} \frac{x^3 - a^3}{x - a}$ ⇒

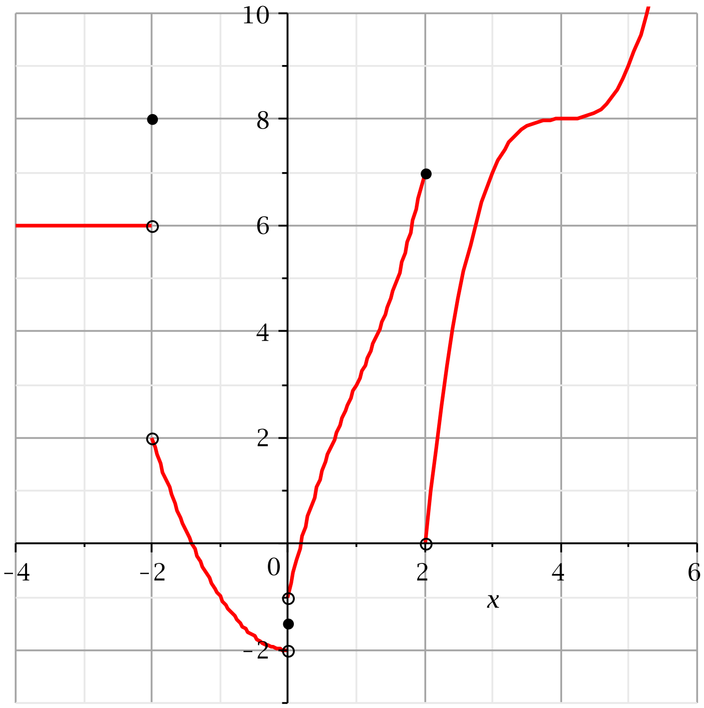
14. $\lim_{x \to 2} (x^2 + 4)^3$ ⇒

15. $\lim_{x \to 1} \begin{cases} x - 5 & x \neq 1, \\ 7 & x = 1. \end{cases}$ ⇒

16. $\lim_{x \to 0} x \sin\left(\frac{1}{x}\right)$ (Hint: Use the fact that $|\sin a| < 1$ for any real number a. You should probably use the definition of a limit here.) ⇒

17. Give an ϵ–δ proof, similar to example 2.3.3, of the fact that $\lim_{x \to 4} (2x - 5) = 3$.

18. Evaluate the expressions by reference to this graph:



(a) $\lim_{x\to 4} f(x)$ (b) $\lim_{x\to -3} f(x)$ (c) $\lim_{x\to 0} f(x)$

(d) $\lim_{x\to 0^-} f(x)$ (e) $\lim_{x\to 0^+} f(x)$ (f) $f(-2)$

(g) $\lim_{x\to 2^-} f(x)$ (h) $\lim_{x\to -2^-} f(x)$ (i) $\lim_{x\to 0} f(x+1)$

(j) $f(0)$ (k) $\lim_{x\to 1^-} f(x-4)$ (l) $\lim_{x\to 0^+} f(x-2)$

⇒

19. Use a calculator to estimate $\lim_{x\to 0} \dfrac{\sin x}{x}$.

20. Use a calculator to estimate $\lim_{x\to 0} \dfrac{\tan(3x)}{\tan(5x)}$.

2.4 The Derivative Function

We have seen how to create, or derive, a new function $f'(x)$ from a function $f(x)$, summarized in the paragraph containing equation 2.1.1. Now that we have the concept of limits, we can make this more precise.

DEFINITION 2.4.1 The derivative of a function f, denoted f', is

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

□

We know that f' carries important information about the original function f. In one example we saw that $f'(x)$ tells us how steep the graph of $f(x)$ is; in another we saw that $f'(x)$ tells us the velocity of an object if $f(x)$ tells us the position of the object at time x. As we said earlier, this same mathematical idea is useful whenever $f(x)$ represents some changing quantity and we want to know something about how it changes, or roughly, the "rate" at which it changes. Most functions encountered in practice are built up from a small collection of "primitive" functions in a few simple ways, for example, by adding or multiplying functions together to get new, more complicated functions. To make good use of the information provided by $f'(x)$ we need to be able to compute it for a variety of such functions.

We will begin to use different notations for the derivative of a function. While initially confusing, each is often useful so it is worth maintaining multiple versions of the same thing.

Consider again the function $f(x) = \sqrt{625 - x^2}$. We have computed the derivative $f'(x) = -x/\sqrt{625 - x^2}$, and have already noted that if we use the alternate notation $y = \sqrt{625 - x^2}$ then we might write $y' = -x/\sqrt{625 - x^2}$. Another notation is quite different, and in time it will become clear why it is often a useful one. Recall that to compute the the derivative of f we computed

$$\lim_{\Delta x \to 0} \frac{\sqrt{625 - (7 + \Delta x)^2} - 24}{\Delta x}.$$

The denominator here measures a distance in the x direction, sometimes called the "run", and the numerator measures a distance in the y direction, sometimes called the "rise," and "rise over run" is the slope of a line. Recall that sometimes such a numerator is abbreviated Δy, exchanging brevity for a more detailed expression. So in general, a derivative is given by

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}.$$

To recall the form of the limit, we sometimes say instead that

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}.$$

In other words, dy/dx is another notation for the derivative, and it reminds us that it is related to an actual slope between two points. This notation is called **Leibniz notation**, after Gottfried Leibniz, who developed the fundamentals of calculus independently, at about the same time that Isaac Newton did. Again, since we often use f and $f(x)$ to mean the original function, we sometimes use df/dx and $df(x)/dx$ to refer to the derivative. If the function $f(x)$ is written out in full we often write the last of these something like this

$$f'(x) = \frac{d}{dx}\sqrt{625 - x^2}$$

with the function written to the side, instead of trying to fit it into the numerator.

EXAMPLE 2.4.2 Find the derivative of $y = f(t) = t^2$.

We compute

$$\begin{aligned} y' = \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} &= \lim_{\Delta t \to 0} \frac{(t + \Delta t)^2 - t^2}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{t^2 + 2t\Delta t + \Delta t^2 - t^2}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{2t\Delta t + \Delta t^2}{\Delta t} \\ &= \lim_{\Delta t \to 0} 2t + \Delta t = 2t. \end{aligned}$$

Remember that Δt is a single quantity, not a "Δ" times a "t", and so Δt^2 is $(\Delta t)^2$ not $\Delta(t^2)$. □

EXAMPLE 2.4.3 Find the derivative of $y = f(x) = 1/x$.

The computation:

$$\begin{aligned} y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} &= \lim_{\Delta x \to 0} \frac{\frac{1}{x+\Delta x} - \frac{1}{x}}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\frac{x}{x(x+\Delta x)} - \frac{x+\Delta x}{x(x+\Delta x)}}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\frac{x-(x+\Delta x)}{x(x+\Delta x)}}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{x - x - \Delta x}{x(x + \Delta x)\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{-\Delta x}{x(x + \Delta x)\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{-1}{x(x + \Delta x)} = \frac{-1}{x^2} \end{aligned}$$

□

Note. If you happen to know some "derivative formulas" from an earlier course, for the time being you should pretend that you do not know them. In examples like the ones above and the exercises below, you are required to know how to find the derivative formula starting from basic principles. We will later develop some formulas so that we do not always need to do such computations, but we will continue to need to know how to do the more involved computations.

Sometimes one encounters a point in the domain of a function $y = f(x)$ where there is *no derivative*, because there is no tangent line. In order for the notion of the tangent line at a point to make sense, the curve must be "smooth" at that point. This means that if you imagine a particle traveling at some steady speed along the curve, then the particle does not experience an abrupt change of direction. There are two types of situations you should be aware of—corners and cusps—where there's a sudden change of direction and hence no derivative.

EXAMPLE 2.4.4 Discuss the derivative of the absolute value function $y = f(x) = |x|$.

If x is positive, then this is the function $y = x$, whose derivative is the constant 1. (Recall that when $y = f(x) = mx + b$, the derivative is the slope m.) If x is negative, then we're dealing with the function $y = -x$, whose derivative is the constant -1. If $x = 0$, then the function has a corner, i.e., there is no tangent line. A tangent line would have to point in the direction of the curve—but there are *two* directions of the curve that come together at the origin. We can summarize this as

$$y' = \begin{cases} 1 & \text{if } x > 0; \\ -1 & \text{if } x < 0; \\ \text{undefined} & \text{if } x = 0. \end{cases}$$

□

EXAMPLE 2.4.5

Discuss the derivative of the function $y = x^{2/3}$, shown in figure 2.4.1. We will later see how to compute this derivative; for now we use the fact that $y' = (2/3)x^{-1/3}$. Visually this looks much like the absolute value function, but it technically has a cusp, not a corner. The absolute value function has no tangent line at 0 because there are (at least) two obvious contenders—the tangent line of the left side of the curve and the tangent line of the right side. The function $y = x^{2/3}$ does not have a tangent line at 0, but unlike the absolute value function it can be said to have a single direction: as we approach 0 from either side the tangent line becomes closer and closer to a vertical line; the curve is vertical at 0. But as before, if you imagine traveling along the curve, an abrupt change in direction is required at 0: a full 180 degree turn. □

In practice we won't worry much about the distinction between these examples; in both cases the function has a "sharp point" where there is no tangent line and no derivative.

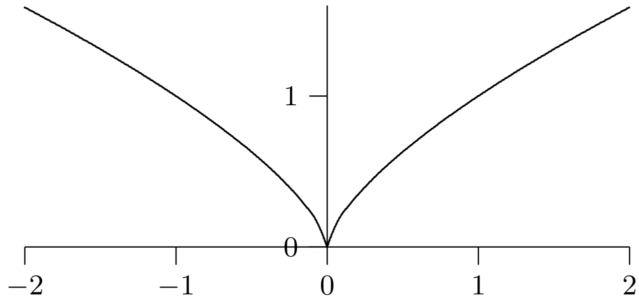
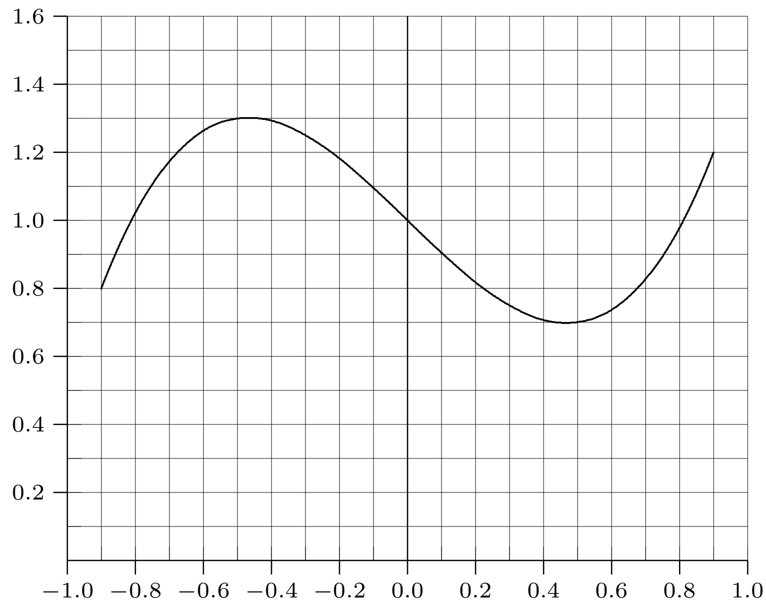


Figure 2.4.1 A cusp on $x^{2/3}$.

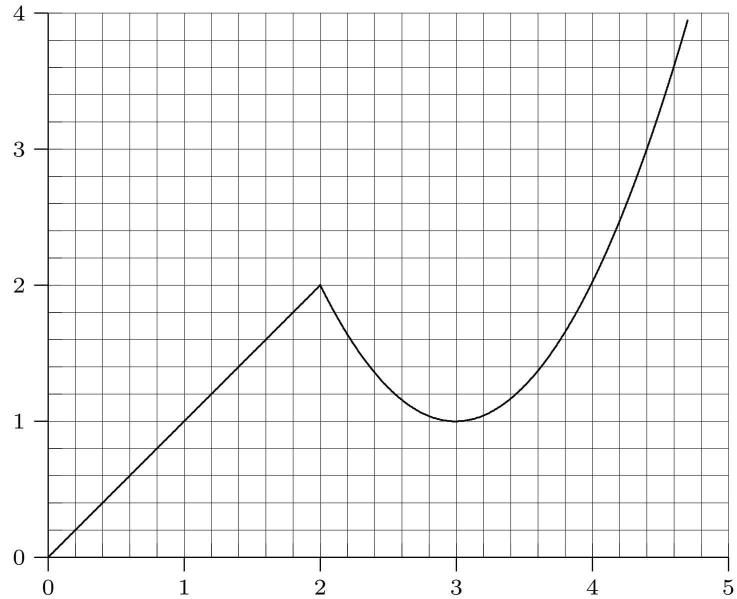
Exercises 2.4.

1. Find the derivative of $y = f(x) = \sqrt{169 - x^2}$. ⇒
2. Find the derivative of $y = f(t) = 80 - 4.9t^2$. ⇒
3. Find the derivative of $y = f(x) = x^2 - (1/x)$. ⇒
4. Find the derivative of $y = f(x) = ax^2 + bx + c$ (where a, b, and c are constants). ⇒
5. Find the derivative of $y = f(x) = x^3$. ⇒
6. Shown is the graph of a function $f(x)$. Sketch the graph of $f'(x)$ by estimating the derivative at a number of points in the interval: estimate the derivative at regular intervals from one end of the interval to the other, and also at "special" points, as when the derivative is zero. Make sure you indicate any places where the derivative does not exist.



7. Shown is the graph of a function $f(x)$. Sketch the graph of $f'(x)$ by estimating the derivative at a number of points in the interval: estimate the derivative at regular intervals from one end of the interval to the other, and also at "special" points, as when the derivative is zero.

Make sure you indicate any places where the derivative does not exist.



8. Find the derivative of $y = f(x) = 2/\sqrt{2x+1}$ ⇒

9. Find the derivative of $y = g(t) = (2t-1)/(t+2)$ ⇒

10. Find an equation for the tangent line to the graph of $f(x) = 5 - x - 3x^2$ at the point $x = 2$ ⇒

11. Find a value for a so that the graph of $f(x) = x^2 + ax - 3$ has a horizontal tangent line at $x = 4$. ⇒

2.5 Adjectives For Functions

As we have defined it in Section 1.3, a function is a very general object. At this point, it is useful to introduce a collection of adjectives to describe certain kinds of functions; these adjectives name useful properties that functions may have. Consider the graphs of the functions in Figure 2.5.1. It would clearly be useful to have words to help us describe the distinct features of each of them. We will point out and define a few adjectives (there are many more) for the functions pictured here. For the sake of the discussion, we will assume that the graphs do not exhibit any unusual behavior off-stage (i.e., outside the view of the graphs).

Functions. Each graph in Figure 2.5.1 certainly represents a function—since each passes the *vertical line test*. In other words, as you sweep a vertical line across the graph of each function, the line never intersects the graph more than once. If it did, then the graph would not represent a function.

Bounded. The graph in (c) appears to approach zero as x goes to both positive and negative infinity. It also never exceeds the value 1 or drops below the value 0. Because the

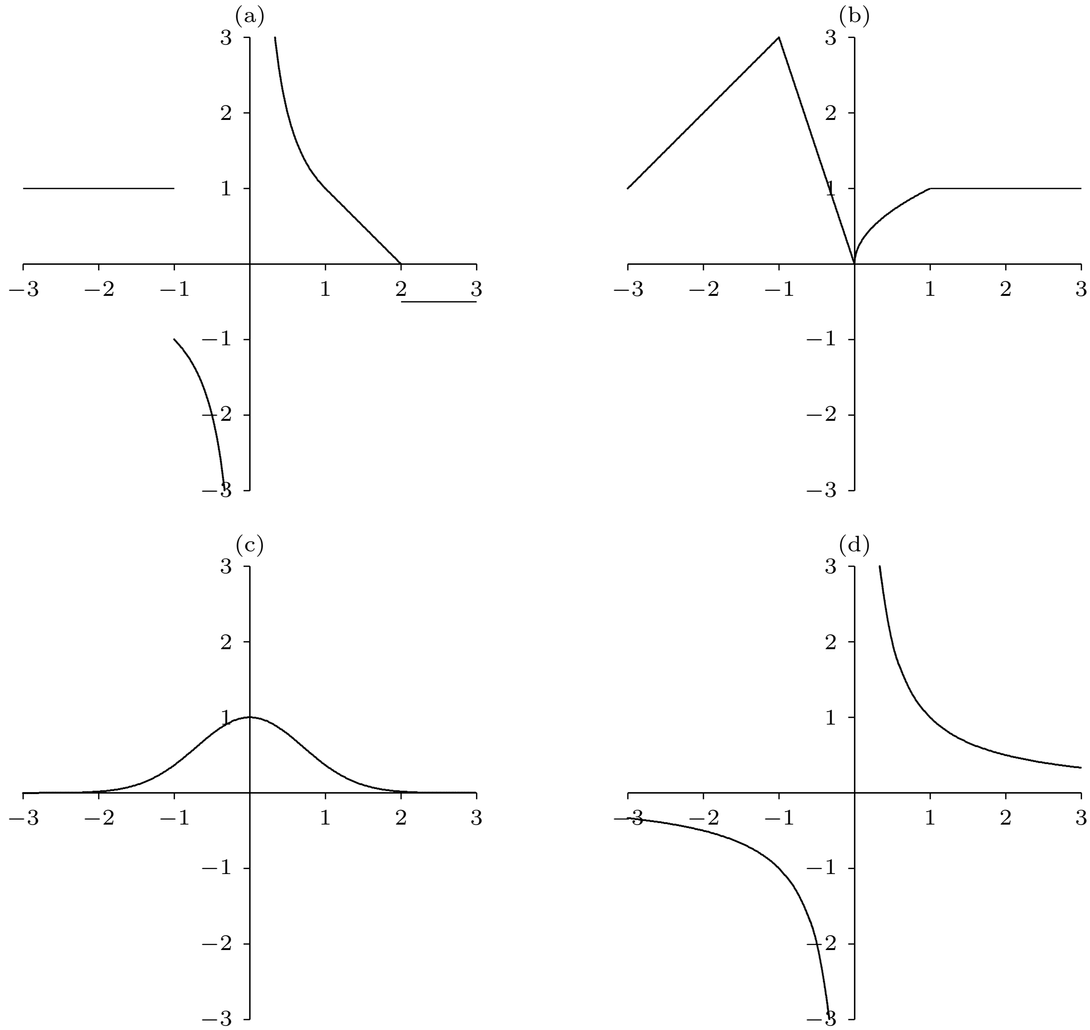


Figure 2.5.1 Function Types: (a) a discontinuous function, (b) a continuous function, (c) a bounded, differentiable function, (d) an unbounded, differentiable function

graph never increases or decreases without bound, we say that the function represented by the graph in (c) is a **bounded** function.

DEFINITION 2.5.1 Bounded A function f is bounded if there is a number M such that $|f(x)| < M$ for every x in the domain of f. □

For the function in (c), one such choice for M would be 10. However, the smallest (optimal) choice would be $M = 1$. In either case, simply finding an M is enough to establish boundedness. No such M exists for the hyperbola in (d) and hence we can say that it is **unbounded**.

Continuity. The graphs shown in (b) and (c) both represent **continuous** functions. Geometrically, this is because there are no jumps in the graphs. That is, if you pick a point on the graph and approach it from the left and right, the values of the function approach the value of the function at that point. For example, we can see that this is not true for function values near $x = -1$ on the graph in (a) which is not continuous at that location.

DEFINITION 2.5.2 Continuous at a Point A function f is continuous at a point a if $\lim_{x \to a} f(x) = f(a)$. □

DEFINITION 2.5.3 Continuous A function f is continuous if it is continuous at every point in its domain. □

Strangely, we can also say that (d) is continuous even though there is a vertical asymptote. A careful reading of the definition of continuous reveals the phrase "*at every point in its domain.*" Because the location of the asymptote, $x = 0$, is not in the domain of the function, and because the rest of the function is *well-behaved*, we can say that (d) is continuous.

Differentiability. Now that we have introduced the derivative of a function at a point, we can begin to use the adjective **differentiable**. We can see that the tangent line is well-defined at every point on the graph in (c). Therefore, we can say that (c) is a differentiable function.

DEFINITION 2.5.4 Differentiable at a Point A function f is differentiable at point a if $f'(a)$ exists. □

DEFINITION 2.5.5 Differentiable A function f is differentiable if is differentiable at every point (excluding endpoints and isolated points in the domain of f) in the domain of f. □

Take note that, for technical reasons not discussed here, both of these definitions exclude endpoints and isolated points in the domain from consideration.

We now have a collection of adjectives to describe the very rich and complex set of objects known as functions.

We close with a useful theorem about continuous functions:

THEOREM 2.5.6 Intermediate Value Theorem If f is continuous on the interval $[a, b]$ and d is between $f(a)$ and $f(b)$, then there is a number c in $[a, b]$ such that $f(c) = d$. ■

This is most frequently used when $d = 0$.

EXAMPLE 2.5.7 Explain why the function $f = x^3 + 3x^2 + x - 2$ has a root between 0 and 1.

By theorem 2.3.6, f is continuous. Since $f(0) = -2$ and $f(1) = 3$, and 0 is between -2 and 3, there is a $c \in [0, 1]$ such that $f(c) = 0$. □

This example also points the way to a simple method for approximating roots.

EXAMPLE 2.5.8 Approximate the root of the previous example to one decimal place.

If we compute $f(0.1)$, $f(0.2)$, and so on, we find that $f(0.6) < 0$ and $f(0.7) > 0$, so by the Intermediate Value Theorem, f has a root between 0.6 and 0.7. Repeating the process with $f(0.61)$, $f(0.62)$, and so on, we find that $f(0.61) < 0$ and $f(0.62) > 0$, so f has a root between 0.61 and 0.62, and the root is 0.6 rounded to one decimal place. □

Exercises 2.5.

1. Along the lines of Figure 2.5.1, for each part below sketch the graph of a function that is:

- **a.** bounded, but not continuous.
- **b.** differentiable and unbounded.
- **c.** continuous at $x = 0$, not continuous at $x = 1$, and bounded.
- **d.** differentiable everywhere except at $x = -1$, continuous, and unbounded.

2. Is $f(x) = \sin(x)$ a bounded function? If so, find the smallest M.

3. Is $s(t) = 1/(1 + t^2)$ a bounded function? If so, find the smallest M.

4. Is $v(u) = 2 \ln |u|$ a bounded function? If so, find the smallest M.

5. Consider the function

$$h(x) = \begin{cases} 2x - 3, & \text{if } x < 1 \\ 0, & \text{if } x \geq 1. \end{cases}$$

Show that it is continuous at the point $x = 0$. Is h a continuous function?

6. Approximate a root of $f = x^3 - 4x^2 + 2x + 2$ to one decimal place.

7. Approximate a root of $f = x^4 + x^3 - 5x + 1$ to one decimal place.

3

Rules for Finding Derivatives

It is tedious to compute a limit every time we need to know the derivative of a function. Fortunately, we can develop a small collection of examples and rules that allow us to compute the derivative of almost any function we are likely to encounter. Many functions involve quantities raised to a constant power, such as polynomials and more complicated combinations like $y = (\sin x)^4$. So we start by examining powers of a single variable; this gives us a building block for more complicated examples.

3.1 The Power Rule

We start with the derivative of a power function, $f(x) = x^n$. Here n is a number of any kind: integer, rational, positive, negative, even irrational, as in x^π. We have already computed some simple examples, so the formula should not be a complete surprise:

$$\frac{d}{dx}x^n = nx^{n-1}.$$

It is not easy to show this is true for any n. We will do some of the easier cases now, and discuss the rest later.

The easiest, and most common, is the case that n is a positive integer. To compute the derivative we need to compute the following limit:

$$\frac{d}{dx}x^n = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^n - x^n}{\Delta x}.$$

For a specific, fairly small value of n, we could do this by straightforward algebra.

EXAMPLE 3.1.1 Find the derivative of $f(x) = x^3$.

$$\begin{aligned}
\frac{d}{dx}x^3 &= \lim_{\Delta x\to 0} \frac{(x+\Delta x)^3 - x^3}{\Delta x}. \\
&= \lim_{\Delta x\to 0} \frac{x^3 + 3x^2\Delta x + 3x\Delta x^2 + \Delta x^3 - x^3}{\Delta x}. \\
&= \lim_{\Delta x\to 0} \frac{3x^2\Delta x + 3x\Delta x^2 + \Delta x^3}{\Delta x}. \\
&= \lim_{\Delta x\to 0} 3x^2 + 3x\Delta x + \Delta x^2 = 3x^2.
\end{aligned}$$

□

The general case is really not much harder as long as we don't try to do too much. The key is understanding what happens when $(x + \Delta x)^n$ is multiplied out:

$$(x+\Delta x)^n = x^n + nx^{n-1}\Delta x + a_2x^{n-2}\Delta x^2 + \cdots + +a_{n-1}x\Delta x^{n-1} + \Delta x^n.$$

We know that multiplying out will give a large number of terms all of the form $x^i\Delta x^j$, and in fact that $i+j=n$ in every term. One way to see this is to understand that one method for multiplying out $(x+\Delta x)^n$ is the following: In every $(x+\Delta x)$ factor, pick either the x or the Δx, then multiply the n choices together; do this in all possible ways. For example, for $(x+\Delta x)^3$, there are eight possible ways to do this:

$$\begin{aligned}
(x+\Delta x)(x+\Delta x)(x+\Delta x) &= xxx + xx\Delta x + x\Delta x x + x\Delta x\Delta x \\
&\quad + \Delta x xx + \Delta x x\Delta x + \Delta x\Delta x x + \Delta x\Delta x\Delta x \\
&= x^3 + x^2\Delta x + x^2\Delta x + x\Delta x^2 \\
&\quad + x^2\Delta x + x\Delta x^2 + x\Delta x^2 + \Delta x^3 \\
&= x^3 + 3x^2\Delta x + 3x\Delta x^2 + \Delta x^3
\end{aligned}$$

No matter what n is, there are n ways to pick Δx in one factor and x in the remaining $n-1$ factors; this means one term is $nx^{n-1}\Delta x$. The other coefficients are somewhat harder to understand, but we don't really need them, so in the formula above they have simply been called a_2, a_3, and so on. We know that every one of these terms contains Δx to at least the power 2. Now let's look at the limit:

$$\begin{aligned}
\frac{d}{dx}x^n &= \lim_{\Delta x\to 0} \frac{(x+\Delta x)^n - x^n}{\Delta x} \\
&= \lim_{\Delta x\to 0} \frac{x^n + nx^{n-1}\Delta x + a_2x^{n-2}\Delta x^2 + \cdots + a_{n-1}x\Delta x^{n-1} + \Delta x^n - x^n}{\Delta x} \\
&= \lim_{\Delta x\to 0} \frac{nx^{n-1}\Delta x + a_2x^{n-2}\Delta x^2 + \cdots + a_{n-1}x\Delta x^{n-1} + \Delta x^n}{\Delta x} \\
&= \lim_{\Delta x\to 0} nx^{n-1} + a_2x^{n-2}\Delta x + \cdots + a_{n-1}x\Delta x^{n-2} + \Delta x^{n-1} = nx^{n-1}.
\end{aligned}$$

Now without much trouble we can verify the formula for negative integers. First let's look at an example:

EXAMPLE 3.1.2 Find the derivative of $y = x^{-3}$. Using the formula, $y' = -3x^{-3-1} = -3x^{-4}$. □

Here is the general computation. Suppose n is a negative integer; the algebra is easier to follow if we use $n = -m$ in the computation, where m is a positive integer.

$$\begin{aligned}
\frac{d}{dx}x^n = \frac{d}{dx}x^{-m} &= \lim_{\Delta x\to 0} \frac{(x+\Delta x)^{-m} - x^{-m}}{\Delta x} \\
&= \lim_{\Delta x\to 0} \frac{\frac{1}{(x+\Delta x)^m} - \frac{1}{x^m}}{\Delta x} \\
&= \lim_{\Delta x\to 0} \frac{x^m - (x+\Delta x)^m}{(x+\Delta x)^m x^m \Delta x} \\
&= \lim_{\Delta x\to 0} \frac{x^m - (x^m + mx^{m-1}\Delta x + a_2 x^{m-2}\Delta x^2 + \cdots + a_{m-1}x\Delta x^{m-1} + \Delta x^m)}{(x+\Delta x)^m x^m \Delta x} \\
&= \lim_{\Delta x\to 0} \frac{-mx^{m-1} - a_2 x^{m-2}\Delta x - \cdots - a_{m-1}x\Delta x^{m-2} - \Delta x^{m-1})}{(x+\Delta x)^m x^m} \\
&= \frac{-mx^{m-1}}{x^m x^m} = \frac{-mx^{m-1}}{x^{2m}} = -mx^{m-1-2m} = nx^{-m-1} = nx^{n-1}.
\end{aligned}$$

We will later see why the other cases of the power rule work, but from now on we will use the power rule whenever n is any real number. Let's note here a simple case in which the power rule applies, or almost applies, but is not really needed. Suppose that $f(x) = 1$; remember that this "1" is a function, not "merely" a number, and that $f(x) = 1$ has a graph that is a horizontal line, with slope zero everywhere. So we know that $f'(x) = 0$. We might also write $f(x) = x^0$, though there is some question about just what this means at $x = 0$. If we apply the power rule, we get $f'(x) = 0x^{-1} = 0/x = 0$, again noting that there is a problem at $x = 0$. So the power rule "works" in this case, but it's really best to just remember that the derivative of any constant function is zero.

Exercises 3.1.

Find the derivatives of the given functions.

1. x^{100} ⇒

2. x^{-100} ⇒

3. $\dfrac{1}{x^5}$ ⇒

4. x^{π} ⇒

5. $x^{3/4}$ ⇒

6. $x^{-9/7}$ ⇒

3.2 Linearity of the Derivative

An operation is linear if it behaves "nicely" with respect to multiplication by a constant and addition. The name comes from the equation of a line through the origin, $f(x) = mx$, and the following two properties of this equation. First, $f(cx) = m(cx) = c(mx) = cf(x)$, so the constant c can be "moved outside" or "moved through" the function f. Second, $f(x+y) = m(x+y) = mx+my = f(x)+f(y)$, so the addition symbol likewise can be moved through the function.

The corresponding properties for the derivative are:

$$(cf(x))' = \frac{d}{dx}cf(x) = c\frac{d}{dx}f(x) = cf'(x),$$

and

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

It is easy to see, or at least to believe, that these are true by thinking of the distance/speed interpretation of derivatives. If one object is at position $f(t)$ at time t, we know its speed is given by $f'(t)$. Suppose another object is at position $5f(t)$ at time t, namely, that it is always 5 times as far along the route as the first object. Then it "must" be going 5 times as fast at all times.

The second rule is somewhat more complicated, but here is one way to picture it. Suppose a flat bed railroad car is at position $f(t)$ at time t, so the car is traveling at a speed of $f'(t)$ (to be specific, let's say that $f(t)$ gives the position on the track of the rear end of the car). Suppose that an ant is crawling from the back of the car to the front so that its position *on the car* is $g(t)$ and its speed *relative to the car* is $g'(t)$. Then in reality, at time t, the ant is at position $f(t)+g(t)$ along the track, and its speed is "obviously" $f'(t)+g'(t)$.

We don't want to rely on some more-or-less obvious physical interpretation to determine what is true mathematically, so let's see how to verify these rules by computation.

We'll do one and leave the other for the exercises.

$$\begin{aligned}
\frac{d}{dx}(f(x)+g(x)) &= \lim_{\Delta x\to 0}\frac{f(x+\Delta x)+g(x+\Delta x)-(f(x)+g(x))}{\Delta x}\\
&= \lim_{\Delta x\to 0}\frac{f(x+\Delta x)+g(x+\Delta x)-f(x)-g(x)}{\Delta x}\\
&= \lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)+g(x+\Delta x)-g(x)}{\Delta x}\\
&= \lim_{\Delta x\to 0}\left(\frac{f(x+\Delta x)-f(x)}{\Delta x}+\frac{g(x+\Delta x)-g(x)}{\Delta x}\right)\\
&= \lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}+\lim_{\Delta x\to 0}\frac{g(x+\Delta x)-g(x)}{\Delta x}\\
&= f'(x)+g'(x)
\end{aligned}$$

This is sometimes called the **sum rule** for derivatives.

EXAMPLE 3.2.1 Find the derivative of $f(x) = x^5 + 5x^2$. We have to invoke linearity twice here:

$$f'(x) = \frac{d}{dx}(x^5+5x^2) = \frac{d}{dx}x^5 + \frac{d}{dx}(5x^2) = 5x^4 + 5\frac{d}{dx}(x^2) = 5x^4 + 5\cdot 2x^1 = 5x^4 + 10x.$$

□

Because it is so easy with a little practice, we can usually combine all uses of linearity into a single step. The following example shows an acceptably detailed computation.

EXAMPLE 3.2.2 Find the derivative of $f(x) = 3/x^4 - 2x^2 + 6x - 7$.

$$f'(x) = \frac{d}{dx}\left(\frac{3}{x^4} - 2x^2 + 6x - 7\right) = \frac{d}{dx}(3x^{-4} - 2x^2 + 6x - 7) = -12x^{-5} - 4x + 6.$$

□

Exercises 3.2.

Find the derivatives of the functions in 1–6.

1. $5x^3 + 12x^2 - 15$ ⇒

2. $-4x^5 + 3x^2 - 5/x^2$ ⇒

3. $5(-3x^2 + 5x + 1)$ ⇒

4. $f(x) + g(x)$, where $f(x) = x^2 - 3x + 2$ and $g(x) = 2x^3 - 5x$ ⇒

5. $(x + 1)(x^2 + 2x - 3)$ ⇒

6. $\sqrt{625 - x^2} + 3x^3 + 12$ (See section 2.1.) ⇒

7. Find an equation for the tangent line to $f(x) = x^3/4 - 1/x$ at $x = -2$. ⇒

8. Find an equation for the tangent line to $f(x) = 3x^2 - \pi^3$ at $x = 4$. ⇒

9. Suppose the position of an object at time t is given by $f(t) = -49t^2/10 + 5t + 10$. Find a function giving the speed of the object at time t. The acceleration of an object is the rate at which its speed is changing, which means it is given by the derivative of the speed function. Find the acceleration of the object at time t. ⇒

10. Let $f(x) = x^3$ and $c = 3$. Sketch the graphs of f, cf, f', and $(cf)'$ on the same diagram.

11. The general polynomial P of degree n in the variable x has the form $P(x) = \sum_{k=0}^{n} a_k x^k = a_0 + a_1 x + \ldots + a_n x^n$. What is the derivative (with respect to x) of P? ⇒

12. Find a cubic polynomial whose graph has horizontal tangents at $(-2, 5)$ and $(2, 3)$. ⇒

13. Prove that $\frac{d}{dx}(cf(x)) = cf'(x)$ using the definition of the derivative.

14. Suppose that f and g are differentiable at x. Show that $f - g$ is differentiable at x using the two linearity properties from this section.

3.3 The Product Rule

Consider the product of two simple functions, say $f(x) = (x^2 + 1)(x^3 - 3x)$. An obvious guess for the derivative of f is the product of the derivatives of the constituent functions: $(2x)(3x^2 - 3) = 6x^3 - 6x$. Is this correct? We can easily check, by rewriting f and doing the calculation in a way that is known to work. First, $f(x) = x^5 - 3x^3 + x^3 - 3x = x^5 - 2x^3 - 3x$, and then $f'(x) = 5x^4 - 6x^2 - 3$. Not even close! What went "wrong"? Well, nothing really, except the guess was wrong.

So the derivative of $f(x)g(x)$ is NOT as simple as $f'(x)g'(x)$. Surely there is some rule for such a situation? There is, and it is instructive to "discover" it by trying to do the general calculation even without knowing the answer in advance.

$$
\begin{aligned}
\frac{d}{dx}(f(x)g(x)) &= \lim_{\Delta x \to 0} \frac{f(x+\Delta x)g(x+\Delta x) - f(x)g(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{f(x+\Delta x)g(x+\Delta x) - f(x+\Delta x)g(x) + f(x+\Delta x)g(x) - f(x)g(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{f(x+\Delta x)g(x+\Delta x) - f(x+\Delta x)g(x)}{\Delta x} + \lim_{\Delta x \to 0} \frac{f(x+\Delta x)g(x) - f(x)g(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} f(x+\Delta x)\frac{g(x+\Delta x) - g(x)}{\Delta x} + \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} g(x) \\
&= f(x)g'(x) + f'(x)g(x)
\end{aligned}
$$

A couple of items here need discussion. First, we used a standard trick, "add and subtract the same thing", to transform what we had into a more useful form. After some rewriting, we realize that we have two limits that produce $f'(x)$ and $g'(x)$. Of course, $f'(x)$ and

$g'(x)$ must actually exist for this to make sense. We also replaced $\lim_{\Delta x \to 0} f(x + \Delta x)$ with $f(x)$—why is this justified?

What we really need to know here is that $\lim_{\Delta x \to 0} f(x + \Delta x) = f(x)$, or in the language of section 2.5, that f is continuous at x. We already know that $f'(x)$ exists (or the whole approach, writing the derivative of fg in terms of f' and g', doesn't make sense). This turns out to imply that f is continuous as well. Here's why:

$$\begin{aligned}
\lim_{\Delta x \to 0} f(x + \Delta x) &= \lim_{\Delta x \to 0} (f(x + \Delta x) - f(x) + f(x)) \\
&= \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \Delta x + \lim_{\Delta x \to 0} f(x) \\
&= f'(x) \cdot 0 + f(x) = f(x)
\end{aligned}$$

To summarize: the product rule says that

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

Returning to the example we started with, let $f(x) = (x^2 + 1)(x^3 - 3x)$. Then $f'(x) = (x^2 + 1)(3x^2 - 3) + (2x)(x^3 - 3x) = 3x^4 - 3x^2 + 3x^2 - 3 + 2x^4 - 6x^2 = 5x^4 - 6x^2 - 3$, as before. In this case it is probably simpler to multiply $f(x)$ out first, then compute the derivative; here's an example for which we really need the product rule.

EXAMPLE 3.3.1 Compute the derivative of $f(x) = x^2\sqrt{625 - x^2}$. We have already computed $\frac{d}{dx}\sqrt{625 - x^2} = \frac{-x}{\sqrt{625 - x^2}}$. Now

$$f'(x) = x^2 \frac{-x}{\sqrt{625 - x^2}} + 2x\sqrt{625 - x^2} = \frac{-x^3 + 2x(625 - x^2)}{\sqrt{625 - x^2}} = \frac{-3x^3 + 1250x}{\sqrt{625 - x^2}}.$$

□

Exercises 3.3.

In 1–4, find the derivatives of the functions using the product rule.

1. $x^3(x^3 - 5x + 10)$ ⇒

2. $(x^2 + 5x - 3)(x^5 - 6x^3 + 3x^2 - 7x + 1)$ ⇒

3. $\sqrt{x}\sqrt{625 - x^2}$ ⇒

4. $\frac{\sqrt{625 - x^2}}{x^{20}}$ ⇒

5. Use the product rule to compute the derivative of $f(x) = (2x - 3)^2$. Sketch the function. Find an equation of the tangent line to the curve at $x = 2$. Sketch the tangent line at $x = 2$. ⇒

6. Suppose that f, g, and h are differentiable functions. Show that $(fgh)'(x) = f'(x)g(x)h(x) + f(x)g'(x)h(x) + f(x)g(x)h'(x)$.

7. State and prove a rule to compute $(fghi)'(x)$, similar to the rule in the previous problem.

Product notation. Suppose $f_1, f_2, \ldots f_n$ are functions. The product of all these functions can be written

$$\prod_{k=1}^{n} f_k.$$

This is similar to the use of $\sum$ to denote a sum. For example,

$$\prod_{k=1}^{5} f_k = f_1 f_2 f_3 f_4 f_5$$

and

$$\prod_{k=1}^{n} k = 1 \cdot 2 \cdot \ldots \cdot n = n!.$$

We sometimes use somewhat more complicated conditions; for example

$$\prod_{k=1,k\neq j}^{n} f_k$$

denotes the product of f_1 through f_n except for f_j. For example,

$$\prod_{k=1,k\neq 4}^{5} x^k = x \cdot x^2 \cdot x^3 \cdot x^5 = x^{11}.$$

8. The **generalized product rule** says that if $f_1, f_2, \ldots, f_n$ are differentiable functions at x then

$$\frac{d}{dx} \prod_{k=1}^{n} f_k(x) = \sum_{j=1}^{n} \left(f_j'(x) \prod_{k=1,k\neq j}^{n} f_k(x) \right).$$

Verify that this is the same as your answer to the previous problem when $n = 4$, and write out what this says when $n = 5$.

3.4 The Quotient Rule

What is the derivative of $(x^2 + 1)/(x^3 - 3x)$? More generally, we'd like to have a formula to compute the derivative of $f(x)/g(x)$ if we already know $f'(x)$ and $g'(x)$. Instead of attacking this problem head-on, let's notice that we've already done part of the problem: $f(x)/g(x) = f(x) \cdot (1/g(x))$, that is, this is "really" a product, and we can compute the derivative if we know $f'(x)$ and $(1/g(x))'$. So really the only new bit of information we need is $(1/g(x))'$ in terms of $g'(x)$. As with the product rule, let's set this up and see how far we can get:

$$\begin{aligned}
\frac{d}{dx}\frac{1}{g(x)} &= \lim_{\Delta x\to 0}\frac{\frac{1}{g(x+\Delta x)}-\frac{1}{g(x)}}{\Delta x}\\
&= \lim_{\Delta x\to 0}\frac{\frac{g(x)-g(x+\Delta x)}{g(x+\Delta x)g(x)}}{\Delta x}\\
&= \lim_{\Delta x\to 0}\frac{g(x)-g(x+\Delta x)}{g(x+\Delta x)g(x)\Delta x}\\
&= \lim_{\Delta x\to 0}-\frac{g(x+\Delta x)-g(x)}{\Delta x}\frac{1}{g(x+\Delta x)g(x)}\\
&= -\frac{g'(x)}{g(x)^2}
\end{aligned}$$

Now we can put this together with the product rule:

$$\frac{d}{dx}\frac{f(x)}{g(x)} = f(x)\frac{-g'(x)}{g(x)^2}+f'(x)\frac{1}{g(x)} = \frac{-f(x)g'(x)+f'(x)g(x)}{g(x)^2} = \frac{f'(x)g(x)-f(x)g'(x)}{g(x)^2}.$$

EXAMPLE 3.4.1 Compute the derivative of $(x^2+1)/(x^3-3x)$.

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

□

It is often possible to calculate derivatives in more than one way, as we have already seen. Since every quotient can be written as a product, it is always possible to use the product rule to compute the derivative, though it is not always simpler.

EXAMPLE 3.4.2 Find the derivative of $\sqrt{625-x^2}/\sqrt{x}$ in two ways: using the quotient rule, and using the product rule.

Quotient rule:

$$\frac{d}{dx}\frac{\sqrt{625-x^2}}{\sqrt{x}} = \frac{\sqrt{x}(-x/\sqrt{625-x^2})-\sqrt{625-x^2}\cdot 1/(2\sqrt{x})}{x}.$$

Note that we have used $\sqrt{x}=x^{1/2}$ to compute the derivative of $\sqrt{x}$ by the power rule.

Product rule:

$$\frac{d}{dx}\sqrt{625-x^2}x^{-1/2} = \sqrt{625-x^2}\frac{-1}{2}x^{-3/2}+\frac{-x}{\sqrt{625-x^2}}x^{-1/2}.$$

With a bit of algebra, both of these simplify to

$$-\frac{x^2+625}{2\sqrt{625-x^2}x^{3/2}}.$$

□

Occasionally you will need to compute the derivative of a quotient with a constant numerator, like $10/x^2$. Of course you can use the quotient rule, but it is usually not the easiest method. If we do use it here, we get

$$\frac{d}{dx}\frac{10}{x^2} = \frac{x^2 \cdot 0 - 10 \cdot 2x}{x^4} = \frac{-20}{x^3},$$

since the derivative of 10 is 0. But it is simpler to do this:

$$\frac{d}{dx}\frac{10}{x^2} = \frac{d}{dx}10x^{-2} = -20x^{-3}.$$

Admittedly, x^2 is a particularly simple denominator, but we will see that a similar calculation is usually possible. Another approach is to remember that

$$\frac{d}{dx}\frac{1}{g(x)} = \frac{-g'(x)}{g(x)^2},$$

but this requires extra memorization. Using this formula,

$$\frac{d}{dx}\frac{10}{x^2} = 10\frac{-2x}{x^4}.$$

Note that we first use linearity of the derivative to pull the 10 out in front.

Exercises 3.4.

Find the derivatives of the functions in 1–4 using the quotient rule.

1. $\dfrac{x^3}{x^3 - 5x + 10}$ ⇒

2. $\dfrac{x^2 + 5x - 3}{x^5 - 6x^3 + 3x^2 - 7x + 1}$ ⇒

3. $\dfrac{\sqrt{x}}{\sqrt{625 - x^2}}$ ⇒

4. $\dfrac{\sqrt{625 - x^2}}{x^{20}}$ ⇒

5. Find an equation for the tangent line to $f(x) = (x^2 - 4)/(5 - x)$ at $x = 3$. ⇒

6. Find an equation for the tangent line to $f(x) = (x - 2)/(x^3 + 4x - 1)$ at $x = 1$. ⇒

7. Let P be a polynomial of degree n and let Q be a polynomial of degree m (with Q not the zero polynomial). Using sigma notation we can write

$$P = \sum_{k=0}^{n} a_k x^k, \qquad Q = \sum_{k=0}^{m} b_k x^k.$$

Use sigma notation to write the derivative of the **rational function** P/Q.

8. The curve $y = 1/(1 + x^2)$ is an example of a class of curves each of which is called a **witch of Agnesi**. Sketch the curve and find the tangent line to the curve at $x = 5$. (The word *witch* here is a mistranslation of the original Italian, as described at

http://mathworld.wolfram.com/WitchofAgnesi.html

and

http://instructional1.calstatela.edu/sgray/Agnesi/WitchHistory/Historynamewitch.html.)

⇒

9. If $f'(4) = 5$, $g'(4) = 12$, $(fg)(4) = f(4)g(4) = 2$, and $g(4) = 6$, compute $f(4)$ and $\dfrac{d}{dx}\dfrac{f}{g}$ at 4. ⇒

3.5 The Chain Rule

So far we have seen how to compute the derivative of a function built up from other functions by addition, subtraction, multiplication and division. There is another very important way that we combine simple functions to make more complicated functions: function composition, as discussed in section 2.3. For example, consider $\sqrt{625-x^2}$. This function has many simpler components, like 625 and x^2, and then there is that square root symbol, so the square root function $\sqrt{x} = x^{1/2}$ is involved. The obvious question is: can we compute the derivative using the derivatives of the constituents $625 - x^2$ and $\sqrt{x}$? We can indeed. In general, if $f(x)$ and $g(x)$ are functions, we can compute the derivatives of $f(g(x))$ and $g(f(x))$ in terms of $f'(x)$ and $g'(x)$.

EXAMPLE 3.5.1 Form the two possible compositions of $f(x) = \sqrt{x}$ and $g(x) = 625 - x^2$ and compute the derivatives. First, $f(g(x)) = \sqrt{625-x^2}$, and the derivative is $-x/\sqrt{625-x^2}$ as we have seen. Second, $g(f(x)) = 625 - (\sqrt{x})^2 = 625 - x$ with derivative -1. Of course, these calculations do not use anything new, and in particular the derivative of $f(g(x))$ was somewhat tedious to compute from the definition. □

Suppose we want the derivative of $f(g(x))$. Again, let's set up the derivative and play some algebraic tricks:

$$\begin{aligned}\frac{d}{dx}f(g(x)) &= \lim_{\Delta x\to 0}\frac{f(g(x+\Delta x)) - f(g(x))}{\Delta x}\\ &= \lim_{\Delta x\to 0}\frac{f(g(x+\Delta x)) - f(g(x))}{g(x+\Delta x)) - g(x)}\,\frac{g(x+\Delta x)) - g(x)}{\Delta x}\end{aligned}$$

Now we see immediately that the second fraction turns into $g'(x)$ when we take the limit. The first fraction is more complicated, but it too looks something like a derivative. The denominator, $g(x + \Delta x)) - g(x)$, is a change in the value of g, so let's abbreviate it as

$\Delta g = g(x+\Delta x)) - g(x)$, which also means $g(x+\Delta x) = g(x) + \Delta g$. This gives us

$$\lim_{\Delta x\to 0} \frac{f(g(x)+\Delta g) - f(g(x))}{\Delta g}.$$

As Δx goes to 0, it is also true that Δg goes to 0, because $g(x+\Delta x)$ goes to $g(x)$. So we can rewrite this limit as

$$\lim_{\Delta g\to 0} \frac{f(g(x)+\Delta g) - f(g(x))}{\Delta g}.$$

Now this looks exactly like a derivative, namely $f'(g(x))$, that is, the function $f'(x)$ with x replaced by $g(x)$. If this all withstands scrutiny, we then get

$$\frac{d}{dx} f(g(x)) = f'(g(x))g'(x).$$

Unfortunately, there is a small flaw in the argument. Recall that what we mean by $\lim_{\Delta x\to 0}$ involves what happens when Δx is close to 0 *but not equal to 0.* The qualification is very important, since we must be able to divide by Δx. But when Δx is close to 0 but not equal to 0, $\Delta g = g(x+\Delta x)) - g(x)$ is close to 0 *and possibly equal to 0.* This means it doesn't really make sense to divide by Δg. Fortunately, it is possible to recast the argument to avoid this difficulty, but it is a bit tricky; we will not include the details, which can be found in many calculus books. Note that many functions g do have the property that $g(x+\Delta x) - g(x) \neq 0$ when Δx is small, and for these functions the argument above is fine.

The chain rule has a particularly simple expression if we use the Leibniz notation for the derivative. The quantity $f'(g(x))$ is the derivative of f with x replaced by g; this can be written df/dg. As usual, $g'(x) = dg/dx$. Then the chain rule becomes

$$\frac{df}{dx} = \frac{df}{dg}\frac{dg}{dx}.$$

This looks like trivial arithmetic, but it is not: dg/dx is not a fraction, that is, not literal division, but a single symbol that means $g'(x)$. Nevertheless, it turns out that what looks like trivial arithmetic, and is therefore easy to remember, is really true.

It will take a bit of practice to make the use of the chain rule come naturally—it is more complicated than the earlier differentiation rules we have seen.

EXAMPLE 3.5.2 Compute the derivative of $\sqrt{625-x^2}$. We already know that the answer is $-x/\sqrt{625-x^2}$, computed directly from the limit. In the context of the chain rule, we have $f(x) = \sqrt{x}$, $g(x) = 625 - x^2$. We know that $f'(x) = (1/2)x^{-1/2}$, so

$f'(g(x)) = (1/2)(625 - x^2)^{-1/2}$. Note that this is a two step computation: first compute $f'(x)$, then replace x by $g(x)$. Since $g'(x) = -2x$ we have

$$f'(g(x))g'(x) = \frac{1}{2\sqrt{625 - x^2}}(-2x) = \frac{-x}{\sqrt{625 - x^2}}.$$

□

EXAMPLE 3.5.3 Compute the derivative of $1/\sqrt{625 - x^2}$. This is a quotient with a constant numerator, so we could use the quotient rule, but it is simpler to use the chain rule. The function is $(625-x^2)^{-1/2}$, the composition of $f(x) = x^{-1/2}$ and $g(x) = 625-x^2$. We compute $f'(x) = (-1/2)x^{-3/2}$ using the power rule, and then

$$f'(g(x))g'(x) = \frac{-1}{2(625 - x^2)^{3/2}}(-2x) = \frac{x}{(625 - x^2)^{3/2}}.$$

□

In practice, of course, you will need to use more than one of the rules we have developed to compute the derivative of a complicated function.

EXAMPLE 3.5.4 Compute the derivative of

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

The "last" operation here is division, so to get started we need to use the quotient rule first. This gives

$$\begin{aligned} f'(x) &= \frac{(x^2-1)'x\sqrt{x^2+1} - (x^2-1)(x\sqrt{x^2+1})'}{x^2(x^2+1)} \\ &= \frac{2x^2\sqrt{x^2+1} - (x^2-1)(x\sqrt{x^2+1})'}{x^2(x^2+1)}. \end{aligned}$$

Now we need to compute the derivative of $x\sqrt{x^2+1}$. This is a product, so we use the product rule:

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

Finally, we use the chain rule:

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

And putting it all together:

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

This can be simplified of course, but we have done all the calculus, so that only algebra is left. □

EXAMPLE 3.5.5 Compute the derivative of $\sqrt{1+\sqrt{1+\sqrt{x}}}$. Here we have a more complicated chain of compositions, so we use the chain rule twice. At the outermost "layer" we have the function $g(x) = 1+\sqrt{1+\sqrt{x}}$ plugged into $f(x)=\sqrt{x}$, so applying the chain rule once gives

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

Now we need the derivative of $\sqrt{1+\sqrt{x}}$. Using the chain rule again:

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

So the original derivative is

$$\begin{aligned} \frac{d}{dx}\sqrt{1+\sqrt{1+\sqrt{x}}} &= \frac{1}{2}\left(1+\sqrt{1+\sqrt{x}}\right)^{-1/2}\frac{1}{2}\left(1+\sqrt{x}\right)^{-1/2}\frac{1}{2}x^{-1/2}. \\ &= \frac{1}{8\sqrt{x}\sqrt{1+\sqrt{x}}\sqrt{1+\sqrt{1+\sqrt{x}}}} \end{aligned}$$

□

Using the chain rule, the power rule, and the product rule, it is possible to avoid using the quotient rule entirely.

EXAMPLE 3.5.6 Compute the derivative of $f(x) = \dfrac{x^3}{x^2+1}$. Write $f(x) = x^3(x^2+1)^{-1}$, then

$$\begin{aligned} f'(x) &= x^3 \frac{d}{dx}(x^2+1)^{-1} + 3x^2(x^2+1)^{-1} \\ &= x^3(-1)(x^2+1)^{-2}(2x) + 3x^2(x^2+1)^{-1} \\ &= -2x^4(x^2+1)^{-2} + 3x^2(x^2+1)^{-1} \\ &= \frac{-2x^4}{(x^2+1)^2} + \frac{3x^2}{x^2+1} \\ &= \frac{-2x^4}{(x^2+1)^2} + \frac{3x^2(x^2+1)}{(x^2+1)^2} \\ &= \frac{-2x^4+3x^4+3x^2}{(x^2+1)^2} = \frac{x^4+3x^2}{(x^2+1)^2} \end{aligned}$$

Note that we already had the derivative on the second line; all the rest is simplification. It is easier to get to this answer by using the quotient rule, so there's a trade off: more work for fewer memorized formulas. □

Exercises 3.5.

Find the derivatives of the functions. For extra practice, and to check your answers, do some of these in more than one way if possible.

1. $x^4 - 3x^3 + (1/2)x^2 + 7x - \pi$ ⇒

2. $x^3 - 2x^2 + 4\sqrt{x}$ ⇒

3. $(x^2+1)^3$ ⇒

4. $x\sqrt{169-x^2}$ ⇒

5. $(x^2-4x+5)\sqrt{25-x^2}$ ⇒

6. $\sqrt{r^2-x^2}$, r is a constant ⇒

7. $\sqrt{1+x^4}$ ⇒

8. $\dfrac{1}{\sqrt{5-\sqrt{x}}}$. ⇒

9. $(1+3x)^2$ ⇒

10. $\dfrac{(x^2+x+1)}{(1-x)}$ ⇒

11. $\dfrac{\sqrt{25-x^2}}{x}$ ⇒

12. $\sqrt{\dfrac{169}{x} - x}$ ⇒

13. $\sqrt{x^3-x^2-(1/x)}$ ⇒

14. $100/(100-x^2)^{3/2}$ ⇒

15. $\sqrt[3]{x+x^3}$ ⇒

16. $\sqrt{(x^2+1)^2+\sqrt{1+(x^2+1)^2}}$ ⇒

17. $(x+8)^5$ ⇒

18. $(4-x)^3$ ⇒

19. $(x^2+5)^3$ ⇒

20. $(6-2x^2)^3$ ⇒

21. $(1-4x^3)^{-2}$ ⇒

22. $5(x+1-1/x)$ ⇒

23. $4(2x^2-x+3)^{-2}$ ⇒

24. $\dfrac{1}{1+1/x}$ ⇒

25. $\dfrac{-3}{4x^2-2x+1}$ ⇒

26. $(x^2+1)(5-2x)/2$ ⇒

27. $(3x^2+1)(2x-4)^3$ ⇒

28. $\dfrac{x+1}{x-1}$ ⇒

29. $\dfrac{x^2-1}{x^2+1}$ ⇒

30. $\dfrac{(x-1)(x-2)}{x-3}$ ⇒

31. $\dfrac{2x^{-1}-x^{-2}}{3x^{-1}-4x^{-2}}$ ⇒

32. $3(x^2+1)(2x^2-1)(2x+3)$ ⇒

33. $\dfrac{1}{(2x+1)(x-3)}$ ⇒

34. $((2x+1)^{-1}+3)^{-1}$ ⇒

35. $(2x+1)^3(x^2+1)^2$ ⇒

36. Find an equation for the tangent line to $f(x)=(x-2)^{1/3}/(x^3+4x-1)^2$ at $x=1$. ⇒

37. Find an equation for the tangent line to $y=9x^{-2}$ at $(3,1)$. ⇒

38. Find an equation for the tangent line to $(x^2-4x+5)\sqrt{25-x^2}$ at $(3,8)$. ⇒

39. Find an equation for the tangent line to $\dfrac{(x^2+x+1)}{(1-x)}$ at $(2,-7)$. ⇒

40. Find an equation for the tangent line to $\sqrt{(x^2+1)^2+\sqrt{1+(x^2+1)^2}}$ at $(1,\sqrt{4+\sqrt{5}})$. ⇒

4

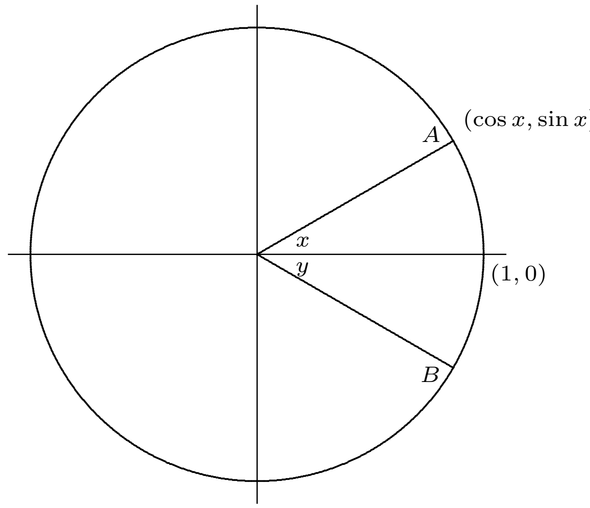
Transcendental Functions

So far we have used only algebraic functions as examples when finding derivatives, that is, functions that can be built up by the usual algebraic operations of addition, subtraction, multiplication, division, and raising to constant powers. Both in theory and practice there are other functions, called transcendental, that are very useful. Most important among these are the trigonometric functions, the inverse trigonometric functions, exponential functions, and logarithms.

4.1 Trigonometric Functions

When you first encountered the trigonometric functions it was probably in the context of "triangle trigonometry," defining, for example, the sine of an angle as the "side opposite over the hypotenuse." While this will still be useful in an informal way, we need to use a more expansive definition of the trigonometric functions. First an important note: while degree measure of angles is sometimes convenient because it is so familiar, it turns out to be ill-suited to mathematical calculation, so (almost) everything we do will be in terms of **radian measure** of angles.

To define the radian measurement system, we consider the unit circle in the xy-plane:

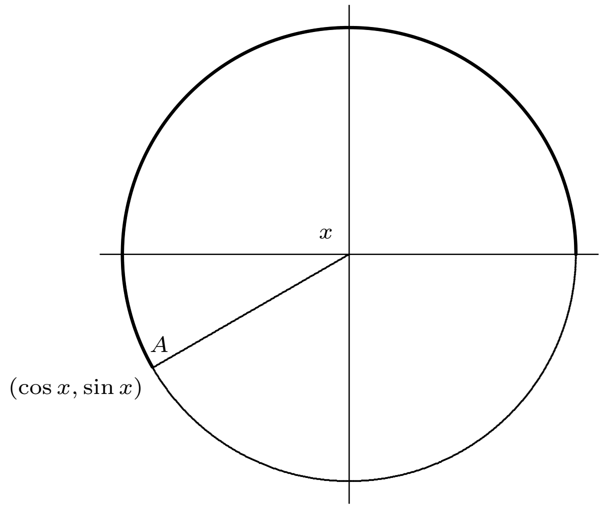


An angle, x, at the center of the circle is associated with an arc of the circle which is said to **subtend** the angle. In the figure, this arc is the portion of the circle from point $(1, 0)$ to point A. The length of this arc is the radian measure of the angle x; the fact that the radian measure is an actual geometric length is largely responsible for the usefulness of radian measure. The circumference of the unit circle is $2\pi r = 2\pi(1) = 2\pi$, so the radian measure of the full circular angle (that is, of the 360 degree angle) is 2π.

While an angle with a particular measure can appear anywhere around the circle, we need a fixed, conventional location so that we can use the coordinate system to define properties of the angle. The standard convention is to place the starting radius for the angle on the positive x-axis, and to measure positive angles counterclockwise around the circle. In the figure, x is the standard location of the angle $\pi/6$, that is, the length of the arc from $(1, 0)$ to A is $\pi/6$. The angle y in the picture is $-\pi/6$, because the distance from $(1, 0)$ to B along the circle is also $\pi/6$, but in a clockwise direction.

Now the fundamental trigonometric definitions are: the cosine of x and the sine of x are the first and second coordinates of the point A, as indicated in the figure. The angle x shown can be viewed as an angle of a right triangle, meaning the usual triangle definitions of the sine and cosine also make sense. Since the hypotenuse of the triangle is 1, the "side opposite over hypotenuse" definition of the sine is the second coordinate of point A over 1, which is just the second coordinate; in other words, both methods give the same value for the sine.

The simple triangle definitions work only for angles that can "fit" in a right triangle, namely, angles between 0 and $\pi/2$. The coordinate definitions, on the other hand, apply to any angles, as indicated in this figure:



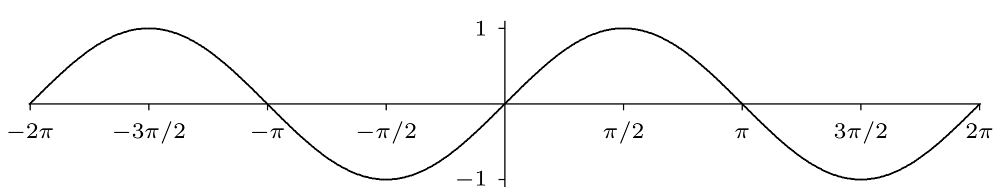
The angle x is subtended by the heavy arc in the figure, that is, $x = 7\pi/6$. Both coordinates of point A in this figure are negative, so the sine and cosine of $7\pi/6$ are both negative.

The remaining trigonometric functions can be most easily defined in terms of the sine and cosine, as usual:

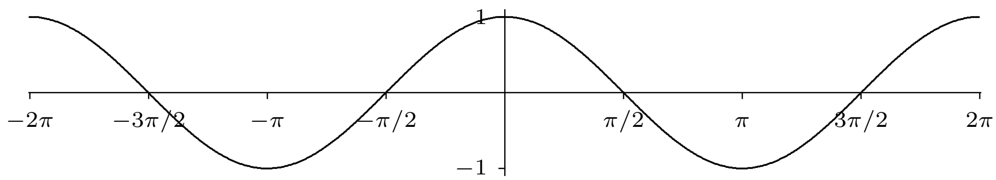
$$\tan x = \frac{\sin x}{\cos x}$$
$$\cot x = \frac{\cos x}{\sin x}$$
$$\sec x = \frac{1}{\cos x}$$
$$\csc x = \frac{1}{\sin x}$$

and they can also be defined as the corresponding ratios of coordinates.

Although the trigonometric functions are defined in terms of the unit circle, the unit circle diagram is not what we normally consider the graph of a trigonometric function. (The unit circle is the graph of, well, the circle.) We can easily get a qualitatively correct idea of the graphs of the trigonometric functions from the unit circle diagram. Consider the sine function, $y = \sin x$. As x increases from 0 in the unit circle diagram, the second coordinate of the point A goes from 0 to a maximum of 1, then back to 0, then to a minimum of -1, then back to 0, and then it obviously repeats itself. So the graph of $y = \sin x$ must look something like this:



Similarly, as angle x increases from 0 in the unit circle diagram, the first coordinate of the point A goes from 1 to 0 then to -1, back to 0 and back to 1, so the graph of $y = \cos x$ must look something like this:



Exercises 4.1.

Some useful trigonometric identities are in appendix B.

1. Find all values of θ such that $\sin(\theta) = -1$; give your answer in radians. ⇒

2. Find all values of θ such that $\cos(2\theta) = 1/2$; give your answer in radians. ⇒

3. Use an angle sum identity to compute $\cos(\pi/12)$. ⇒

4. Use an angle sum identity to compute $\tan(5\pi/12)$. ⇒

5. Verify the identity $\cos^2(t)/(1 - \sin(t)) = 1 + \sin(t)$.

6. Verify the identity $2\csc(2\theta) = \sec(\theta)\csc(\theta)$.

7. Verify the identity $\sin(3\theta) - \sin(\theta) = 2\cos(2\theta)\sin(\theta)$.

8. Sketch $y = 2\sin(x)$.

9. Sketch $y = \sin(3x)$.

10. Sketch $y = \sin(-x)$.

11. Find all of the solutions of $2\sin(t) - 1 - \sin^2(t) = 0$ in the interval $[0, 2\pi]$. ⇒

4.2 The Derivative of $\sin x$

What about the derivative of the sine function? The rules for derivatives that we have are no help, since $\sin x$ is not an algebraic function. We need to return to the definition of the derivative, set up a limit, and try to compute it. Here's the definition:

$$\frac{d}{dx}\sin x = \lim_{\Delta x \to 0} \frac{\sin(x + \Delta x) - \sin x}{\Delta x}.$$

Using some trigonometric identities, we can make a little progress on the quotient:

$$\begin{aligned}\frac{\sin(x + \Delta x) - \sin x}{\Delta x} &= \frac{\sin x \cos \Delta x + \sin \Delta x \cos x - \sin x}{\Delta x} \\ &= \sin x \frac{\cos \Delta x - 1}{\Delta x} + \cos x \frac{\sin \Delta x}{\Delta x}.\end{aligned}$$

This isolates the difficult bits in the two limits

$$\lim_{\Delta x \to 0} \frac{\cos \Delta x - 1}{\Delta x} \quad \text{and} \quad \lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x}.$$

Here we get a little lucky: it turns out that once we know the second limit the first is quite easy. The second is quite tricky, however. Indeed, it is the hardest limit we will actually compute, and we devote a section to it.

4.3 A hard limit

We want to compute this limit:

$$\lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x}.$$

Equivalently, to make the notation a bit simpler, we can compute

$$\lim_{x \to 0} \frac{\sin x}{x}.$$

In the original context we need to keep x and Δx separate, but here it doesn't hurt to rename Δx to something more convenient.

To do this we need to be quite clever, and to employ some indirect reasoning. The indirect reasoning is embodied in a theorem, frequently called the **squeeze theorem**.

THEOREM 4.3.1 Squeeze Theorem Suppose that $g(x) \le f(x) \le h(x)$ for all x close to a but not equal to a. If $\lim_{x\to a} g(x) = L = \lim_{x\to a} h(x)$, then $\lim_{x\to a} f(x) = L$. ■

This theorem can be proved using the official definition of limit. We won't prove it here, but point out that it is easy to understand and believe graphically. The condition says that $f(x)$ is trapped between $g(x)$ below and $h(x)$ above, and that at $x = a$, both g and h approach the same value. This means the situation looks something like figure 4.3.1. The wiggly curve is $x^2 \sin(\pi/x)$, the upper and lower curves are x^2 and $-x^2$. Since the sine function is always between -1 and 1, $-x^2 \le x^2 \sin(\pi/x) \le x^2$, and it is easy to see that $\lim_{x\to 0} -x^2 = 0 = \lim_{x\to 0} x^2$. It is not so easy to see directly, that is algebraically, that $\lim_{x\to 0} x^2 \sin(\pi/x) = 0$, because the π/x prevents us from simply plugging in $x = 0$. The squeeze theorem makes this "hard limit" as easy as the trivial limits involving x^2.

To do the hard limit that we want, $\lim_{x\to 0}(\sin x)/x$, we will find two simpler functions g and h so that $g(x) \le (\sin x)/x \le h(x)$, and so that $\lim_{x\to 0} g(x) = \lim_{x\to 0} h(x)$. Not too surprisingly, this will require some trigonometry and geometry. Referring to figure 4.3.2, x is the measure of the angle in radians. Since the circle has radius 1, the coordinates of

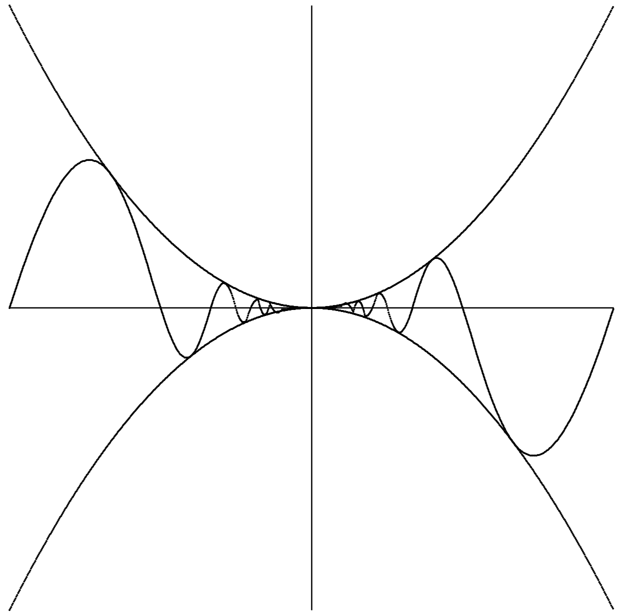


Figure 4.3.1 The squeeze theorem.

point A are $(\cos x, \sin x)$, and the area of the small triangle is $(\cos x \sin x)/2$. This triangle is completely contained within the circular wedge-shaped region bordered by two lines and the circle from $(1,0)$ to point A. Comparing the areas of the triangle and the wedge we see $(\cos x \sin x)/2 \le x/2$, since the area of a circular region with angle θ and radius r is $\theta r^2/2$. With a little algebra this turns into $(\sin x)/x \le 1/\cos x$, giving us the h we seek.

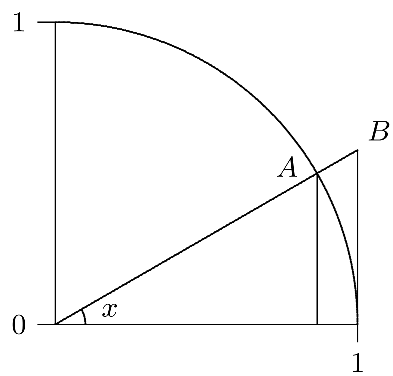


Figure 4.3.2 Visualizing $\sin x/x$.

To find g, we note that the circular wedge is completely contained inside the larger triangle. The height of the triangle, from $(1,0)$ to point B, is $\tan x$, so comparing areas we get $x/2 \le (\tan x)/2 = \sin x/(2\cos x)$. With a little algebra this becomes $\cos x \le (\sin x)/x$. So now we have

$$\cos x \le \frac{\sin x}{x} \le \frac{1}{\cos x}.$$

Finally, the two limits $\lim_{x\to 0}\cos x$ and $\lim_{x\to 0} 1/\cos x$ are easy, because $\cos(0) = 1$. By the squeeze theorem, $\lim_{x\to 0}(\sin x)/x = 1$ as well.

Before we can complete the calculation of the derivative of the sine, we need one other limit:

$$\lim_{x\to 0}\frac{\cos x - 1}{x}.$$

This limit is just as hard as $\sin x/x$, but closely related to it, so that we don't have to do a similar calculation; instead we can do a bit of tricky algebra.

$$\frac{\cos x - 1}{x} = \frac{\cos x - 1}{x}\frac{\cos x + 1}{\cos x + 1} = \frac{\cos^2 x - 1}{x(\cos x + 1)} = \frac{-\sin^2 x}{x(\cos x + 1)} = -\frac{\sin x}{x}\frac{\sin x}{\cos x + 1}.$$

To compute the desired limit it is sufficient to compute the limits of the two final fractions, as x goes to 0. The first of these is the hard limit we've just done, namely 1. The second turns out to be simple, because the denominator presents no problem:

$$\lim_{x\to 0}\frac{\sin x}{\cos x + 1} = \frac{\sin 0}{\cos 0 + 1} = \frac{0}{2} = 0.$$

Thus,

$$\lim_{x\to 0}\frac{\cos x - 1}{x} = 0.$$

Exercises 4.3.

1. Compute $\displaystyle\lim_{x\to 0}\frac{\sin(5x)}{x}$ ⇒

2. Compute $\displaystyle\lim_{x\to 0}\frac{\sin(7x)}{\sin(2x)}$ ⇒

3. Compute $\displaystyle\lim_{x\to 0}\frac{\cot(4x)}{\csc(3x)}$ ⇒

4. Compute $\displaystyle\lim_{x\to 0}\frac{\tan x}{x}$ ⇒

5. Compute $\displaystyle\lim_{x\to \pi/4}\frac{\sin x - \cos x}{\cos(2x)}$ ⇒

6. For all $x \ge 0$, $4x - 9 \le f(x) \le x^2 - 4x + 7$. Find $\displaystyle\lim_{x\to 4} f(x)$. ⇒

7. For all x, $2x \le g(x) \le x^4 - x^2 + 2$. Find $\displaystyle\lim_{x\to 1} g(x)$. ⇒

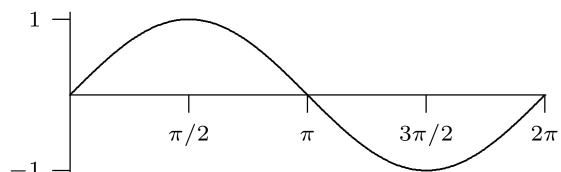
8. Use the Squeeze Theorem to show that $\displaystyle\lim_{x\to 0} x^4\cos(2/x) = 0$.

4.4 The Derivative of $\sin x$, continued

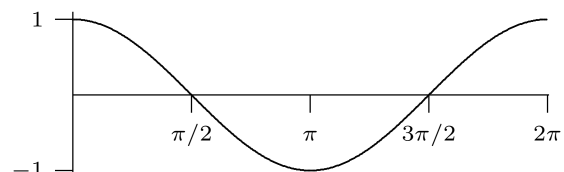
Now we can complete the calculation of the derivative of the sine:

$$\begin{aligned}\frac{d}{dx}\sin x &= \lim_{\Delta x\to 0}\frac{\sin(x+\Delta x)-\sin x}{\Delta x}\\ &= \lim_{\Delta x\to 0}\sin x\frac{\cos\Delta x-1}{\Delta x}+\cos x\frac{\sin\Delta x}{\Delta x}\\ &= \sin x\cdot 0+\cos x\cdot 1=\cos x.\end{aligned}$$

The derivative of a function measures the slope or steepness of the function; if we examine the graphs of the sine and cosine side by side, it should be that the latter appears to accurately describe the slope of the former, and indeed this is true:



sin x



cos x

Notice that where the cosine is zero the sine does appear to have a horizontal tangent line, and that the sine appears to be steepest where the cosine takes on its extreme values of 1 and −1.

Of course, now that we know the derivative of the sine, we can compute derivatives of more complicated functions involving the sine.

EXAMPLE 4.4.1 Compute the derivative of $\sin(x^2)$.

$$\frac{d}{dx}\sin(x^2)=\cos(x^2)\cdot 2x=2x\cos(x^2).$$

□

EXAMPLE 4.4.2 Compute the derivative of $\sin^2(x^3-5x)$.

$$\begin{aligned}\frac{d}{dx}\sin^2(x^3-5x) &= \frac{d}{dx}(\sin(x^3-5x))^2\\ &= 2(\sin(x^3-5x))^1\cos(x^3-5x)(3x^2-5)\\ &= 2(3x^2-5)\cos(x^3-5x)\sin(x^3-5x).\end{aligned}$$

□

Exercises 4.4.

Find the derivatives of the following functions.

1. $\sin^2(\sqrt{x})$ ⇒

2. $\sqrt{x}\sin x$ ⇒

3. $\dfrac{1}{\sin x}$ ⇒

4. $\dfrac{x^2+x}{\sin x}$ ⇒

5. $\sqrt{1-\sin^2 x}$ ⇒

4.5 Derivatives of the Trigonometric Functions

All of the other trigonometric functions can be expressed in terms of the sine, and so their derivatives can easily be calculated using the rules we already have. For the cosine we need to use two identities,

$$\cos x = \sin(x+\frac{\pi}{2}),$$
$$\sin x = -\cos(x+\frac{\pi}{2}).$$

Now:

$$\frac{d}{dx}\cos x = \frac{d}{dx}\sin(x+\frac{\pi}{2}) = \cos(x+\frac{\pi}{2})\cdot 1 = -\sin x$$
$$\frac{d}{dx}\tan x = \frac{d}{dx}\frac{\sin x}{\cos x} = \frac{\cos^2 x+\sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x$$
$$\frac{d}{dx}\sec x = \frac{d}{dx}(\cos x)^{-1} = -1(\cos x)^{-2}(-\sin x) = \frac{\sin x}{\cos^2 x} = \sec x\tan x$$

The derivatives of the cotangent and cosecant are similar and left as exercises.

Exercises 4.5.

Find the derivatives of the following functions.

1. $\sin x\cos x$ ⇒

2. $\sin(\cos x)$ ⇒

3. $\sqrt{x\tan x}$ ⇒

4. $\tan x/(1+\sin x)$ ⇒

5. $\cot x$ ⇒

6. $\csc x$ ⇒

7. $x^3\sin(23x^2)$ ⇒

8. $\sin^2 x+\cos^2 x$ ⇒

9. $\sin(\cos(6x))$ ⇒

10. Compute $\dfrac{d}{d\theta}\dfrac{\sec\theta}{1+\sec\theta}$. ⇒

11. Compute $\dfrac{d}{dt}t^5\cos(6t)$. ⇒

12. Compute $\dfrac{d}{dt}\dfrac{t^3\sin(3t)}{\cos(2t)}$. ⇒

13. Find all points on the graph of $f(x)=\sin^2(x)$ at which the tangent line is horizontal. ⇒

14. Find all points on the graph of $f(x) = 2\sin(x) - \sin^2(x)$ at which the tangent line is horizontal. ⇒

15. Find an equation for the tangent line to $\sin^2(x)$ at $x = \pi/3$. ⇒

16. Find an equation for the tangent line to $\sec^2 x$ at $x = \pi/3$. ⇒

17. Find an equation for the tangent line to $\cos^2 x - \sin^2(4x)$ at $x = \pi/6$. ⇒

18. Find the points on the curve $y = x + 2\cos x$ that have a horizontal tangent line. ⇒

19. Let C be a circle of radius r. Let A be an arc on C subtending a central angle θ. Let B be the chord of C whose endpoints are the endpoints of A. (Hence, B also subtends θ.) Let s be the length of A and let d be the length of B. Sketch a diagram of the situation and compute $\lim_{\theta\to 0^+} s/d$.

4.6 Exponential and Logarithmic functions

An exponential function has the form a^x, where a is a constant; examples are 2^x, 10^x, e^x. The logarithmic functions are the **inverses** of the exponential functions, that is, functions that "undo" the exponential functions, just as, for example, the cube root function "undoes" the cube function: $\sqrt[3]{2^3} = 2$. Note that the original function also undoes the inverse function: $(\sqrt[3]{8})^3 = 8$.

Let $f(x) = 2^x$. The inverse of this function is called the logarithm base 2, denoted $\log_2(x)$ or (especially in computer science circles) $\lg(x)$. What does this really mean? The logarithm must undo the action of the exponential function, so for example it must be that $\lg(2^3) = 3$—starting with 3, the exponential function produces $2^3 = 8$, and the logarithm of 8 must get us back to 3. A little thought shows that it is not a coincidence that $\lg(2^3)$ simply gives the exponent—the exponent *is* the original value that we must get back to. In other words, *the logarithm is the exponent.* Remember this catchphrase, and what it means, and you won't go wrong. (You *do* have to remember what it means. Like any good mnemonic, "the logarithm is the exponent" leaves out a lot of detail, like "Which exponent?" and "Exponent of what?")

EXAMPLE 4.6.1 What is the value of $\log_{10}(1000)$? The "10" tells us the appropriate number to use for the base of the exponential function. The logarithm is the exponent, so the question is, what exponent E makes $10^E = 1000$? If we can find such an E, then $\log_{10}(1000) = \log_{10}(10^E) = E$; finding the appropriate exponent is the same as finding the logarithm. In this case, of course, it is easy: $E = 3$ so $\log_{10}(1000) = 3$. □

Let's review some laws of exponents and logarithms; let a be a positive number. Since $a^5 = a\cdot a\cdot a\cdot a\cdot a$ and $a^3 = a\cdot a\cdot a$, it's clear that $a^5\cdot a^3 = a\cdot a\cdot a\cdot a\cdot a\cdot a\cdot a\cdot a = a^8 = a^{5+3}$, and in general that $a^m a^n = a^{m+n}$. Since "the logarithm is the exponent," it's no surprise that this translates directly into a fact about the logarithm function. Here are three facts from the example: $\log_a(a^5) = 5$, $\log_a(a^3) = 3$, $\log_a(a^8) = 8$. So $\log_a(a^5a^3) = \log_a(a^8) = 8 = 5 + 3 = \log_a(a^5) + \log_a(a^3)$. Now let's make this a bit more general. Suppose A and B are two numbers, $A = a^x$, and $B = a^y$. Then $\log_a(AB) = \log_a(a^x a^y) = \log_a(a^{x+y}) = x + y = \log_a(A) + \log_a(B)$.

Now consider $(a^5)^3 = a^5 \cdot a^5 \cdot a^5 = a^{5+5+5} = a^{5\cdot 3} = a^{15}$. Again it's clear that more generally $(a^m)^n = a^{mn}$, and again this gives us a fact about logarithms. If $A = a^x$ then $A^y = (a^x)^y = a^{xy}$, so $\log_a(A^y) = xy = y\log_a(A)$—the exponent can be "pulled out in front."

We have cheated a bit in the previous two paragraphs. It is obvious that $a^5 = a\cdot a\cdot a\cdot a\cdot a$ and $a^3 = a \cdot a \cdot a$ and that the rest of the example follows; likewise for the second example. But when we consider an exponential function a^x we can't be limited to substituting integers for x. What does $a^{2.5}$ or $a^{-1.3}$ or a^π mean? And is it really true that $a^{2.5}a^{-1.3} = a^{2.5-1.3}$? The answer to the first question is actually quite difficult, so we will evade it; the answer to the second question is "yes."

We'll evade the full answer to the hard question, but we have to know something about exponential functions. You need first to understand that since it's not "obvious" what 2^x should mean, we are really free to make it mean whatever we want, so long as we keep the behavior that *is* obvious, namely, when x is a positive integer. What else do we want to be true about 2^x? We want the properties of the previous two paragraphs to be true for all exponents: $2^x2^y = 2^{x+y}$ and $(2^x)^y = 2^{xy}$.

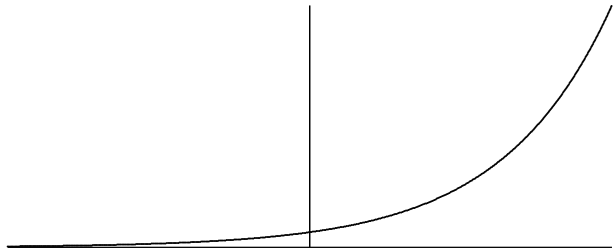
After the positive integers, the next easiest number to understand is 0: $2^0 = 1$. You have presumably learned this fact in the past; why is it true? It is true precisely because we want $2^a2^b = 2^{a+b}$ to be true about the function 2^x. We need it to be true that $2^02^x = 2^{0+x} = 2^x$, and this only works if $2^0 = 1$. The same argument implies that $a^0 = 1$ for any a.

The next easiest set of numbers to understand is the negative integers: for example, $2^{-3} = 1/2^3$. We know that whatever 2^{-3} means it must be that $2^{-3}2^3 = 2^{-3+3} = 2^0 = 1$, which means that 2^{-3} must be $1/2^3$. In fact, by the same argument, once we know what 2^x means for some value of x, 2^{-x} must be $1/2^x$ and more generally $a^{-x} = 1/a^x$.

Next, consider an exponent $1/q$, where q is a positive integer. We want it to be true that $(2^x)^y = 2^{xy}$, so $(2^{1/q})^q = 2$. This means that $2^{1/q}$ is a q-th root of 2, $2^{1/q} = \sqrt[q]{2}$. This is all we need to understand that $2^{p/q} = (2^{1/q})^p = (\sqrt[q]{2})^p$ and $a^{p/q} = (a^{1/q})^p = (\sqrt[q]{a})^p$.

What's left is the hard part: what does 2^x mean when x cannot be written as a fraction, like $x = \sqrt{2}$ or $x = \pi$? What we know so far is how to assign meaning to 2^x

whenever $x = p/q$; if we were to graph this we'd see something like this:



But this is a poor picture, because you can't see that the "curve" is really a whole lot of individual points, above the rational numbers on the x-axis. There are really a lot of "holes" in the curve, above $x = \pi$, for example. But (this is the hard part) it is possible to prove that the holes can be "filled in", and that the resulting function, called 2^x, really does have the properties we want, namely that $2^x 2^y = 2^{x+y}$ and $(2^x)^y = 2^{xy}$.

Exercises 4.6.

1. Expand $\log_{10}((x+45)^7(x-2))$.

2. Expand $\log_2 \dfrac{x^3}{3x-5+(7/x)}$.

3. Write $\log_2 3x + 17\log_2(x-2) - 2\log_2(x^2+4x+1)$ as a single logarithm.

4. Solve $\log_2(1+\sqrt{x}) = 6$ for x.

5. Solve $2^{x^2} = 8$ for x.

6. Solve $\log_2(\log_3(x)) = 1$ for x.

4.7 Derivatives of the exponential and logarithmic functions

As with the sine, we don't know anything about derivatives that allows us to compute the derivatives of the exponential and logarithmic functions without going back to basics. Let's do a little work with the definition again:

$$\begin{aligned}
\frac{d}{dx}a^x &= \lim_{\Delta x\to 0}\frac{a^{x+\Delta x}-a^x}{\Delta x}\\
&= \lim_{\Delta x\to 0}\frac{a^x a^{\Delta x}-a^x}{\Delta x}\\
&= \lim_{\Delta x\to 0} a^x\frac{a^{\Delta x}-1}{\Delta x}\\
&= a^x \lim_{\Delta x\to 0}\frac{a^{\Delta x}-1}{\Delta x}
\end{aligned}$$

There are two interesting things to note here: As in the case of the sine function we are left with a limit that involves Δx but not x, which means that whatever $\lim_{\Delta x \to 0}(a^{\Delta x} - 1)/\Delta x$ is, we know that it is a number, that is, a constant. This means that a^x has a remarkable property: its derivative is a constant times itself.

We earlier remarked that the hardest limit we would compute is $\lim_{x \to 0} \sin x/x = 1$; we now have a limit that is just a bit too hard to include here. In fact the hard part is to see that $\lim_{\Delta x \to 0}(a^{\Delta x} - 1)/\Delta x$ even exists—does this fraction really get closer and closer to some fixed value? Yes it does, but we will not prove this fact.

We can look at some examples. Consider $(2^x - 1)/x$ for some small values of x: 1, 0.828427124, 0.756828460, 0.724061864, 0.70838051, 0.70070877 when x is 1, 1/2, 1/4, 1/8, 1/16, 1/32, respectively. It looks like this is settling in around 0.7, which turns out to be true (but the limit is not exactly 0.7). Consider next $(3^x - 1)/x$: 2, 1.464101616, 1.264296052, 1.177621520, 1.13720773, 1.11768854, at the same values of x. It turns out to be true that in the limit this is about 1.1. Two examples don't establish a pattern, but if you do more examples you will find that the limit varies directly with the value of a: bigger a, bigger limit; smaller a, smaller limit. As we can already see, some of these limits will be less than 1 and some larger than 1. Somewhere between $a = 2$ and $a = 3$ the limit will be exactly 1; the value at which this happens is called e, so that

$$\lim_{\Delta x \to 0} \frac{e^{\Delta x} - 1}{\Delta x} = 1.$$

As you might guess from our two examples, e is closer to 3 than to 2, and in fact $e \approx 2.718$.

Now we see that the function e^x has a truly remarkable property:

$$\begin{aligned}
\frac{d}{dx} e^x &= \lim_{\Delta x \to 0} \frac{e^{x+\Delta x} - e^x}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{e^x e^{\Delta x} - e^x}{\Delta x} \\
&= \lim_{\Delta x \to 0} e^x \frac{e^{\Delta x} - 1}{\Delta x} \\
&= e^x \lim_{\Delta x \to 0} \frac{e^{\Delta x} - 1}{\Delta x} \\
&= e^x
\end{aligned}$$

That is, e^x is its own derivative, or in other words the slope of e^x is the same as its height, or the same as its second coordinate: The function $f(x) = e^x$ goes through the point (z, e^z) and has slope e^z there, no matter what z is. It is sometimes convenient to express the function e^x without an exponent, since complicated exponents can be hard to read. In such cases we use $\exp(x)$, e.g., $\exp(1 + x^2)$ instead of e^{1+x^2}.

What about the logarithm function? This too is hard, but as the cosine function was easier to do once the sine was done, so the logarithm is easier to do now that we know the derivative of the exponential function. Let's start with $\log_e x$, which as you probably know is often abbreviated $\ln x$ and called the "natural logarithm" function.

Consider the relationship between the two functions, namely, that they are inverses, that one "undoes" the other. Graphically this means that they have the same graph except that one is "flipped" or "reflected" through the line $y = x$, as shown in figure 4.7.1.

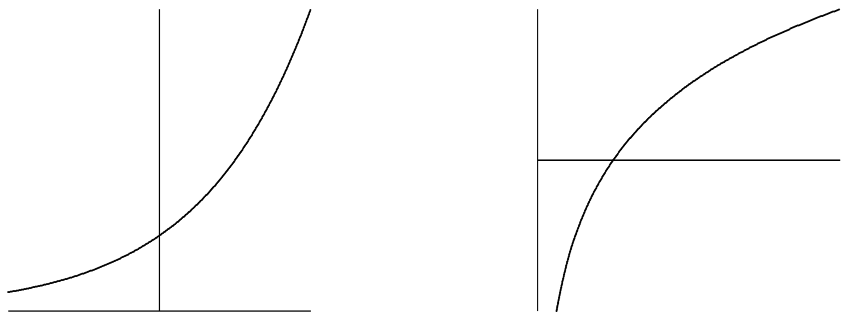


Figure 4.7.1 The exponential and logarithm functions.

This means that the slopes of these two functions are closely related as well: For example, the slope of e^x is e at $x = 1$; at the corresponding point on the $\ln(x)$ curve, the slope must be $1/e$, because the "rise" and the "run" have been interchanged. Since the slope of e^x is e at the point $(1, e)$, the slope of $\ln(x)$ is $1/e$ at the point $(e, 1)$.

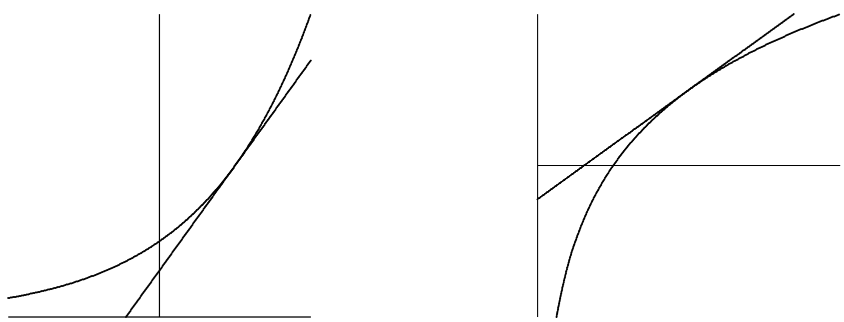


Figure 4.7.2 Slope of the exponential and logarithm functions.

More generally, we know that the slope of e^x is e^z at the point (z, e^z), so the slope of $\ln(x)$ is $1/e^z$ at (e^z, z), as indicated in figure 4.7.2. In other words, the slope of $\ln x$ is the reciprocal of the first coordinate at any point; this means that the slope of $\ln x$ at $(x, \ln x)$ is $1/x$. The upshot is:

$$\frac{d}{dx} \ln x = \frac{1}{x}.$$

We have discussed this from the point of view of the graphs, which is easy to understand but is not normally considered a rigorous proof—it is too easy to be led astray by pictures that seem reasonable but that miss some hard point. It is possible to do this derivation without resorting to pictures, and indeed we will see an alternate approach soon.

Note that $\ln x$ is defined only for $x > 0$. It is sometimes useful to consider the function $\ln |x|$, a function defined for $x \neq 0$. When $x < 0$, $\ln |x| = \ln(-x)$ and

$$\frac{d}{dx} \ln |x| = \frac{d}{dx} \ln(-x) = \frac{1}{-x}(-1) = \frac{1}{x}.$$

Thus whether x is positive or negative, the derivative is the same.

What about the functions a^x and $\log_a x$? We know that the derivative of a^x is some constant times a^x itself, but what constant? Remember that "the logarithm is the exponent" and you will see that $a = e^{\ln a}$. Then

$$a^x = (e^{\ln a})^x = e^{x \ln a},$$

and we can compute the derivative using the chain rule:

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

The constant is simply $\ln a$. Likewise we can compute the derivative of the logarithm function $\log_a x$. Since

$$x = e^{\ln x}$$

we can take the logarithm base a of both sides to get

$$\log_a(x) = \log_a(e^{\ln x}) = \ln x \log_a e.$$

Then

$$\frac{d}{dx} \log_a x = \frac{1}{x} \log_a e.$$

This is a perfectly good answer, but we can improve it slightly. Since

$$\begin{aligned} a &= e^{\ln a} \\ \log_a(a) &= \log_a(e^{\ln a}) = \ln a \log_a e \\ 1 &= \ln a \log_a e \\ \frac{1}{\ln a} &= \log_a e, \end{aligned}$$

we can replace $\log_a e$ to get

$$\frac{d}{dx}\log_a x = \frac{1}{x\ln a}.$$

You may if you wish memorize the formulas

$$\frac{d}{dx}a^x = (\ln a)a^x \quad \text{and} \quad \frac{d}{dx}\log_a x = \frac{1}{x\ln a}.$$

Because the "trick" $a = e^{\ln a}$ is often useful, and sometimes essential, it may be better to remember the trick, not the formula.

EXAMPLE 4.7.1 Compute the derivative of $f(x) = 2^x$.

$$\begin{aligned}
\frac{d}{dx}2^x &= \frac{d}{dx}(e^{\ln 2})^x \\
&= \frac{d}{dx}e^{x\ln 2} \\
&= \left(\frac{d}{dx}x\ln 2\right)e^{x\ln 2} \\
&= (\ln 2)e^{x\ln 2} = 2^x \ln 2
\end{aligned}$$

□

EXAMPLE 4.7.2 Compute the derivative of $f(x) = 2^{x^2} = 2^{(x^2)}$.

$$\begin{aligned}
\frac{d}{dx}2^{x^2} &= \frac{d}{dx}e^{x^2\ln 2} \\
&= \left(\frac{d}{dx}x^2\ln 2\right)e^{x^2\ln 2} \\
&= (2\ln 2)xe^{x^2\ln 2} \\
&= (2\ln 2)x2^{x^2}
\end{aligned}$$

□

EXAMPLE 4.7.3 Compute the derivative of $f(x) = x^x$. At first this appears to be a new kind of function: it is not a constant power of x, and it does not seem to be an exponential function, since the base is not constant. But in fact it is no harder than the previous example.

$$\begin{aligned}
\frac{d}{dx}x^x &= \frac{d}{dx}e^{x\ln x} \\
&= \left(\frac{d}{dx}x\ln x\right)e^{x\ln x} \\
&= (x\frac{1}{x} + \ln x)x^x \\
&= (1 + \ln x)x^x
\end{aligned}$$

□

EXAMPLE 4.7.4 Recall that we have not justified the power rule except when the exponent is a positive or negative integer. We can use the exponential function to take care of other exponents.

$$\begin{aligned}\frac{d}{dx}x^r &= \frac{d}{dx}e^{r\ln x}\\ &= \left(\frac{d}{dx}r\ln x\right)e^{r\ln x}\\ &= (r\frac{1}{x})x^r\\ &= rx^{r-1}\end{aligned}$$

□

Exercises 4.7.

In 1–19, find the derivatives of the functions.

1. 3^{x^2} ⇒

2. $\dfrac{\sin x}{e^x}$ ⇒

3. $(e^x)^2$ ⇒

4. $\sin(e^x)$ ⇒

5. $e^{\sin x}$ ⇒

6. $x^{\sin x}$ ⇒

7. x^3e^x ⇒

8. $x + 2^x$ ⇒

9. $(1/3)^{x^2}$ ⇒

10. e^{4x}/x ⇒

11. $\ln(x^3 + 3x)$ ⇒

12. $\ln(\cos(x))$ ⇒

13. $\sqrt{\ln(x^2)}/x$ ⇒

14. $\ln(\sec(x) + \tan(x))$ ⇒

15. $x^{\cos(x)}$ ⇒

16. $x\ln x$

17. $\ln(\ln(3x))$

18. $\dfrac{1 + \ln(3x^2)}{1 + \ln(4x)}$

19. $\dfrac{x^8(x-23)^{1/2}}{27x^6(4x-6)^8}$

20. Find the value of a so that the tangent line to $y = \ln(x)$ at $x = a$ is a line through the origin. Sketch the resulting situation. ⇒

21. If $f(x) = \ln(x^3 + 2)$ compute $f'(e^{1/3})$.

4.8 Implicit Differentiation

As we have seen, there is a close relationship between the derivatives of e^x and $\ln x$ because these functions are inverses. Rather than relying on pictures for our understanding, we would like to be able to exploit this relationship computationally. In fact this technique can help us find derivatives in many situations, not just when we seek the derivative of an inverse function.

We will begin by illustrating the technique to find what we already know, the derivative of $\ln x$. Let's write $y = \ln x$ and then $x = e^{\ln x} = e^y$, that is, $x = e^y$. We say that this equation defines the function $y = \ln x$ implicitly because while it is not an explicit expression $y = \ldots$, it is true that if $x = e^y$ then y is in fact the natural logarithm function. Now, for the time being, pretend that all we know of y is that $x = e^y$; what can we say about derivatives? We can take the derivative of both sides of the equation:

$$\frac{d}{dx}x = \frac{d}{dx}e^y.$$

Then using the chain rule on the right hand side:

$$1 = \left(\frac{d}{dx}y\right)e^y = y'e^y.$$

Then we can solve for y':

$$y' = \frac{1}{e^y} = \frac{1}{x}.$$

There is one little difficulty here. To use the chain rule to compute $d/dx(e^y) = y'e^y$ we need to know that the function y *has* a derivative. All we have shown is that *if* it has a derivative then that derivative must be $1/x$. When using this method we will always have to assume that the desired derivative exists, but fortunately this is a safe assumption for most such problems.

The example $y = \ln x$ involved an inverse function defined implicitly, but other functions can be defined implicitly, and sometimes a single equation can be used to implicitly define more than one function. Here's a familiar example. The equation $r^2 = x^2 + y^2$ describes a circle of radius r. The circle is not a function $y = f(x)$ because for some values of x there are two corresponding values of y. If we want to work with a function, we can break the circle into two pieces, the upper and lower semicircles, each of which is a function. Let's call these $y = U(x)$ and $y = L(x)$; in fact this is a fairly simple example, and it's possible to give explicit expressions for these: $U(x) = \sqrt{r^2 - x^2}$ and $L(x) = -\sqrt{r^2 - x^2}$. But it's somewhat easier, and quite useful, to view both functions as given implicitly by $r^2 = x^2 + y^2$: both $r^2 = x^2 + U(x)^2$ and $r^2 = x^2 + L(x)^2$ are true, and we can think of $r^2 = x^2 + y^2$ as defining both $U(x)$ and $L(x)$.

Now we can take the derivative of both sides as before, remembering that y is not simply a variable but a function—in this case, y is either $U(x)$ or $L(x)$ but we're not yet specifying which one. When we take the derivative we just have to remember to apply the chain rule where y appears.

$$\begin{aligned}\frac{d}{dx}r^2 &= \frac{d}{dx}(x^2+y^2)\\ 0 &= 2x + 2yy'\\ y' &= \frac{-2x}{2y} = -\frac{x}{y}\end{aligned}$$

Now we have an expression for y', but it contains y as well as x. This means that if we want to compute y' for some particular value of x we'll have to know or compute y at that value of x as well. It is at this point that we will need to know whether y is $U(x)$ or $L(x)$. Occasionally it will turn out that we can avoid explicit use of $U(x)$ or $L(x)$ by the nature of the problem

EXAMPLE 4.8.1 Find the slope of the circle $4 = x^2 + y^2$ at the point $(1, -\sqrt{3})$. Since we know both the x and y coordinates of the point of interest, we do not need to explicitly recognize that this point is on $L(x)$, and we do not need to use $L(x)$ to compute y—but we could. Using the calculation of y' from above,

$$y' = -\frac{x}{y} = -\frac{1}{-\sqrt{3}} = \frac{1}{\sqrt{3}}.$$

It is instructive to compare this approach to others.

We might have recognized at the start that $(1, -\sqrt{3})$ is on the function $y = L(x) = -\sqrt{4-x^2}$. We could then take the derivative of $L(x)$, using the power rule and the chain rule, to get

$$L'(x) = -\frac{1}{2}(4-x^2)^{-1/2}(-2x) = \frac{x}{\sqrt{4-x^2}}.$$

Then we could compute $L'(1) = 1/\sqrt{3}$ by substituting $x = 1$.

Alternately, we could realize that the point is on $L(x)$, but use the fact that $y' = -x/y$. Since the point is on $L(x)$ we can replace y by $L(x)$ to get

$$y' = -\frac{x}{L(x)} = \frac{x}{\sqrt{4-x^2}},$$

without computing the derivative of $L(x)$ explicitly. Then we substitute $x = 1$ and get the same answer as before. □

In the case of the circle it is possible to find the functions $U(x)$ and $L(x)$ explicitly, but there are potential advantages to using implicit differentiation anyway. In some cases it is more difficult or impossible to find an explicit formula for y and implicit differentiation is the only way to find the derivative.

EXAMPLE 4.8.2 Find the derivative of any function defined implicitly by $yx^2+e^y = x$. We treat y as an unspecified function and use the chain rule:

$$\begin{aligned}
\frac{d}{dx}(yx^2 + e^y) &= \frac{d}{dx}x \\
(y \cdot 2x + y' \cdot x^2) + y'e^y &= 1 \\
y'x^2 + y'e^y &= 1 - 2xy \\
y'(x^2 + e^y) &= 1 - 2xy \\
y' &= \frac{1 - 2xy}{x^2 + e^y}
\end{aligned}$$

□

You might think that the step in which we solve for y' could sometimes be difficult—after all, we're using implicit differentiation here because we can't solve the equation $yx^2 + e^y = x$ for y, so maybe after taking the derivative we get something that is hard to solve for y'. In fact, *this never happens.* All occurrences y' come from applying the chain rule, and whenever the chain rule is used it deposits a single y' multiplied by some other expression. So it will always be possible to group the terms containing y' together and factor out the y', just as in the previous example. If you ever get anything more difficult you have made a mistake and should fix it before trying to continue.

It is sometimes the case that a situation leads naturally to an equation that defines a function implicitly.

EXAMPLE 4.8.3 Consider all the points (x, y) that have the property that the distance from (x, y) to (x_1, y_1) plus the distance from (x, y) to (x_2, y_2) is $2a$ (a is some constant). These points form an ellipse, which like a circle is not a function but can viewed as two functions pasted together. Because we know how to write down the distance between two points, we can write down an implicit equation for the ellipse:

$$\sqrt{(x - x_1)^2 + (y - y_1)^2} + \sqrt{(x - x_2)^2 + (y - y_2)^2} = 2a.$$

Then we can use implicit differentiation to find the slope of the ellipse at any point, though the computation is rather messy. □

EXAMPLE 4.8.4 We have already justified the power rule by using the exponential function, but we could also do it for rational exponents by using implicit differentiation. Suppose that $y = x^{m/n}$, where m and n are positive integers. We can write this implicitly as $y^n = x^m$, then because we justified the power rule for integers, we can take the derivative of each side:

$$\begin{aligned} ny^{n-1}y' &= mx^{m-1} \\ y' &= \frac{m}{n}\frac{x^{m-1}}{y^{n-1}} \\ y' &= \frac{m}{n}\frac{x^{m-1}}{(x^{m/n})^{n-1}} \\ y' &= \frac{m}{n}x^{m-1-(m/n)(n-1)} \\ y' &= \frac{m}{n}x^{m-1-m+(m/n)} \\ y' &= \frac{m}{n}x^{(m/n)-1} \end{aligned}$$

□

Exercises 4.8.

In exercises 1–8, find a formula for the derivative y' at the point (x, y):

1. $y^2 = 1 + x^2$ ⇒
2. $x^2 + xy + y^2 = 7$ ⇒
3. $x^3 + xy^2 = y^3 + yx^2$ ⇒
4. $4\cos x \sin y = 1$ ⇒
5. $\sqrt{x} + \sqrt{y} = 9$ ⇒
6. $\tan(x/y) = x + y$ ⇒
7. $\sin(x + y) = xy$ ⇒
8. $\dfrac{1}{x} + \dfrac{1}{y} = 7$ ⇒
9. A hyperbola passing through $(8, 6)$ consists of all points whose distance from the origin is a constant more than its distance from the point (5,2). Find the slope of the tangent line to the hyperbola at $(8, 6)$. ⇒
10. Compute y' for the ellipse of example 4.8.3.
11. If $y = \log_a x$ then $a^y = x$. Use implicit differentiation to find y'.
12. The graph of the equation $x^2 - xy + y^2 = 9$ is an ellipse. Find the lines tangent to this curve at the two points where it intersects the x-axis. Show that these lines are parallel. ⇒
13. Repeat the previous problem for the points at which the ellipse intersects the y-axis. ⇒
14. Find the points on the ellipse from the previous two problems where the slope is horizontal and where it is vertical. ⇒
15. Find an equation for the tangent line to $x^4 = y^2 + x^2$ at $(2, \sqrt{12})$. (This curve is the **kampyle of Eudoxus**.) ⇒
16. Find an equation for the tangent line to $x^{2/3} + y^{2/3} = a^{2/3}$ at a point (x_1, y_1) on the curve, with $x_1 \neq 0$ and $y_1 \neq 0$. (This curve is an **astroid**.) ⇒
17. Find an equation for the tangent line to $(x^2 + y^2)^2 = x^2 - y^2$ at a point (x_1, y_1) on the curve, with $x_1 \neq 0, -1, 1$. (This curve is a **lemniscate**.) ⇒

Definition. Two curves are **orthogonal** if at each point of intersection, the angle between their tangent lines is $\pi/2$. Two families of curves, $\mathcal{A}$ and $\mathcal{B}$, are **orthogonal trajectories** of each other if given any curve C in $\mathcal{A}$ and any curve D in $\mathcal{B}$ the curves C and D are orthogonal. For example, the family of horizontal lines in the plane is orthogonal to the family of vertical lines in the plane.

18. Show that $x^2 - y^2 = 5$ is orthogonal to $4x^2 + 9y^2 = 72$. (Hint: You need to find the intersection points of the two curves and then show that the product of the derivatives at each intersection point is -1.)

19. Show that $x^2 + y^2 = r^2$ is orthogonal to $y = mx$. Conclude that the family of circles centered at the origin is an orthogonal trajectory of the family of lines that pass through the origin.

Note that there is a technical issue when $m = 0$. The circles fail to be differentiable when they cross the x-axis. However, the circles are orthogonal to the x-axis. Explain why. Likewise, the vertical line through the origin requires a separate argument.

20. For $k \neq 0$ and $c \neq 0$ show that $y^2 - x^2 = k$ is orthogonal to $yx = c$. In the case where k and c are both zero, the curves intersect at the origin. Are the curves $y^2 - x^2 = 0$ and $yx = 0$ orthogonal to each other?

21. Suppose that $m \neq 0$. Show that the family of curves $\{y = mx + b \mid b \in \mathbb{R}\}$ is orthogonal to the family of curves $\{y = -(x/m) + c \mid c \in \mathbb{R}\}$.

4.9 Inverse Trigonometric Functions

The trigonometric functions frequently arise in problems, and often it is necessary to invert the functions, for example, to find an angle with a specified sine. Of course, there are many angles with the same sine, so the sine function doesn't actually have an inverse that reliably "undoes" the sine function. If you know that $\sin x = 0.5$, you can't reverse this to discover x, that is, you can't solve for x, as there are infinitely many angles with sine 0.5. Nevertheless, it is useful to have something like an inverse to the sine, however imperfect. The usual approach is to pick out some collection of angles that produce all possible values of the sine exactly once. If we "discard" all other angles, the resulting function does have a proper inverse.

The sine takes on all values between -1 and 1 exactly once on the interval $[-\pi/2, \pi/2]$. If we truncate the sine, keeping only the interval $[-\pi/2, \pi/2]$, as shown in figure 4.9.1, then this truncated sine has an inverse function. We call this the inverse sine or the arcsine, and write $y = \arcsin(x)$.

Recall that a function and its inverse undo each other in either order, for example, $(\sqrt[3]{x})^3 = x$ and $\sqrt[3]{x^3} = x$. This does not work with the sine and the "inverse sine" because the inverse sine is the inverse of the truncated sine function, not the real sine function. It is true that $\sin(\arcsin(x)) = x$, that is, the sine undoes the arcsine. It is not true that the arcsine undoes the sine, for example, $\sin(5\pi/6) = 1/2$ and $\arcsin(1/2) = \pi/6$, so doing first the sine then the arcsine does not get us back where we started. This is because $5\pi/6$

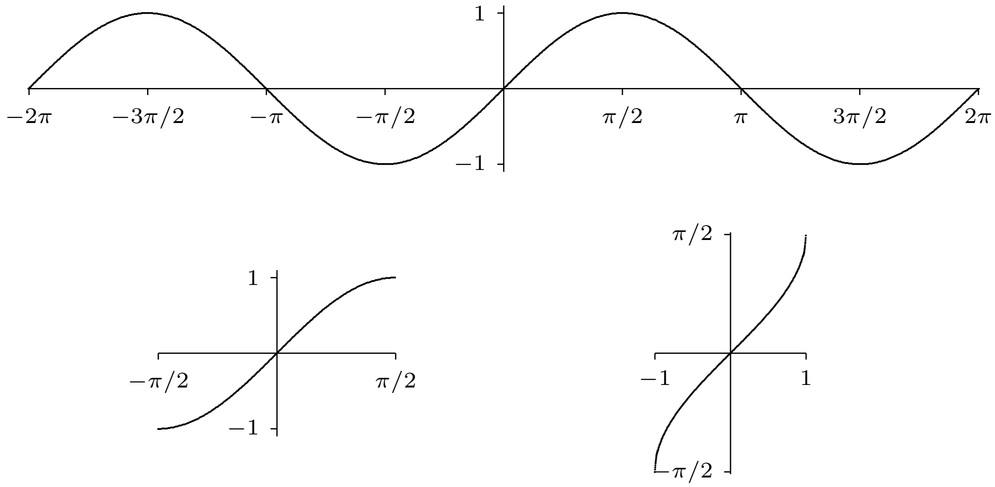


Figure 4.9.1 The sine, the truncated sine, the inverse sine.

is not in the domain of the truncated sine. If we start with an angle between $-\pi/2$ and $\pi/2$ then the arcsine does reverse the sine: $\sin(\pi/6) = 1/2$ and $\arcsin(1/2) = \pi/6$.

What is the derivative of the arcsine? Since this is an inverse function, we can discover the derivative by using implicit differentiation. Suppose $y = \arcsin(x)$. Then

$$\sin(y) = \sin(\arcsin(x)) = x.$$

Now taking the derivative of both sides, we get

$$\begin{aligned} y' \cos y &= 1 \\ y' &= \frac{1}{\cos y} \end{aligned}$$

As we expect when using implicit differentiation, y appears on the right hand side here. We would certainly prefer to have y' written in terms of x, and as in the case of $\ln x$ we can actually do that here. Since $\sin^2 y + \cos^2 y = 1$, $\cos^2 y = 1 - \sin^2 y = 1 - x^2$. So $\cos y = \pm\sqrt{1 - x^2}$, but which is it—plus or minus? It could in general be either, but this isn't "in general": since $y = \arcsin(x)$ we know that $-\pi/2 \le y \le \pi/2$, and the cosine of an angle in this interval is always positive. Thus $\cos y = \sqrt{1 - x^2}$ and

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

Note that this agrees with figure 4.9.1: the graph of the arcsine has positive slope everywhere.

We can do something similar for the cosine. As with the sine, we must first truncate the cosine so that it can be inverted, as shown in figure 4.9.2. Then we use implicit

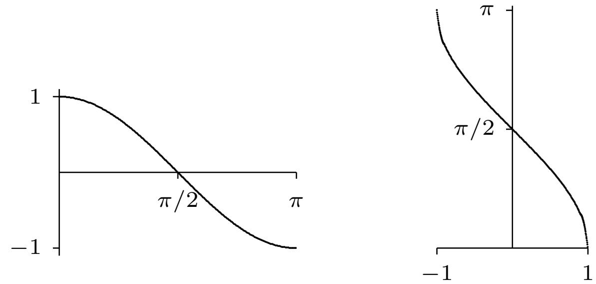


Figure 4.9.2 The truncated cosine, the inverse cosine.

differentiation to find that

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

Note that the truncated cosine uses a different interval than the truncated sine, so that if $y = \arccos(x)$ we know that $0 \le y \le \pi$. The computation of the derivative of the arccosine is left as an exercise.

Finally we look at the tangent; the other trigonometric functions also have "partial inverses" but the sine, cosine and tangent are enough for most purposes. The tangent, truncated tangent and inverse tangent are shown in figure 4.9.3; the derivative of the arctangent is left as an exercise.

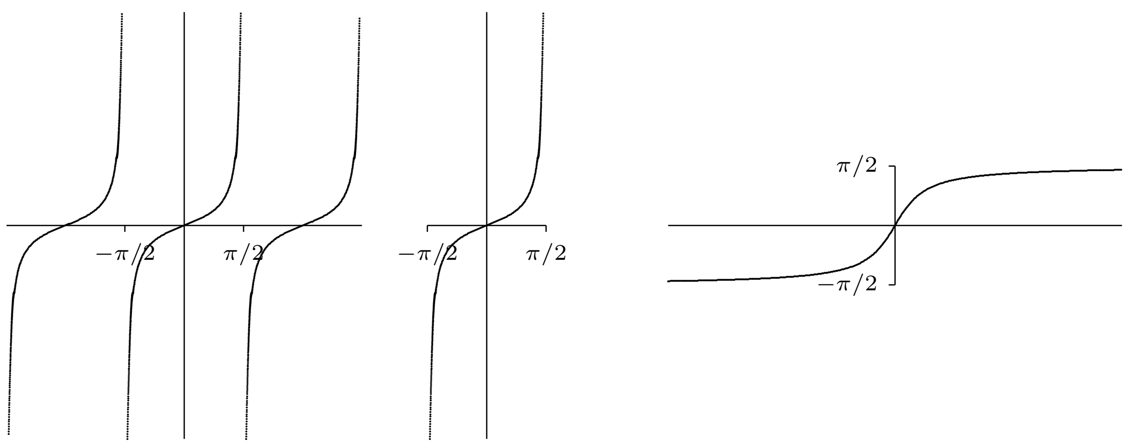


Figure 4.9.3 The tangent, the truncated tangent, the inverse tangent.

Exercises 4.9.

1. Show that the derivative of $\arccos x$ is $-\dfrac{1}{\sqrt{1-x^2}}$.

2. Show that the derivative of $\arctan x$ is $\dfrac{1}{1+x^2}$.

3. The inverse of cot is usually defined so that the range of arccot is $(0, \pi)$. Sketch the graph of $y = \operatorname{arccot} x$. In the process you will make it clear what the domain of arccot is. Find the derivative of the arccotangent. ⇒
4. Show that $\operatorname{arccot} x + \arctan x = \pi/2$.
5. Find the derivative of $\arcsin(x^2)$. ⇒
6. Find the derivative of $\arctan(e^x)$. ⇒
7. Find the derivative of $\arccos(\sin x^3)$ ⇒
8. Find the derivative of $\ln((\arcsin x)^2)$ ⇒
9. Find the derivative of $\arccos e^x$ ⇒
10. Find the derivative of $\arcsin x + \arccos x$ ⇒
11. Find the derivative of $\log_5(\arctan(x^x))$ ⇒

4.10 Limits revisited

We have defined and used the concept of limit, primarily in our development of the derivative. Recall that $\lim_{x\to a} f(x) = L$ is true if, in a precise sense, $f(x)$ gets closer and closer to L as x gets closer and closer to a. While some limits are easy to see, others take some ingenuity; in particular, the limits that define derivatives are always difficult on their face, since in

$$\lim_{\Delta x\to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

both the numerator and denominator approach zero. Typically this difficulty can be resolved when f is a “nice” function and we are trying to compute a derivative. Occasionally such limits are interesting for other reasons, and the limit of a fraction in which both numerator and denominator approach zero can be difficult to analyze. Now that we have the derivative available, there is another technique that can sometimes be helpful in such circumstances.

Before we introduce the technique, we will also expand our concept of limit, in two ways. When the limit of $f(x)$ as x approaches a does not exist, it may be useful to note in what way it does not exist. We have already talked about one such case: one-sided limits. Another case is when “f goes to infinity”. We also will occasionally want to know what happens to f when x “goes to infinity”.

EXAMPLE 4.10.1 What happens to $1/x$ as x goes to 0? From the right, $1/x$ gets bigger and bigger, or goes to infinity. From the left it goes to negative infinity. □

EXAMPLE 4.10.2 What happens to the function $\cos(1/x)$ as x goes to infinity? It seems clear that as x gets larger and larger, $1/x$ gets closer and closer to zero, so $\cos(1/x)$ should be getting closer and closer to $\cos(0) = 1$. □

As with ordinary limits, these concepts can be made precise. Roughly, we want $\lim_{x\to a} f(x) = \infty$ to mean that we can make $f(x)$ arbitrarily large by making x close enough to a, and $\lim_{x\to\infty} f(x) = L$ should mean we can make $f(x)$ as close as we want to L by making x large enough. Compare this definition to the definition of limit in section 2.3, definition 2.3.2.

DEFINITION 4.10.3 If f is a function, we say that $\lim_{x\to a} f(x) = \infty$ if for every $N > 0$ there is a $\delta > 0$ such that whenever $|x - a| < \delta$, $f(x) > N$. We can extend this in the obvious ways to define $\lim_{x\to a} f(x) = -\infty$, $\lim_{x\to a^-} f(x) = \pm\infty$, and $\lim_{x\to a^+} f(x) = \pm\infty$. □

DEFINITION 4.10.4 Limit at infinity If f is a function, we say that $\lim_{x\to\infty} f(x) = L$ if for every $\epsilon > 0$ there is an $N > 0$ so that whenever $x > N$, $|f(x) - L| < \epsilon$. We may similarly define $\lim_{x\to-\infty} f(x) = L$, and using the idea of the previous definition, we may define $\lim_{x\to\pm\infty} f(x) = \pm\infty$. □

We include these definitions for completeness, but we will not explore them in detail. Suffice it to say that such limits behave in much the same way that ordinary limits do; in particular there are some analogs of theorem 2.3.6.

Now consider this limit:

$$\lim_{x\to\pi} \frac{x^2 - \pi^2}{\sin x}.$$

As x approaches π, both the numerator and denominator approach zero, so it is not obvious what, if anything, the quotient approaches. We can often compute such limits by application of the following theorem.

THEOREM 4.10.5 L'Hôpital's Rule For "sufficiently nice" functions $f(x)$ and $g(x)$, if $\lim_{x\to a} f(x) = 0 = \lim_{x\to a} g(x)$ or both $\lim_{x\to a} f(x) = \pm\infty$ and $\lim_{x\to a} g(x) = \pm\infty$, and if $\lim_{x\to a} \frac{f'(x)}{g'(x)}$ exists, then $\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$. This remains true if "$x \to a$" is replaced by "$x \to \infty$" or "$x \to -\infty$". ■

This theorem is somewhat difficult to prove, in part because it incorporates so many different possibilities, so we will not prove it here. We also will not need to worry about the precise definition of "sufficiently nice", as the functions we encounter will be suitable.

EXAMPLE 4.10.6 Compute $\lim_{x\to\pi} \frac{x^2 - \pi^2}{\sin x}$ in two ways.

First we use L'Hôpital's Rule: Since the numerator and denominator both approach zero,

$$\lim_{x\to\pi} \frac{x^2-\pi^2}{\sin x} = \lim_{x\to\pi} \frac{2x}{\cos x},$$

provided the latter exists. But in fact this is an easy limit, since the denominator now approaches -1, so

$$\lim_{x\to\pi} \frac{x^2-\pi^2}{\sin x} = \frac{2\pi}{-1} = -2\pi.$$

We don't really need L'Hôpital's Rule to do this limit. Rewrite it as

$$\lim_{x\to\pi} (x+\pi)\frac{x-\pi}{\sin x}$$

and note that

$$\lim_{x\to\pi} \frac{x-\pi}{\sin x} = \lim_{x\to\pi} \frac{x-\pi}{-\sin(x-\pi)} = \lim_{x\to 0} -\frac{x}{\sin x}$$

since $x-\pi$ approaches zero as x approaches π. Now

$$\lim_{x\to\pi} (x+\pi)\frac{x-\pi}{\sin x} = \lim_{x\to\pi} (x+\pi) \lim_{x\to 0} -\frac{x}{\sin x} = 2\pi(-1) = -2\pi$$

as before. □

EXAMPLE 4.10.7 Compute $\displaystyle\lim_{x\to\infty} \frac{2x^2-3x+7}{x^2+47x+1}$ in two ways.

As x goes to infinity both the numerator and denominator go to infinity, so we may apply L'Hôpital's Rule:

$$\lim_{x\to\infty} \frac{2x^2-3x+7}{x^2+47x+1} = \lim_{x\to\infty} \frac{4x-3}{2x+47}.$$

In the second quotient, it is still the case that the numerator and denominator both go to infinity, so we are allowed to use L'Hôpital's Rule again:

$$\lim_{x\to\infty} \frac{4x-3}{2x+47} = \lim_{x\to\infty} \frac{4}{2} = 2.$$

So the original limit is 2 as well.

Again, we don't really need L'Hôpital's Rule, and in fact a more elementary approach is easier—we divide the numerator and denominator by x^2:

$$\lim_{x\to\infty} \frac{2x^2-3x+7}{x^2+47x+1} = \lim_{x\to\infty} \frac{2x^2-3x+7}{x^2+47x+1}\frac{\frac{1}{x^2}}{\frac{1}{x^2}} = \lim_{x\to\infty} \frac{2-\frac{3}{x}+\frac{7}{x^2}}{1+\frac{47}{x}+\frac{1}{x^2}}.$$

Now as x approaches infinity, all the quotients with some power of x in the denominator approach zero, leaving 2 in the numerator and 1 in the denominator, so the limit again is 2. □

EXAMPLE 4.10.8 Compute $\lim_{x\to 0}\frac{\sec x - 1}{\sin x}$.

Both the numerator and denominator approach zero, so applying L'Hôpital's Rule:

$$\lim_{x\to 0}\frac{\sec x - 1}{\sin x} = \lim_{x\to 0}\frac{\sec x \tan x}{\cos x} = \frac{1\cdot 0}{1} = 0.$$

□

EXAMPLE 4.10.9 Compute $\lim_{x\to 0^+} x\ln x$.

This doesn't appear to be suitable for L'Hôpital's Rule, but it also is not "obvious". As x approaches zero, $\ln x$ goes to $-\infty$, so the product looks like (something very small) $\cdot$ (something very large and negative). But this could be anything: it depends on *how small* and *how large*. For example, consider $(x^2)(1/x)$, $(x)(1/x)$, and $(x)(1/x^2)$. As x approaches zero, each of these is (something very small) $\cdot$ (something very large), yet the limits are respectively zero, 1, and ∞.

We can in fact turn this into a L'Hôpital's Rule problem:

$$x\ln x = \frac{\ln x}{1/x} = \frac{\ln x}{x^{-1}}.$$

Now as x approaches zero, both the numerator and denominator approach infinity (one $-\infty$ and one $+\infty$, but only the size is important). Using L'Hôpital's Rule:

$$\lim_{x\to 0^+}\frac{\ln x}{x^{-1}} = \lim_{x\to 0^+}\frac{1/x}{-x^{-2}} = \lim_{x\to 0^+}\frac{1}{x}(-x^2) = \lim_{x\to 0^+} -x = 0.$$

One way to interpret this is that since $\lim_{x\to 0^+} x\ln x = 0$, the x approaches zero much faster than the $\ln x$ approaches $-\infty$.

□

Exercises 4.10.

Compute the limits.

1. $\lim_{x\to 0}\frac{\cos x - 1}{\sin x}$ ⇒

2. $\lim_{x\to \infty}\frac{e^x}{x^3}$ ⇒

3. $\lim_{x\to \infty}\sqrt{x^2+x} - \sqrt{x^2-x}$ ⇒

4. $\lim_{x\to \infty}\frac{\ln x}{x}$ ⇒

5. $\lim_{x\to \infty}\frac{\ln x}{\sqrt{x}}$ ⇒

6. $\lim_{x\to \infty}\frac{e^x + e^{-x}}{e^x - e^{-x}}$ ⇒

7. $\lim_{x\to 0}\frac{\sqrt{9+x} - 3}{x}$ ⇒

8. $\lim_{t\to 1^+}\frac{(1/t) - 1}{t^2 - 2t + 1}$ ⇒

9. $\lim_{x\to 2}\frac{2 - \sqrt{x+2}}{4 - x^2}$ ⇒

10. $\lim_{t\to \infty}\frac{t + 5 - 2/t - 1/t^3}{3t + 12 - 1/t^2}$ ⇒

11. $\lim_{y\to \infty}\frac{\sqrt{y+1} + \sqrt{y-1}}{y}$ ⇒

12. $\lim_{x\to 1}\frac{\sqrt{x} - 1}{\sqrt[3]{x} - 1}$ ⇒

13. $\displaystyle\lim_{x\to 0}\frac{(1-x)^{1/4}-1}{x}$ ⇒

14. $\displaystyle\lim_{t\to 0}\left(t+\frac{1}{t}\right)\left((4-t)^{3/2}-8\right)$ ⇒

15. $\displaystyle\lim_{t\to 0^+}\left(\frac{1}{t}+\frac{1}{\sqrt{t}}\right)(\sqrt{t+1}-1)$ ⇒

16. $\displaystyle\lim_{x\to 0}\frac{x^2}{\sqrt{2x+1}-1}$ ⇒

17. $\displaystyle\lim_{u\to 1}\frac{(u-1)^3}{(1/u)-u^2+3u-3}$ ⇒

18. $\displaystyle\lim_{x\to 0}\frac{2+(1/x)}{3-(2/x)}$ ⇒

19. $\displaystyle\lim_{x\to 0^+}\frac{1+5/\sqrt{x}}{2+1/\sqrt{x}}$ ⇒

20. $\displaystyle\lim_{x\to 0^+}\frac{3+x^{-1/2}+x^{-1}}{2+4x^{-1/2}}$ ⇒

21. $\displaystyle\lim_{x\to\infty}\frac{x+x^{1/2}+x^{1/3}}{x^{2/3}+x^{1/4}}$ ⇒

22. $\displaystyle\lim_{t\to\infty}\frac{1-\sqrt{\frac{t}{t+1}}}{2-\sqrt{\frac{4t+1}{t+2}}}$ ⇒

23. $\displaystyle\lim_{t\to\infty}\frac{1-\frac{t}{t-1}}{1-\sqrt{\frac{t}{t-1}}}$ ⇒

24. $\displaystyle\lim_{x\to-\infty}\frac{x+x^{-1}}{1+\sqrt{1-x}}$ ⇒

25. $\displaystyle\lim_{x\to\pi/2}\frac{\cos x}{(\pi/2)-x}$ ⇒

26. $\displaystyle\lim_{x\to 0}\frac{e^x-1}{x}$ ⇒

27. $\displaystyle\lim_{x\to 0}\frac{x^2}{e^x-x-1}$ ⇒

28. $\displaystyle\lim_{x\to 1}\frac{\ln x}{x-1}$ ⇒

29. $\displaystyle\lim_{x\to 0}\frac{\ln(x^2+1)}{x}$ ⇒

30. $\displaystyle\lim_{x\to 1}\frac{x\ln x}{x^2-1}$ ⇒

31. $\displaystyle\lim_{x\to 0}\frac{\sin(2x)}{\ln(x+1)}$ ⇒

32. $\displaystyle\lim_{x\to 1}\frac{x^{1/4}-1}{x}$ ⇒

33. $\displaystyle\lim_{x\to 1^+}\frac{\sqrt{x}}{x-1}$ ⇒

34. $\displaystyle\lim_{x\to 1}\frac{\sqrt{x}-1}{x-1}$ ⇒

35. $\displaystyle\lim_{x\to\infty}\frac{x^{-1}+x^{-1/2}}{x+x^{-1/2}}$ ⇒

36. $\displaystyle\lim_{x\to\infty}\frac{x+x^{-2}}{2x+x^{-2}}$ ⇒

37. $\displaystyle\lim_{x\to\infty}\frac{5+x^{-1}}{1+2x^{-1}}$ ⇒

38. $\displaystyle\lim_{x\to\infty}\frac{4x}{\sqrt{2x^2+1}}$ ⇒

39. $\displaystyle\lim_{x\to 0}\frac{3x^2+x+2}{x-4}$ ⇒

40. $\displaystyle\lim_{x\to 0}\frac{\sqrt{x+1}-1}{\sqrt{x+4}-2}$ ⇒

41. $\displaystyle\lim_{x\to 0}\frac{\sqrt{x+1}-1}{\sqrt{x+2}-2}$ ⇒

42. $\displaystyle\lim_{x\to 0^+}\frac{\sqrt{x+1}+1}{\sqrt{x+1}-1}$ ⇒

43. $\displaystyle\lim_{x\to 0}\frac{\sqrt{x^2+1}-1}{\sqrt{x+1}-1}$ ⇒

44. $\displaystyle\lim_{x\to\infty}(x+5)\left(\frac{1}{2x}+\frac{1}{x+2}\right)$ ⇒

45. $\displaystyle\lim_{x\to 0^+}(x+5)\left(\frac{1}{2x}+\frac{1}{x+2}\right)$ ⇒

46. $\displaystyle\lim_{x\to 1}(x+5)\left(\frac{1}{2x}+\frac{1}{x+2}\right)$ ⇒

47. $\displaystyle\lim_{x\to 2}\frac{x^3-6x-2}{x^3+4}$ ⇒

48. $\displaystyle\lim_{x\to 2}\frac{x^3-6x-2}{x^3-4x}$ ⇒

49. $\displaystyle\lim_{x\to 1^+}\frac{x^3+4x+8}{2x^3-2}$ ⇒

50. The function $f(x) = \dfrac{x}{\sqrt{x^2+1}}$ has two horizontal asymptotes. Find them and give a rough sketch of f with its horizontal asymptotes. ⇒

4.11 Hyperbolic Functions

The hyperbolic functions appear with some frequency in applications, and are quite similar in many respects to the trigonometric functions. This is a bit surprising given our initial definitions.

DEFINITION 4.11.1 The **hyperbolic cosine** is the function

$$\cosh x = \frac{e^x + e^{-x}}{2},$$

and the **hyperbolic sine** is the function

$$\sinh x = \frac{e^x - e^{-x}}{2}.$$

□

Notice that cosh is even (that is, $\cosh(-x) = \cosh(x)$) while sinh is odd ($\sinh(-x) = -\sinh(x)$), and $\cosh x + \sinh x = e^x$. Also, for all x, $\cosh x > 0$, while $\sinh x = 0$ if and only if $e^x - e^{-x} = 0$, which is true precisely when $x = 0$.

LEMMA 4.11.2 The range of $\cosh x$ is $[1, \infty)$.

Proof. Let $y = \cosh x$. We solve for x:

$$\begin{aligned} y &= \frac{e^x + e^{-x}}{2} \\ 2y &= e^x + e^{-x} \\ 2ye^x &= e^{2x} + 1 \\ 0 &= e^{2x} - 2ye^x + 1 \\ e^x &= \frac{2y \pm \sqrt{4y^2 - 4}}{2} \\ e^x &= y \pm \sqrt{y^2 - 1} \end{aligned}$$

From the last equation, we see $y^2 \geq 1$, and since $y \geq 0$, it follows that $y \geq 1$.

Now suppose $y \geq 1$, so $y \pm \sqrt{y^2-1} > 0$. Then $x = \ln(y \pm \sqrt{y^2-1})$ is a real number, and $y = \cosh x$, so y is in the range of $\cosh(x)$. ■

DEFINITION 4.11.3 The other hyperbolic functions are

$$\tanh x = \frac{\sinh x}{\cosh x}$$
$$\coth x = \frac{\cosh x}{\sinh x}$$
$$\operatorname{sech} x = \frac{1}{\cosh x}$$
$$\operatorname{csch} x = \frac{1}{\sinh x}$$

The domain of coth and csch is $x \neq 0$ while the domain of the other hyperbolic functions is all real numbers. Graphs are shown in figure 4.11.1 □

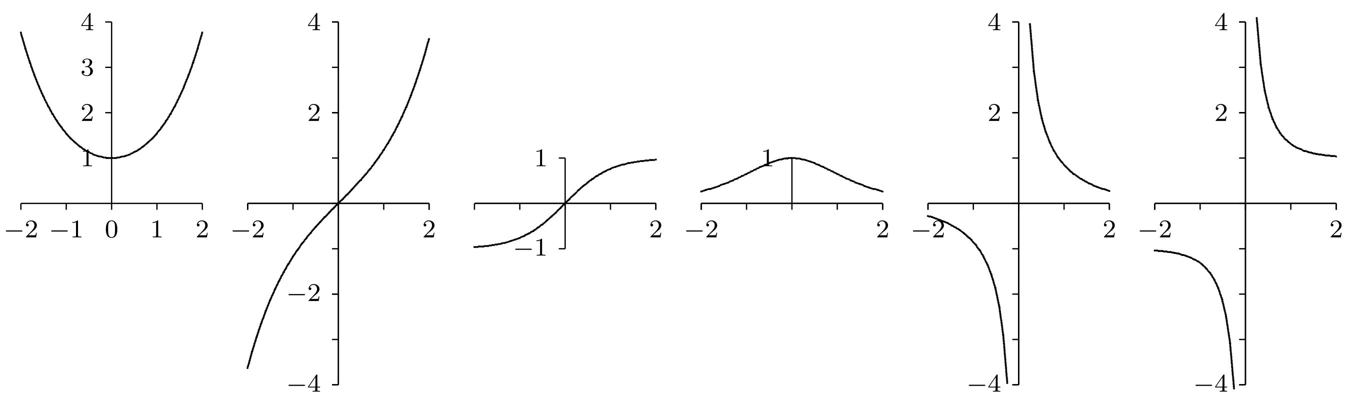


Figure 4.11.1 The hyperbolic functions: cosh, sinh, tanh, sech, csch, coth.

Certainly the hyperbolic functions do not closely resemble the trigonometric functions graphically. But they do have analogous properties, beginning with the following identity.

THEOREM 4.11.4 For all x in $\mathbb{R}$, $\cosh^2 x - \sinh^2 x = 1$.

Proof. The proof is a straightforward computation:

$$\cosh^2 x - \sinh^2 x = \frac{(e^x + e^{-x})^2}{4} - \frac{(e^x - e^{-x})^2}{4} = \frac{e^{2x} + 2 + e^{-2x} - e^{2x} + 2 - e^{-2x}}{4} = \frac{4}{4} = 1.$$

■

This immediately gives two additional identities:

$$1 - \tanh^2 x = \operatorname{sech}^2 x \qquad \text{and} \qquad \coth^2 x - 1 = \operatorname{csch}^2 x.$$

The identity of the theorem also helps to provide a geometric motivation. Recall that the graph of $x^2 - y^2 = 1$ is a hyperbola with asymptotes $x = \pm y$ whose x-intercepts are

± 1. If (x,y) is a point on the right half of the hyperbola, and if we let $x = \cosh t$, then $y = \pm\sqrt{x^2-1} = \pm\sqrt{\cosh^2 x - 1} = \pm\sinh t$. So for some suitable t, $\cosh t$ and $\sinh t$ are the coordinates of a typical point on the hyperbola. In fact, it turns out that t is twice the area shown in the first graph of figure 4.11.2. Even this is analogous to trigonometry; $\cos t$ and $\sin t$ are the coordinates of a typical point on the unit circle, and t is twice the area shown in the second graph of figure 4.11.2.

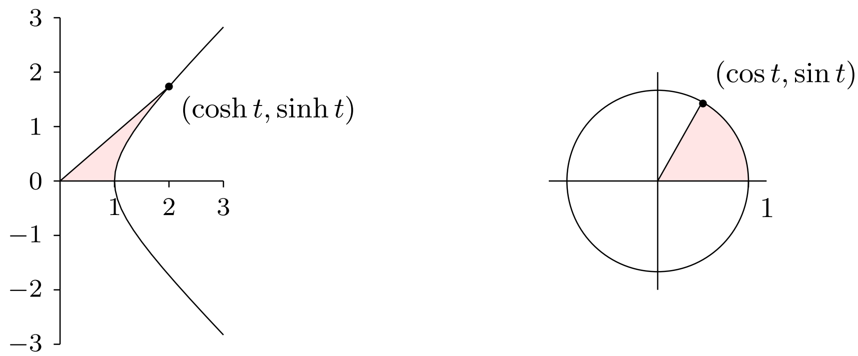


Figure 4.11.2 Geometric definitions of sin, cos, sinh, cosh: t is twice the shaded area in each figure.

Given the definitions of the hyperbolic functions, finding their derivatives is straightforward. Here again we see similarities to the trigonometric functions.

THEOREM 4.11.5 $\dfrac{d}{dx}\cosh x = \sinh x$ and $\dfrac{d}{dx}\sinh x = \cosh x$.

Proof. $\dfrac{d}{dx}\cosh x = \dfrac{d}{dx}\dfrac{e^x+e^{-x}}{2} = \dfrac{e^x-e^{-x}}{2} = \sinh x$, and $\dfrac{d}{dx}\sinh x = \dfrac{d}{dx}\dfrac{e^x-e^{-x}}{2} = \dfrac{e^x+e^{-x}}{2} = \cosh x$. ∎

Since $\cosh x > 0$, $\sinh x$ is increasing and hence injective, so $\sinh x$ has an inverse, $\operatorname{arcsinh} x$. Also, $\sinh x > 0$ when $x > 0$, so $\cosh x$ is injective on $[0,\infty)$ and has a (partial) inverse, $\operatorname{arccosh} x$. The other hyperbolic functions have inverses as well, though $\operatorname{arcsech} x$ is only a partial inverse. We may compute the derivatives of these functions as we have other inverse functions.

THEOREM 4.11.6 $\dfrac{d}{dx}\operatorname{arcsinh} x = \dfrac{1}{\sqrt{1+x^2}}$.

Proof. Let $y = \operatorname{arcsinh} x$, so $\sinh y = x$. Then $\dfrac{d}{dx}\sinh y = \cosh(y)\cdot y' = 1$, and so $y' = \dfrac{1}{\cosh y} = \dfrac{1}{\sqrt{1+\sinh^2 y}} = \dfrac{1}{\sqrt{1+x^2}}$. ∎

The other derivatives are left to the exercises.

Exercises 4.11.

1. Show that the range of $\sinh x$ is all real numbers. (Hint: show that if $y = \sinh x$ then $x = \ln(y + \sqrt{y^2+1})$.)

2. Compute the following limits:

a. $\lim_{x\to\infty} \cosh x$

b. $\lim_{x\to\infty} \sinh x$

c. $\lim_{x\to\infty} \tanh x$

d. $\lim_{x\to\infty} (\cosh x - \sinh x)$

3. Show that the range of $\tanh x$ is $(-1, 1)$. What are the ranges of coth, sech, and csch? (Use the fact that they are reciprocal functions.)

4. Prove that for every $x, y \in \mathbb{R}$, $\sinh(x+y) = \sinh x \cosh y + \cosh x \sinh y$. Obtain a similar identity for $\sinh(x-y)$.

5. Prove that for every $x, y \in \mathbb{R}$, $\cosh(x+y) = \cosh x \cosh y + \sinh x \sinh y$. Obtain a similar identity for $\cosh(x-y)$.

6. Use exercises 4 and 5 to show that $\sinh(2x) = 2\sinh x \cosh x$ and $\cosh(2x) = \cosh^2 x + \sinh^2 x$ for every x. Conclude also that $(\cosh(2x) - 1)/2 = \sinh^2 x$.

7. Show that $\frac{d}{dx}(\tanh x) = \operatorname{sech}^2 x$. Compute the derivatives of the remaining hyperbolic functions as well.

8. What are the domains of the six inverse hyperbolic functions?

9. Sketch the graphs of all six inverse hyperbolic functions.

5

Curve Sketching

Whether we are interested in a function as a purely mathematical object or in connection with some application to the real world, it is often useful to know what the graph of the function looks like. We can obtain a good picture of the graph using certain crucial information provided by derivatives of the function and certain limits.

5.1 Maxima and Minima

A **local maximum point** on a function is a point (x, y) on the graph of the function whose y coordinate is larger than all other y coordinates on the graph at points "close to" (x, y). More precisely, $(x, f(x))$ is a local maximum if there is an interval (a, b) with $a < x < b$ and $f(x) \geq f(z)$ for every z in (a, b). Similarly, (x, y) is a **local minimum point** if it has locally the smallest y coordinate. Again being more precise: $(x, f(x))$ is a local minimum if there is an interval (a, b) with $a < x < b$ and $f(x) \leq f(z)$ for every z in (a, b). A **local extremum** is either a local minimum or a local maximum.

Local maximum and minimum points are quite distinctive on the graph of a function, and are therefore useful in understanding the shape of the graph. In many applied problems we want to find the largest or smallest value that a function achieves (for example, we might want to find the minimum cost at which some task can be performed) and so identifying maximum and minimum points will be useful for applied problems as well. Some examples of local maximum and minimum points are shown in figure 5.1.1.

If $(x, f(x))$ is a point where $f(x)$ reaches a local maximum or minimum, and if the derivative of f exists at x, then the graph has a tangent line and the tangent line must be horizontal. This is important enough to state as a theorem, though we will not prove it.

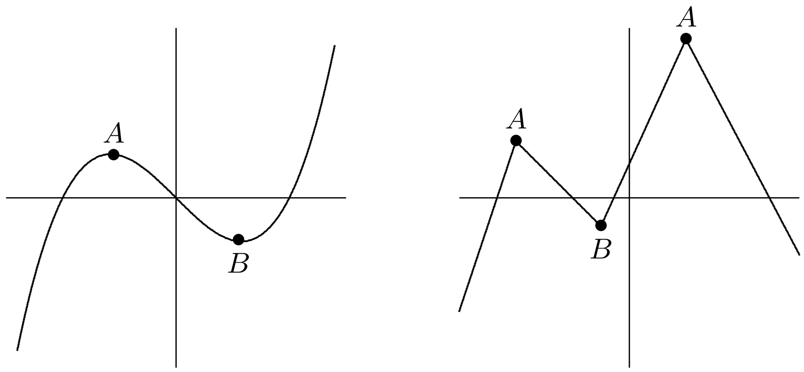


Figure 5.1.1 Some local maximum points (A) and minimum points (B).

THEOREM 5.1.1 Fermat's Theorem If $f(x)$ has a local extremum at $x = a$ and f is differentiable at a, then $f'(a) = 0$. ■

Thus, the only points at which a function can have a local maximum or minimum are points at which the derivative is zero, as in the left hand graph in figure 5.1.1, or the derivative is undefined, as in the right hand graph. Any value of x for which $f'(x)$ is zero or undefined is called a **critical value** for f. When looking for local maximum and minimum points, you are likely to make two sorts of mistakes: You may forget that a maximum or minimum can occur where the derivative does not exist, and so forget to check whether the derivative exists everywhere. You might also assume that any place that the derivative is zero is a local maximum or minimum point, but this is not true. A portion of the graph of $f(x) = x^3$ is shown in figure 5.1.2. The derivative of f is $f'(x) = 3x^2$, and $f'(0) = 0$, but there is neither a maximum nor minimum at $(0, 0)$.

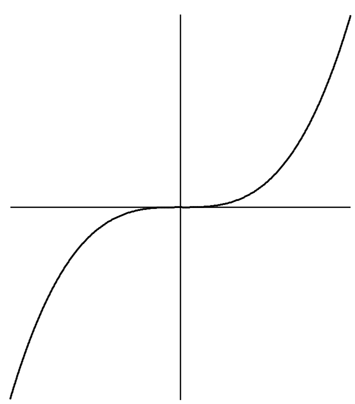


Figure 5.1.2 No maximum or minimum even though the derivative is zero.

Since the derivative is zero or undefined at both local maximum and local minimum points, we need a way to determine which, if either, actually occurs. The most elementary approach, but one that is often tedious or difficult, is to test directly whether the y coordinates "near" the potential maximum or minimum are above or below the y coordinate at the point of interest. Of course, there are too many points "near" the point to test, but a little thought shows we need only test two provided we know that f is continuous (recall that this means that the graph of f has no jumps or gaps).

Suppose, for example, that we have identified three points at which f' is zero or nonexistent: (x_1, y_1), (x_2, y_2), (x_3, y_3), and $x_1 < x_2 < x_3$ (see figure 5.1.3). Suppose that we compute the value of $f(a)$ for $x_1 < a < x_2$, and that $f(a) < f(x_2)$. What can we say about the graph between a and x_2? Could there be a point $(b, f(b))$, $a < b < x_2$ with $f(b) > f(x_2)$? No: if there were, the graph would go up from $(a, f(a))$ to $(b, f(b))$ then down to $(x_2, f(x_2))$ and somewhere in between would have a local maximum point. (This is not obvious; it is a result of the Extreme Value Theorem, theorem 6.1.2.) But at that local maximum point the derivative of f would be zero or nonexistent, yet we already know that the derivative is zero or nonexistent only at x_1, x_2, and x_3. The upshot is that one computation tells us that $(x_2, f(x_2))$ has the largest y coordinate of any point on the graph near x_2 and to the left of x_2. We can perform the same test on the right. If we find that on both sides of x_2 the values are smaller, then there must be a local maximum at $(x_2, f(x_2))$; if we find that on both sides of x_2 the values are larger, then there must be a local minimum at $(x_2, f(x_2))$; if we find one of each, then there is neither a local maximum or minimum at x_2.

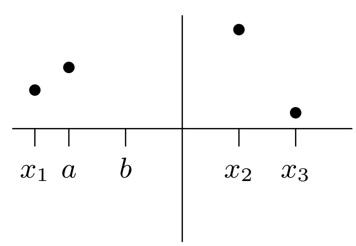


Figure 5.1.3 Testing for a maximum or minimum.

It is not always easy to compute the value of a function at a particular point. The task is made easier by the availability of calculators and computers, but they have their own drawbacks—they do not always allow us to distinguish between values that are very close together. Nevertheless, because this method is conceptually simple and sometimes easy to perform, you should always consider it.

EXAMPLE 5.1.2 Find all local maximum and minimum points for the function $f(x) = x^3 - x$. The derivative is $f'(x) = 3x^2 - 1$. This is defined everywhere and is zero at $x = \pm\sqrt{3}/3$. Looking first at $x = \sqrt{3}/3$, we see that $f(\sqrt{3}/3) = -2\sqrt{3}/9$. Now we test two points on either side of $x = \sqrt{3}/3$, making sure that neither is farther away than the nearest critical value; since $\sqrt{3} < 3$, $\sqrt{3}/3 < 1$ and we can use $x = 0$ and $x = 1$. Since $f(0) = 0 > -2\sqrt{3}/9$ and $f(1) = 0 > -2\sqrt{3}/9$, there must be a local minimum at

$x = \sqrt{3}/3$. For $x = -\sqrt{3}/3$, we see that $f(-\sqrt{3}/3) = 2\sqrt{3}/9$. This time we can use $x = 0$ and $x = -1$, and we find that $f(-1) = f(0) = 0 < 2\sqrt{3}/9$, so there must be a local maximum at $x = -\sqrt{3}/3$. □

Of course this example is made very simple by our choice of points to test, namely $x = -1, 0, 1$. We could have used other values, say $-5/4$, $1/3$, and $3/4$, but this would have made the calculations considerably more tedious.

EXAMPLE 5.1.3 Find all local maximum and minimum points for $f(x) = \sin x + \cos x$. The derivative is $f'(x) = \cos x - \sin x$. This is always defined and is zero whenever $\cos x = \sin x$. Recalling that the $\cos x$ and $\sin x$ are the x and y coordinates of points on a unit circle, we see that $\cos x = \sin x$ when x is $\pi/4$, $\pi/4 \pm \pi$, $\pi/4 \pm 2\pi$, $\pi/4 \pm 3\pi$, etc. Since both sine and cosine have a period of 2π, we need only determine the status of $x = \pi/4$ and $x = 5\pi/4$. We can use 0 and $\pi/2$ to test the critical value $x = \pi/4$. We find that $f(\pi/4) = \sqrt{2}$, $f(0) = 1 < \sqrt{2}$ and $f(\pi/2) = 1$, so there is a local maximum when $x = \pi/4$ and also when $x = \pi/4 \pm 2\pi$, $\pi/4 \pm 4\pi$, etc. We can summarize this more neatly by saying that there are local maxima at $\pi/4 \pm 2k\pi$ for every integer k.

We use π and 2π to test the critical value $x = 5\pi/4$. The relevant values are $f(5\pi/4) = -\sqrt{2}$, $f(\pi) = -1 > -\sqrt{2}$, $f(2\pi) = 1 > -\sqrt{2}$, so there is a local minimum at $x = 5\pi/4$, $5\pi/4 \pm 2\pi$, $5\pi/4 \pm 4\pi$, etc. More succinctly, there are local minima at $5\pi/4 \pm 2k\pi$ for every integer k. □

Exercises 5.1.

In problems 1–12, find all local maximum and minimum points (x, y) by the method of this section.

1. $y = x^2 - x$ ⇒

2. $y = 2 + 3x - x^3$ ⇒

3. $y = x^3 - 9x^2 + 24x$ ⇒

4. $y = x^4 - 2x^2 + 3$ ⇒

5. $y = 3x^4 - 4x^3$ ⇒

6. $y = (x^2 - 1)/x$ ⇒

7. $y = 3x^2 - (1/x^2)$ ⇒

8. $y = \cos(2x) - x$ ⇒

9. $f(x) = \begin{cases} x - 1 & x < 2 \\ x^2 & x \ge 2 \end{cases}$ ⇒

10. $f(x) = \begin{cases} x - 3 & x < 3 \\ x^3 & 3 \le x \le 5 \\ 1/x & x > 5 \end{cases}$ ⇒

11. $f(x) = x^2 - 98x + 4$ ⇒

12. $f(x) = \begin{cases} -2 & x = 0 \\ 1/x^2 & x \ne 0 \end{cases}$ ⇒

13. For any real number x there is a unique integer n such that $n \le x < n + 1$, and the greatest integer function is defined as $\lfloor x \rfloor = n$. Where are the critical values of the greatest integer function? Which are local maxima and which are local minima?

14. Explain why the function $f(x) = 1/x$ has no local maxima or minima.

15. How many critical points can a quadratic polynomial function have? ⇒

16. Show that a cubic polynomial can have at most two critical points. Give examples to show that a cubic polynomial can have zero, one, or two critical points.

17. Explore the family of functions $f(x) = x^3 + cx + 1$ where c is a constant. How many and what types of local extremes are there? Your answer should depend on the value of c, that is, different values of c will give different answers.

18. We generalize the preceding two questions. Let n be a positive integer and let f be a polynomial of degree n. How many critical points can f have? (Hint: Recall the **Fundamental Theorem of Algebra**, which says that a polynomial of degree n has at most n roots.)

5.2 The first derivative test

The method of the previous section for deciding whether there is a local maximum or minimum at a critical value is not always convenient. We can instead use information about the derivative $f'(x)$ to decide; since we have already had to compute the derivative to find the critical values, there is often relatively little extra work involved in this method.

How can the derivative tell us whether there is a maximum, minimum, or neither at a point? Suppose that $f'(a) = 0$. If there is a local maximum when $x = a$, the function must be lower near $x = a$ than it is right at $x = a$. If the derivative exists near $x = a$, this means $f'(x) > 0$ when x is near a and $x < a$, because the function must "slope up" just to the left of a. Similarly, $f'(x) < 0$ when x is near a and $x > a$, because f slopes down from the local maximum as we move to the right. Using the same reasoning, if there is a local minimum at $x = a$, the derivative of f must be negative just to the left of a and positive just to the right. If the derivative exists near a but does not change from positive to negative or negative to positive, that is, it is positive on both sides or negative on both sides, then there is neither a maximum nor minimum when $x = a$. See the first graph in figure 5.1.1 and the graph in figure 5.1.2 for examples.

EXAMPLE 5.2.1 Find all local maximum and minimum points for $f(x) = \sin x + \cos x$ using the first derivative test. The derivative is $f'(x) = \cos x - \sin x$ and from example 5.1.3 the critical values we need to consider are $\pi/4$ and $5\pi/4$.

The graphs of $\sin x$ and $\cos x$ are shown in figure 5.2.1. Just to the left of $\pi/4$ the cosine is larger than the sine, so $f'(x)$ is positive; just to the right the cosine is smaller than the sine, so $f'(x)$ is negative. This means there is a local maximum at $\pi/4$. Just to the left of $5\pi/4$ the cosine is smaller than the sine, and to the right the cosine is larger than the sine. This means that the derivative $f'(x)$ is negative to the left and positive to the right, so f has a local minimum at $5\pi/4$. □

Exercises 5.2.

In 1–13, find all critical points and identify them as local maximum points, local minimum points, or neither.

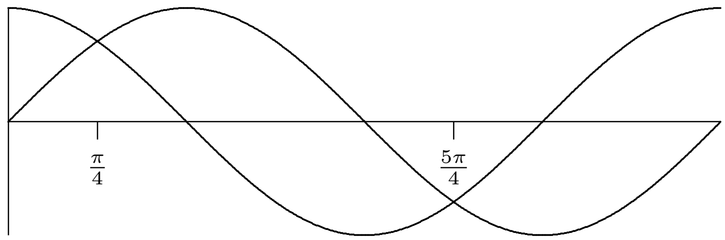


Figure 5.2.1 The sine and cosine.

1. $y = x^2 - x$ ⇒

2. $y = 2 + 3x - x^3$ ⇒

3. $y = x^3 - 9x^2 + 24x$ ⇒

4. $y = x^4 - 2x^2 + 3$ ⇒

5. $y = 3x^4 - 4x^3$ ⇒

6. $y = (x^2 - 1)/x$ ⇒

7. $y = 3x^2 - (1/x^2)$ ⇒

8. $y = \cos(2x) - x$ ⇒

9. $f(x) = (5 - x)/(x + 2)$ ⇒

10. $f(x) = |x^2 - 121|$ ⇒

11. $f(x) = x^3/(x + 1)$ ⇒

12. $f(x) = \begin{cases} x^2 \sin(1/x) & x \neq 0 \\ 0 & x = 0 \end{cases}$

13. $f(x) = \sin^2 x$ ⇒

14. Find the maxima and minima of $f(x) = \sec x$. ⇒

15. Let $f(\theta) = \cos^2(\theta) - 2\sin(\theta)$. Find the intervals where f is increasing and the intervals where f is decreasing in $[0, 2\pi]$. Use this information to classify the critical points of f as either local maximums, local minimums, or neither. ⇒

16. Let $r > 0$. Find the local maxima and minima of the function $f(x) = \sqrt{r^2 - x^2}$ on its domain $[-r, r]$.

17. Let $f(x) = ax^2 + bx + c$ with $a \neq 0$. Show that f has exactly one critical point. Give conditions on a and b which guarantee that the critical point will be a maximum. It is possible to see this without using calculus at all; explain.

5.3 The second derivative test

derivative!secondThe basis of the first derivative test is that if the derivative changes from positive to negative at a point at which the derivative is zero then there is a local maximum at the point, and similarly for a local minimum. If f' changes from positive to negative it is decreasing; this means that the derivative of f', f'', might be negative, and if in fact f'' is negative then f' is definitely decreasing, so there is a local maximum at the point in question. Note well that f' might change from positive to negative while f'' is zero, in which case f'' gives us no information about the critical value. Similarly, if f' changes from negative to positive there is a local minimum at the point, and f' is increasing. If $f'' > 0$ at the point, this tells us that f' is increasing, and so there is a local minimum.

EXAMPLE 5.3.1 Consider again $f(x) = \sin x + \cos x$, with $f'(x) = \cos x - \sin x$ and $f''(x) = -\sin x - \cos x$. Since $f''(\pi/4) = -\sqrt{2}/2 - \sqrt{2}/2 = -\sqrt{2} < 0$, we know there is a local maximum at $\pi/4$. Since $f''(5\pi/4) = --\sqrt{2}/2 - -\sqrt{2}/2 = \sqrt{2} > 0$, there is a local minimum at $5\pi/4$. □

When it works, the second derivative test is often the easiest way to identify local maximum and minimum points. Sometimes the test fails, and sometimes the second derivative is quite difficult to evaluate; in such cases we must fall back on one of the previous tests.

EXAMPLE 5.3.2 Let $f(x) = x^4$. The derivatives are $f'(x) = 4x^3$ and $f''(x) = 12x^2$. Zero is the only critical value, but $f''(0) = 0$, so the second derivative test tells us nothing. However, $f(x)$ is positive everywhere except at zero, so clearly $f(x)$ has a local minimum at zero. On the other hand, $f(x) = -x^4$ also has zero as its only critical value, and the second derivative is again zero, but $-x^4$ has a local maximum at zero. □

Exercises 5.3.

Find all local maximum and minimum points by the second derivative test.

1. $y = x^2 - x$ ⇒

2. $y = 2 + 3x - x^3$ ⇒

3. $y = x^3 - 9x^2 + 24x$ ⇒

4. $y = x^4 - 2x^2 + 3$ ⇒

5. $y = 3x^4 - 4x^3$ ⇒

6. $y = (x^2 - 1)/x$ ⇒

7. $y = 3x^2 - (1/x^2)$ ⇒

8. $y = \cos(2x) - x$ ⇒

9. $y = 4x + \sqrt{1 - x}$ ⇒

10. $y = (x + 1)/\sqrt{5x^2 + 35}$ ⇒

11. $y = x^5 - x$ ⇒

12. $y = 6x + \sin 3x$ ⇒

13. $y = x + 1/x$ ⇒

14. $y = x^2 + 1/x$ ⇒

15. $y = (x + 5)^{1/4}$ ⇒

16. $y = \tan^2 x$ ⇒

17. $y = \cos^2 x - \sin^2 x$ ⇒

18. $y = \sin^3 x$ ⇒

5.4 Concavity and inflection points

We know that the sign of the derivative tells us whether a function is increasing or decreasing; for example, when $f'(x) > 0$, $f(x)$ is increasing. The sign of the second derivative $f''(x)$ tells us whether f' is increasing or decreasing; we have seen that if f' is zero and increasing at a point then there is a local minimum at the point, and if f' is zero and decreasing at a point then there is a local maximum at the point. Thus, we extracted information about f from information about f''.

We can get information from the sign of f'' even when f' is not zero. Suppose that $f''(a) > 0$. This means that near $x = a$, f' is increasing. If $f'(a) > 0$, this means that f slopes up and is getting steeper; if $f'(a) < 0$, this means that f slopes down and is getting

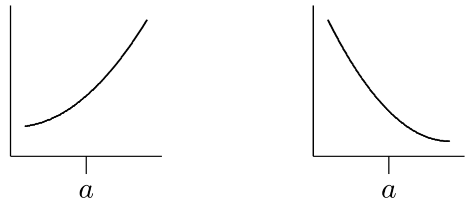


Figure 5.4.1 $f''(a) > 0$: $f'(a)$ positive and increasing, $f'(a)$ negative and increasing.

less steep. The two situations are shown in figure 5.4.1. A curve that is shaped like this is called **concave up**.

Now suppose that $f''(a) < 0$. This means that near $x = a$, f' is decreasing. If $f'(a) > 0$, this means that f slopes up and is getting less steep; if $f'(a) < 0$, this means that f slopes down and is getting steeper. The two situations are shown in figure 5.4.2. A curve that is shaped like this is called **concave down**.

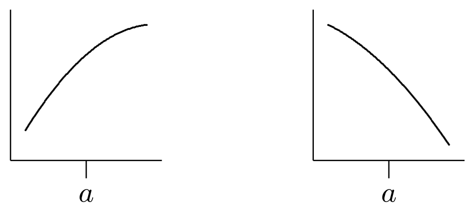


Figure 5.4.2 $f''(a) < 0$: $f'(a)$ positive and decreasing, $f'(a)$ negative and decreasing.

If we are trying to understand the shape of the graph of a function, knowing where it is concave up and concave down helps us to get a more accurate picture. Of particular interest are points at which the concavity changes from up to down or down to up; such points are called **inflection points**. If the concavity changes from up to down at $x = a$, f'' changes from positive to the left of a to negative to the right of a, and usually $f''(a) = 0$. We can identify such points by first finding where $f''(x)$ is zero and then checking to see whether $f''(x)$ does in fact go from positive to negative or negative to positive at these points. Note that it is possible that $f''(a) = 0$ but the concavity is the same on both sides; $f(x) = x^4$ at $x = 0$ is an example.

EXAMPLE 5.4.1 Describe the concavity of $f(x) = x^3 - x$. $f'(x) = 3x^2 - 1$, $f''(x) = 6x$. Since $f''(0) = 0$, there is potentially an inflection point at zero. Since $f''(x) > 0$ when $x > 0$ and $f''(x) < 0$ when $x < 0$ the concavity does change from down to up at zero, and the curve is concave down for all $x < 0$ and concave up for all $x > 0$. □

Note that we need to compute and analyze the second derivative to understand concavity, so we may as well try to use the second derivative test for maxima and minima. If for some reason this fails we can then try one of the other tests.

Exercises 5.4.

Describe the concavity of the functions in 1–18.

1. $y = x^2 - x$ ⇒

2. $y = 2 + 3x - x^3$ ⇒

3. $y = x^3 - 9x^2 + 24x$ ⇒

4. $y = x^4 - 2x^2 + 3$ ⇒

5. $y = 3x^4 - 4x^3$ ⇒

6. $y = (x^2 - 1)/x$ ⇒

7. $y = 3x^2 - (1/x^2)$ ⇒

8. $y = \sin x + \cos x$ ⇒

9. $y = 4x + \sqrt{1-x}$ ⇒

10. $y = (x+1)/\sqrt{5x^2+35}$ ⇒

11. $y = x^5 - x$ ⇒

12. $y = 6x + \sin 3x$ ⇒

13. $y = x + 1/x$ ⇒

14. $y = x^2 + 1/x$ ⇒

15. $y = (x+5)^{1/4}$ ⇒

16. $y = \tan^2 x$ ⇒

17. $y = \cos^2 x - \sin^2 x$ ⇒

18. $y = \sin^3 x$ ⇒

19. Identify the intervals on which the graph of the function $f(x) = x^4 - 4x^3 + 10$ is of one of these four shapes: concave up and increasing; concave up and decreasing; concave down and increasing; concave down and decreasing. ⇒

20. Describe the concavity of $y = x^3 + bx^2 + cx + d$. You will need to consider different cases, depending on the values of the coefficients.

21. Let n be an integer greater than or equal to two, and suppose f is a polynomial of degree n. How many inflection points can f have? Hint: Use the second derivative test and the fundamental theorem of algebra.

5.5 Asymptotes and Other Things to Look For

A vertical asymptote is a place where the function becomes infinite, typically because the formula for the function has a denominator that becomes zero. For example, the reciprocal function $f(x) = 1/x$ has a vertical asymptote at $x = 0$, and the function $\tan x$ has a vertical asymptote at $x = \pi/2$ (and also at $x = -\pi/2$, $x = 3\pi/2$, etc.). Whenever the formula for a function contains a denominator it is worth looking for a vertical asymptote by checking to see if the denominator can ever be zero, and then checking the limit at such points. Note that there is not always a vertical asymptote where the denominator is zero: $f(x) = (\sin x)/x$ has a zero denominator at $x = 0$, but since $\lim_{x\to 0}(\sin x)/x = 1$ there is no asymptote there.

A horizontal asymptote is a horizontal line to which $f(x)$ gets closer and closer as x approaches ∞ (or as x approaches $-\infty$). For example, the reciprocal function has the x-axis for a horizontal asymptote. Horizontal asymptotes can be identified by computing the limits $\lim_{x\to\infty} f(x)$ and $\lim_{x\to-\infty} f(x)$. Since $\lim_{x\to\infty} 1/x = \lim_{x\to-\infty} 1/x = 0$, the line $y = 0$ (that is, the x-axis) is a horizontal asymptote in both directions.

Some functions have asymptotes that are neither horizontal nor vertical, but some other line. Such asymptotes are somewhat more difficult to identify and we will ignore them.

If the domain of the function does not extend out to infinity, we should also ask what happens as x approaches the boundary of the domain. For example, the function $y = f(x) = 1/\sqrt{r^2 - x^2}$ has domain $-r < x < r$, and y becomes infinite as x approaches either r or $-r$. In this case we might also identify this behavior because when $x = \pm r$ the denominator of the function is zero.

If there are any points where the derivative fails to exist (a cusp or corner), then we should take special note of what the function does at such a point.

Finally, it is worthwhile to notice any symmetry. A function $f(x)$ that has the same value for $-x$ as for x, i.e., $f(-x) = f(x)$, is called an "even function." Its graph is symmetric with respect to the y-axis. Some examples of even functions are: x^n when n is an even number, $\cos x$, and $\sin^2 x$. On the other hand, a function that satisfies the property $f(-x) = -f(x)$ is called an "odd function." Its graph is symmetric with respect to the origin. Some examples of odd functions are: x^n when n is an odd number, $\sin x$, and $\tan x$. Of course, most functions are neither even nor odd, and do not have any particular symmetry.

Exercises 5.5.

Sketch the curves. Identify clearly any interesting features, including local maximum and minimum points, inflection points, asymptotes, and intercepts.

1. $y = x^5 - 5x^4 + 5x^3$

2. $y = x^3 - 3x^2 - 9x + 5$

3. $y = (x-1)^2(x+3)^{2/3}$

4. $x^2 + x^2y^2 = a^2y^2$, $a > 0$.

5. $y = xe^x$

6. $y = (e^x + e^{-x})/2$

7. $y = e^{-x}\cos x$

8. $y = e^x - \sin x$

9. $y = e^x/x$

10. $y = 4x + \sqrt{1-x}$

11. $y = (x+1)/\sqrt{5x^2 + 35}$

12. $y = x^5 - x$

13. $y = 6x + \sin 3x$

14. $y = x + 1/x$

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

16. $y = (x+5)^{1/4}$

17. $y = \tan^2 x$

18. $y = \cos^2 x - \sin^2 x$

19. $y = \sin^3 x$

20. $y = x(x^2 + 1)$

21. $y = x^3 + 6x^2 + 9x$

22. $y = x/(x^2 - 9)$

23. $y = x^2/(x^2 + 9)$

24. $y = 2\sqrt{x} - x$

25. $y = 3\sin(x) - \sin^3(x)$, for $x \in [0, 2\pi]$

26. $y = (x-1)/(x^2)$

For each of the following five functions, identify any vertical and horizontal asymptotes, and identify intervals on which the function is concave up and increasing; concave up and decreasing; concave down and increasing; concave down and decreasing.

27. $f(\theta) = \sec(\theta)$

28. $f(x) = 1/(1 + x^2)$

29. $f(x) = (x - 3)/(2x - 2)$

30. $f(x) = 1/(1 - x^2)$

31. $f(x) = 1 + 1/(x^2)$

32. Let $f(x) = 1/(x^2 - a^2)$, where $a \geq 0$. Find any vertical and horizontal asymptotes and the intervals upon which the given function is concave up and increasing; concave up and decreasing; concave down and increasing; concave down and decreasing. Discuss how the value of a affects these features.

6

Applications of the Derivative

6.1 Optimization

Many important applied problems involve finding the best way to accomplish some task. Often this involves finding the maximum or minimum value of some function: the minimum time to make a certain journey, the minimum cost for doing a task, the maximum power that can be generated by a device, and so on. Many of these problems can be solved by finding the appropriate function and then using techniques of calculus to find the maximum or the minimum value required.

Generally such a problem will have the following mathematical form: Find the largest (or smallest) value of $f(x)$ when $a \le x \le b$. Sometimes a or b are infinite, but frequently the real world imposes some constraint on the values that x may have.

Such a problem differs in two ways from the local maximum and minimum problems we encountered when graphing functions: We are interested only in the function between a and b, and we want to know the largest or smallest value that $f(x)$ takes on, not merely values that are the largest or smallest in a small interval. That is, we seek not a local maximum or minimum but a **global** maximum or minimum, sometimes also called an **absolute** maximum or minimum.

Any global maximum or minimum must of course be a local maximum or minimum. If we find all possible local extrema, then the global maximum, *if it exists*, must be the largest of the local maxima and the global minimum, *if it exists*, must be the smallest of the local minima. We already know where local extrema can occur: only at those points at which $f'(x)$ is zero or undefined. Actually, there are two additional points at which a maximum or minimum can occur if the endpoints a and b are not infinite, namely, at a

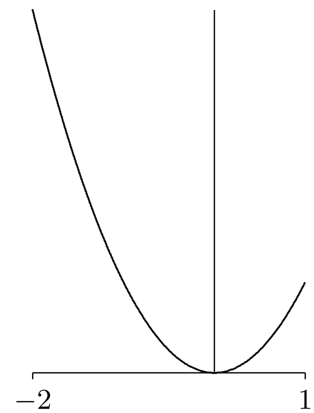


Figure 6.1.1 The function $f(x) = x^2$ restricted to $[-2, 1]$

and b. We have not previously considered such points because we have not been interested in limiting a function to a small interval. An example should make this clear.

EXAMPLE 6.1.1 Find the maximum and minimum values of $f(x) = x^2$ on the interval $[-2, 1]$, shown in figure 6.1.1. We compute $f'(x) = 2x$, which is zero at $x = 0$ and is always defined.

Since $f'(1) = 2$ we would not normally flag $x = 1$ as a point of interest, but it is clear from the graph that *when $f(x)$ is restricted to $[-2, 1]$ there is a local maximum at $x = 1$.* Likewise we would not normally pay attention to $x = -2$, but since we have truncated f at -2 we have introduced a new local maximum there as well. In a technical sense nothing new is going on here: When we truncate f we actually create a new function, let's call it g, that is defined only on the interval $[-2, 1]$. If we try to compute the derivative of this new function we actually find that it does not have a derivative at -2 or 1. Why? Because to compute the derivative at 1 we must compute the limit

$$\lim_{\Delta x \to 0} \frac{g(1 + \Delta x) - g(1)}{\Delta x}.$$

This limit does not exist because when $\Delta x > 0$, $g(1 + \Delta x)$ is not defined. It is simpler, however, simply to remember that we must always check the endpoints.

So the function g, that is, f restricted to $[-2, 1]$, has one critical value and two finite endpoints, any of which might be the global maximum or minimum. We could first determine which of these are local maximum or minimum points (or neither); then the largest local maximum must be the global maximum and the smallest local minimum must be the global minimum. It is usually easier, however, to compute the value of f at every point at which the global maximum or minimum might occur; the largest of these is the global maximum, the smallest is the global minimum.

So we compute $f(-2) = 4$, $f(0) = 0$, $f(1) = 1$. The global maximum is 4 at $x = -2$ and the global minimum is 0 at $x = 0$. □

It is possible that there is no global maximum or minimum. It is difficult, and not particularly useful, to express a complete procedure for determining whether this is the case. Generally, the best approach is to gain enough understanding of the shape of the graph to decide. Fortunately, only a rough idea of the shape is usually needed.

There are some particularly nice cases that are easy. A continuous function on a closed interval $[a, b]$ *always* has both a global maximum and a global minimum, so examining the critical values and the endpoints is enough:

THEOREM 6.1.2 Extreme value theorem If f is continuous on a closed interval $[a, b]$, then it has both a minimum and a maximum point. That is, there are real numbers c and d in $[a, b]$ so that for every x in $[a, b]$, $f(x) \le f(c)$ and $f(x) \ge f(d)$. ■

Another easy case: If a function is continuous and has a single critical value, then if there is a local maximum at the critical value it is a global maximum, and if it is a local minimum it is a global minimum. There may also be a global minimum in the first case, or a global maximum in the second case, but that will generally require more effort to determine.

EXAMPLE 6.1.3 Let $f(x) = -x^2 + 4x - 3$. Find the maximum value of $f(x)$ on the interval $[0, 4]$. First note that $f'(x) = -2x+4 = 0$ when $x = 2$, and $f(2) = 1$. Next observe that $f'(x)$ is defined for all x, so there are no other critical values. Finally, $f(0) = -3$ and $f(4) = -3$. The largest value of $f(x)$ on the interval $[0, 4]$ is $f(2) = 1$. □

EXAMPLE 6.1.4 Let $f(x) = -x^2 + 4x - 3$. Find the maximum value of $f(x)$ on the interval $[-1, 1]$.

First note that $f'(x) = -2x + 4 = 0$ when $x = 2$. But $x = 2$ is not in the interval, so we don't use it. Thus the only two points to be checked are the endpoints; $f(-1) = -8$ and $f(1) = 0$. So the largest value of $f(x)$ on $[-1, 1]$ is $f(1) = 0$. □

EXAMPLE 6.1.5 Find the maximum and minimum values of the function $f(x) = 7 + |x - 2|$ for x between 1 and 4 inclusive. The derivative $f'(x)$ is never zero, but $f'(x)$ is undefined at $x = 2$, so we compute $f(2) = 7$. Checking the end points we get $f(1) = 8$ and $f(4) = 9$. The smallest of these numbers is $f(2) = 7$, which is, therefore, the minimum value of $f(x)$ on the interval $1 \le x \le 4$, and the maximum is $f(4) = 9$. □

EXAMPLE 6.1.6 Find all local maxima and minima for $f(x) = x^3 - x$, and determine whether there is a global maximum or minimum on the open interval $(-2, 2)$. In example 5.1.2 we found a local maximum at $(-\sqrt{3}/3, 2\sqrt{3}/9)$ and a local minimum at $(\sqrt{3}/3, -2\sqrt{3}/9)$. Since the endpoints are not in the interval $(-2, 2)$ they cannot be con-

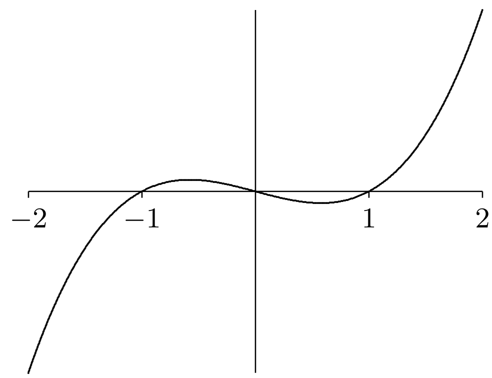


Figure 6.1.2 $f(x) = x^3 - x$

sidered. Is the lone local maximum a global maximum? Here we must look more closely at the graph. We know that on the closed interval $[-\sqrt{3}/3, \sqrt{3}/3]$ there is a global maximum at $x = -\sqrt{3}/3$ and a global minimum at $x = \sqrt{3}/3$. So the question becomes: what happens between -2 and $-\sqrt{3}/3$, and between $\sqrt{3}/3$ and 2? Since there is a local minimum at $x = \sqrt{3}/3$, the graph must continue up to the right, since there are no more critical values. This means no value of f will be less than $-2\sqrt{3}/9$ between $\sqrt{3}/3$ and 2, but it says nothing about whether we might find a value larger than the local maximum $2\sqrt{3}/9$. How can we tell? Since the function increases to the right of $\sqrt{3}/3$, we need to know what the function values do "close to" 2. Here the easiest test is to pick a number and do a computation to get some idea of what's going on. Since $f(1.9) = 4.959 > 2\sqrt{3}/9$, there is no global maximum at $-\sqrt{3}/3$, and hence no global maximum at all. (How can we tell that $4.959 > 2\sqrt{3}/9$? We can use a calculator to approximate the right hand side; if it is not even close to 4.959 we can take this as decisive. Since $2\sqrt{3}/9 \approx 0.3849$, there's really no question. Funny things can happen in the rounding done by computers and calculators, however, so we might be a little more careful, especially if the values come out quite close. In this case we can convert the relation $4.959 > 2\sqrt{3}/9$ into $(9/2)4.959 > \sqrt{3}$ and ask whether this is true. Since the left side is clearly larger than $4 \cdot 4$ which is clearly larger than $\sqrt{3}$, this settles the question.)

A similar analysis shows that there is also no global minimum. The graph of $f(x)$ on $(-2, 2)$ is shown in figure 6.1.2. □

EXAMPLE 6.1.7 Of all rectangles of area 100, which has the smallest perimeter?

First we must translate this into a purely mathematical problem in which we want to find the minimum value of a function. If x denotes one of the sides of the rectangle, then the adjacent side must be $100/x$ (in order that the area be 100). So the function we want to minimize is

$$f(x) = 2x + 2\frac{100}{x}$$

since the perimeter is twice the length plus twice the width of the rectangle. Not all values of x make sense in this problem: lengths of sides of rectangles must be positive, so $x > 0$. If $x > 0$ then so is $100/x$, so we need no second condition on x.

We next find $f'(x)$ and set it equal to zero: $0 = f'(x) = 2 - 200/x^2$. Solving $f'(x) = 0$ for x gives us $x = \pm 10$. We are interested only in $x > 0$, so only the value $x = 10$ is of interest. Since $f'(x)$ is defined everywhere on the interval $(0, \infty)$, there are no more critical values, and there are no endpoints. Is there a local maximum, minimum, or neither at $x = 10$? The second derivative is $f''(x) = 400/x^3$, and $f''(10) > 0$, so there is a local minimum. Since there is only one critical value, this is also the global minimum, so the rectangle with smallest perimeter is the 10×10 square. □

EXAMPLE 6.1.8 You want to sell a certain number n of items in order to maximize your profit. Market research tells you that if you set the price at \$1.50, you will be able to sell 5000 items, and for every 10 cents you lower the price below \$1.50 you will be able to sell another 1000 items. Suppose that your fixed costs ("start-up costs") total \$2000, and the per item cost of production ("marginal cost") is \$0.50. Find the price to set per item and the number of items sold in order to maximize profit, and also determine the maximum profit you can get.

The first step is to convert the problem into a function maximization problem. Since we want to maximize profit by setting the price per item, we should look for a function $P(x)$ representing the profit when the price per item is x. Profit is revenue minus costs, and revenue is number of items sold times the price per item, so we get $P = nx - 2000 - 0.50n$. The number of items sold is itself a function of x, $n = 5000 + 1000(1.5 - x)/0.10$, because $(1.5 - x)/0.10$ is the number of multiples of 10 cents that the price is below \$1.50. Now we substitute for n in the profit function:

$$\begin{aligned} P(x) &= (5000 + 1000(1.5 - x)/0.10)x - 2000 - 0.5(5000 + 1000(1.5 - x)/0.10) \\ &= -10000x^2 + 25000x - 12000 \end{aligned}$$

We want to know the maximum value of this function when x is between 0 and 1.5. The derivative is $P'(x) = -20000x + 25000$, which is zero when $x = 1.25$. Since $P''(x) = -20000 < 0$, there must be a local maximum at $x = 1.25$, and since this is the only critical value it must be a global maximum as well. (Alternately, we could compute $P(0) = -12000$, $P(1.25) = 3625$, and $P(1.5) = 3000$ and note that $P(1.25)$ is the maximum of these.) Thus the maximum profit is \$3625, attained when we set the price at \$1.25 and sell 7500 items. □

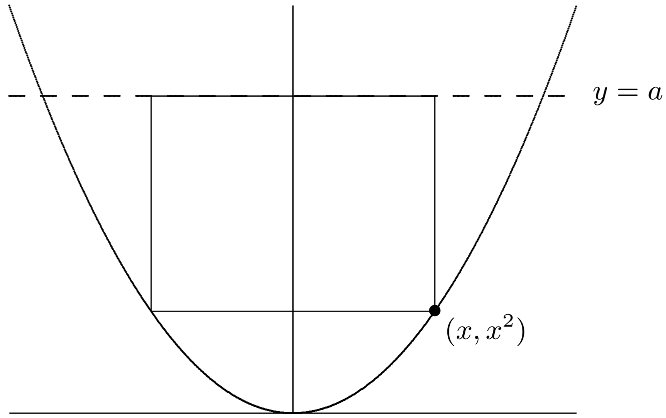


Figure 6.1.3 Rectangle in a parabola.

EXAMPLE 6.1.9 Find the largest rectangle (that is, the rectangle with largest area) that fits inside the graph of the parabola $y = x^2$ below the line $y = a$ (a is an unspecified constant value), with the top side of the rectangle on the horizontal line $y = a$; see figure 6.1.3.)

We want to find the maximum value of some function $A(x)$ representing area. Perhaps the hardest part of this problem is deciding what x should represent. The lower right corner of the rectangle is at (x, x^2), and once this is chosen the rectangle is completely determined. So we can let the x in $A(x)$ be the x of the parabola $f(x) = x^2$. Then the area is $A(x) = (2x)(a - x^2) = -2x^3 + 2ax$. We want the maximum value of $A(x)$ when x is in $[0, \sqrt{a}]$. (You might object to allowing $x = 0$ or $x = \sqrt{a}$, since then the "rectangle" has either no width or no height, so is not "really" a rectangle. But the problem is somewhat easier if we simply allow such rectangles, which have zero area.)

Setting $0 = A'(x) = -6x^2 + 2a$ we get $x = \sqrt{a/3}$ as the only critical value. Testing this and the two endpoints, we have $A(0) = A(\sqrt{a}) = 0$ and $A(\sqrt{a/3}) = (4/9)\sqrt{3}a^{3/2}$. The maximum area thus occurs when the rectangle has dimensions $2\sqrt{a/3} \times (2/3)a$. □

EXAMPLE 6.1.10 If you fit the largest possible cone inside a sphere, what fraction of the volume of the sphere is occupied by the cone? (Here by "cone" we mean a right circular cone, i.e., a cone for which the base is perpendicular to the axis of symmetry, and for which the cross-section cut perpendicular to the axis of symmetry at any point is a circle.)

Let R be the radius of the sphere, and let r and h be the base radius and height of the cone inside the sphere. What we want to maximize is the volume of the cone: $\pi r^2 h/3$. Here R is a fixed value, but r and h can vary. Namely, we could choose r to be as large as possible—equal to R—by taking the height equal to R; or we could make the cone's height h larger at the expense of making r a little less than R. See the cross-section depicted in

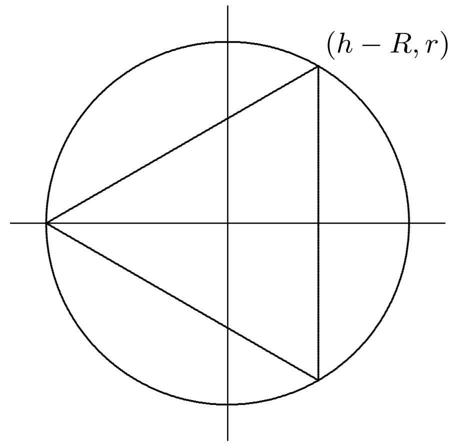


Figure 6.1.4 Cone in a sphere.

figure 6.1.4. We have situated the picture in a convenient way relative to the x and y axes, namely, with the center of the sphere at the origin and the vertex of the cone at the far left on the x-axis.

Notice that the function we want to maximize, $\pi r^2 h/3$, depends on *two* variables. This is frequently the case, but often the two variables are related in some way so that "really" there is only one variable. So our next step is to find the relationship and use it to solve for one of the variables in terms of the other, so as to have a function of only one variable to maximize. In this problem, the condition is apparent in the figure: the upper corner of the triangle, whose coordinates are $(h-R, r)$, must be on the circle of radius R. That is,

$$(h-R)^2 + r^2 = R^2.$$

We can solve for h in terms of r or for r in terms of h. Either involves taking a square root, but we notice that the volume function contains r^2, not r by itself, so it is easiest to solve for r^2 directly: $r^2 = R^2 - (h-R)^2$. Then we substitute the result into $\pi r^2 h/3$:

$$\begin{aligned} V(h) &= \pi(R^2 - (h-R)^2)h/3 \\ &= -\frac{\pi}{3}h^3 + \frac{2}{3}\pi h^2 R \end{aligned}$$

We want to maximize $V(h)$ when h is between 0 and $2R$. Now we solve $0 = f'(h) = -\pi h^2 + (4/3)\pi h R$, getting $h = 0$ or $h = 4R/3$. We compute $V(0) = V(2R) = 0$ and $V(4R/3) = (32/81)\pi R^3$. The maximum is the latter; since the volume of the sphere is $(4/3)\pi R^3$, the fraction of the sphere occupied by the cone is

$$\frac{(32/81)\pi R^3}{(4/3)\pi R^3} = \frac{8}{27} \approx 30\%.$$

□

EXAMPLE 6.1.11 You are making cylindrical containers to contain a given volume. Suppose that the top and bottom are made of a material that is N times as expensive (cost per unit area) as the material used for the lateral side of the cylinder. Find (in terms of N) the ratio of height to base radius of the cylinder that minimizes the cost of making the containers.

Let us first choose letters to represent various things: h for the height, r for the base radius, V for the volume of the cylinder, and c for the cost per unit area of the lateral side of the cylinder; V and c are constants, h and r are variables. Now we can write the cost of materials:

$$c(2\pi rh) + Nc(2\pi r^2).$$

Again we have two variables; the relationship is provided by the fixed volume of the cylinder: $V = \pi r^2 h$. We use this relationship to eliminate h (we could eliminate r, but it's a little easier if we eliminate h, which appears in only one place in the above formula for cost). The result is

$$f(r) = 2c\pi r\frac{V}{\pi r^2} + 2Nc\pi r^2 = \frac{2cV}{r} + 2Nc\pi r^2.$$

We want to know the minimum value of this function when r is in $(0,\infty)$. We now set $0 = f'(r) = -2cV/r^2 + 4Nc\pi r$, giving $r = \sqrt[3]{V/(2N\pi)}$. Since $f''(r) = 4cV/r^3 + 4Nc\pi$ is positive when r is positive, there is a local minimum at the critical value, and hence a global minimum since there is only one critical value.

Finally, since $h = V/(\pi r^2)$,

$$\frac{h}{r} = \frac{V}{\pi r^3} = \frac{V}{\pi(V/(2N\pi))} = 2N,$$

so the minimum cost occurs when the height h is $2N$ times the radius. If, for example, there is no difference in the cost of materials, the height is twice the radius (or the height is equal to the diameter). □

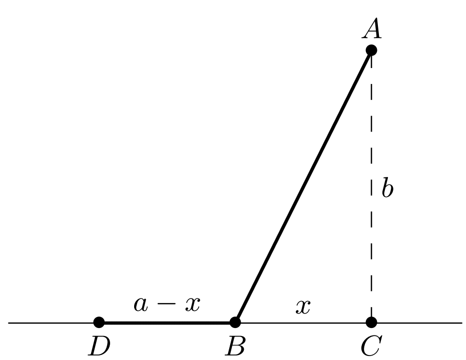


Figure 6.1.5 Minimizing travel time.

EXAMPLE 6.1.12 Suppose you want to reach a point A that is located across the sand from a nearby road (see figure 6.1.5). Suppose that the road is straight, and b is the distance from A to the closest point C on the road. Let v be your speed on the road, and let w, which is less than v, be your speed on the sand. Right now you are at the point D, which is a distance a from C. At what point B should you turn off the road and head across the sand in order to minimize your travel time to A?

Let x be the distance short of C where you turn off, i.e., the distance from B to C. We want to minimize the total travel time. Recall that when traveling at constant velocity, time is distance divided by velocity.

You travel the distance $\overline{DB}$ at speed v, and then the distance $\overline{BA}$ at speed w. Since $\overline{DB} = a - x$ and, by the Pythagorean theorem, $\overline{BA} = \sqrt{x^2 + b^2}$, the total time for the trip is

$$f(x) = \frac{a-x}{v} + \frac{\sqrt{x^2+b^2}}{w}.$$

We want to find the minimum value of f when x is between 0 and a. As usual we set $f'(x) = 0$ and solve for x:

$$\begin{aligned}
0 = f'(x) &= -\frac{1}{v} + \frac{x}{w\sqrt{x^2+b^2}} \\
w\sqrt{x^2+b^2} &= vx \\
w^2(x^2+b^2) &= v^2x^2 \\
w^2b^2 &= (v^2-w^2)x^2 \\
x &= \frac{wb}{\sqrt{v^2-w^2}}
\end{aligned}$$

Notice that a does not appear in the last expression, but a is not irrelevant, since we are interested only in critical values that are in $[0, a]$, and $wb/\sqrt{v^2 - w^2}$ is either in this interval or not. If it is, we can use the second derivative to test it:

$$f''(x) = \frac{b^2}{(x^2+b^2)^{3/2}w}.$$

Since this is always positive there is a local minimum at the critical point, and so it is a global minimum as well.

If the critical value is not in $[0, a]$ it is larger than a. In this case the minimum must occur at one of the endpoints. We can compute

$$\begin{aligned}
f(0) &= \frac{a}{v} + \frac{b}{w} \\
f(a) &= \frac{\sqrt{a^2+b^2}}{w}
\end{aligned}$$

but it is difficult to determine which of these is smaller by direct comparison. If, as is likely in practice, we know the values of v, w, a, and b, then it is easy to determine this. With a little cleverness, however, we can determine the minimum in general. We have seen that $f''(x)$ is always positive, so the derivative $f'(x)$ is always increasing. We know that at $wb/\sqrt{v^2 - w^2}$ the derivative is zero, so for values of x less than that critical value, the derivative is negative. This means that $f(0) > f(a)$, so the minimum occurs when $x = a$.

So the upshot is this: If you start farther away from C than $wb/\sqrt{v^2 - w^2}$ then you always want to cut across the sand when you are a distance $wb/\sqrt{v^2 - w^2}$ from point C. If you start closer than this to C, you should cut directly across the sand. □

Summary—Steps to solve an optimization problem.

1. Decide what the variables are and what the constants are, draw a diagram if appropriate, understand clearly what it is that is to be maximized or minimized.
2. Write a formula for the function for which you wish to find the maximum or minimum.
3. Express that formula in terms of only one variable, that is, in the form $f(x)$.
4. Set $f'(x) = 0$ and solve. Check all critical values and endpoints to determine the extreme value.

Exercises 6.1.

1. Let $f(x) = \begin{cases} 1 + 4x - x^2 & \text{for } x \le 3 \\ (x + 5)/2 & \text{for } x > 3 \end{cases}$

Find the maximum value and minimum values of $f(x)$ for x in $[0, 4]$. Graph $f(x)$ to check your answers. ⇒

2. Find the dimensions of the rectangle of largest area having fixed perimeter 100. ⇒

3. Find the dimensions of the rectangle of largest area having fixed perimeter P. ⇒

4. A box with square base and no top is to hold a volume 100. Find the dimensions of the box that requires the least material for the five sides. Also find the ratio of height to side of the base. ⇒

5. A box with square base is to hold a volume 200. The bottom and top are formed by folding in flaps from all four sides, so that the bottom and top consist of two layers of cardboard. Find the dimensions of the box that requires the least material. Also find the ratio of height to side of the base. ⇒

6. A box with square base and no top is to hold a volume V. Find (in terms of V) the dimensions of the box that requires the least material for the five sides. Also find the ratio of height to side of the base. (This ratio will not involve V.) ⇒

7. You have 100 feet of fence to make a rectangular play area alongside the wall of your house. The wall of the house bounds one side. What is the largest size possible (in square feet) for the play area? ⇒

8. You have l feet of fence to make a rectangular play area alongside the wall of your house. The wall of the house bounds one side. What is the largest size possible (in square feet) for the play area? ⇒

9. Marketing tells you that if you set the price of an item at \$10 then you will be unable to sell it, but that you can sell 500 items for each dollar below \$10 that you set the price. Suppose your fixed costs total \$3000, and your marginal cost is \$2 per item. What is the most profit you can make? ⇒

10. Find the area of the largest rectangle that fits inside a semicircle of radius 10 (one side of the rectangle is along the diameter of the semicircle). ⇒

11. Find the area of the largest rectangle that fits inside a semicircle of radius r (one side of the rectangle is along the diameter of the semicircle). ⇒

12. For a cylinder with surface area 50, including the top and the bottom, find the ratio of height to base radius that maximizes the volume. ⇒

13. For a cylinder with given surface area S, including the top and the bottom, find the ratio of height to base radius that maximizes the volume. ⇒

14. You want to make cylindrical containers to hold 1 liter (1000 cubic centimeters) using the least amount of construction material. The side is made from a rectangular piece of material, and this can be done with no material wasted. However, the top and bottom are cut from squares of side $2r$, so that $2(2r)^2 = 8r^2$ of material is needed (rather than $2\pi r^2$, which is the total area of the top and bottom). Find the dimensions of the container using the least amount of material, and also find the ratio of height to radius for this container. ⇒

15. You want to make cylindrical containers of a given volume V using the least amount of construction material. The side is made from a rectangular piece of material, and this can be done with no material wasted. However, the top and bottom are cut from squares of side $2r$, so that $2(2r)^2 = 8r^2$ of material is needed (rather than $2\pi r^2$, which is the total area of the top and bottom). Find the optimal ratio of height to radius. ⇒

16. Given a right circular cone, you put an upside-down cone inside it so that its vertex is at the center of the base of the larger cone and its base is parallel to the base of the larger cone. If you choose the upside-down cone to have the largest possible volume, what fraction of the volume of the larger cone does it occupy? (Let H and R be the height and base radius of the larger cone, and let h and r be the height and base radius of the smaller cone. Hint: Use similar triangles to get an equation relating h and r.) ⇒

17. In example 6.1.12, what happens if $w \ge v$ (i.e., your speed on sand is at least your speed on the road)? ⇒

18. A container holding a fixed volume is being made in the shape of a cylinder with a hemispherical top. (The hemispherical top has the same radius as the cylinder.) Find the ratio of height to radius of the cylinder which minimizes the cost of the container if (a) the cost per unit area of the top is twice as great as the cost per unit area of the side, and the container is made with no bottom; (b) the same as in (a), except that the container is made with a circular bottom, for which the cost per unit area is 1.5 times the cost per unit area of the side. ⇒

19. A piece of cardboard is 1 meter by 1/2 meter. A square is to be cut from each corner and the sides folded up to make an open-top box. What are the dimensions of the box with maximum possible volume? ⇒

20. (a) A square piece of cardboard of side a is used to make an open-top box by cutting out a small square from each corner and bending up the sides. How large a square should be cut from each corner in order that the box have maximum volume? (b) What if the piece of cardboard used to make the box is a rectangle of sides a and b? ⇒

21. A window consists of a rectangular piece of clear glass with a semicircular piece of colored glass on top; the colored glass transmits only 1/2 as much light per unit area as the the clear glass. If the distance from top to bottom (across both the rectangle and the semicircle) is 2 meters and the window may be no more than 1.5 meters wide, find the dimensions of the rectangular portion of the window that lets through the most light. ⇒

22. A window consists of a rectangular piece of clear glass with a semicircular piece of colored glass on top. Suppose that the colored glass transmits only k times as much light per unit area as the clear glass (k is between 0 and 1). If the distance from top to bottom (across both the rectangle and the semicircle) is a fixed distance H, find (in terms of k) the ratio of vertical side to horizontal side of the rectangle for which the window lets through the most light. ⇒

23. You are designing a poster to contain a fixed amount A of printing (measured in square centimeters) and have margins of a centimeters at the top and bottom and b centimeters at the sides. Find the ratio of vertical dimension to horizontal dimension of the printed area on the poster if you want to minimize the amount of posterboard needed. ⇒

24. The strength of a rectangular beam is proportional to the product of its width w times the square of its depth d. Find the dimensions of the strongest beam that can be cut from a cylindrical log of radius r. ⇒

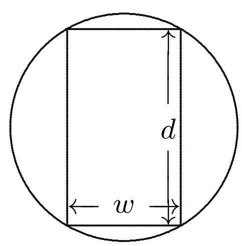


Figure 6.1.6 Cutting a beam.

25. What fraction of the volume of a sphere is taken up by the largest cylinder that can be fit inside the sphere? ⇒

26. The U.S. post office will accept a box for shipment only if the sum of the length and girth (distance around) is at most 108 in. Find the dimensions of the largest acceptable box with square front and back. ⇒

27. Find the dimensions of the lightest cylindrical can containing 0.25 liter (=250 cm^3) if the top and bottom are made of a material that is twice as heavy (per unit area) as the material used for the side. ⇒

28. A conical paper cup is to hold 1/4 of a liter. Find the height and radius of the cone which minimizes the amount of paper needed to make the cup. Use the formula $\pi r\sqrt{r^2+h^2}$ for the area of the side of a cone. ⇒

29. A conical paper cup is to hold a fixed volume of water. Find the ratio of height to base radius of the cone which minimizes the amount of paper needed to make the cup. Use the formula $\pi r\sqrt{r^2+h^2}$ for the area of the side of a cone, called the **lateral area** of the cone. ⇒

30. If you fit the cone with the largest possible surface area (lateral area plus area of base) into a sphere, what percent of the volume of the sphere is occupied by the cone? ⇒

31. Two electrical charges, one a positive charge A of magnitude a and the other a negative charge B of magnitude b, are located a distance c apart. A positively charged particle P is situated on the line between A and B. Find where P should be put so that the pull away from A towards B is minimal. Here assume that the force from each charge is proportional to the strength of the source and inversely proportional to the square of the distance from the source. ⇒

32. Find the fraction of the area of a triangle that is occupied by the largest rectangle that can be drawn in the triangle (with one of its sides along a side of the triangle). Show that this fraction does not depend on the dimensions of the given triangle. ⇒

33. How are your answers to Problem 9 affected if the cost per item for the x items, instead of being simply \$2, decreases below \$2 in proportion to x (because of economy of scale and volume discounts) by 1 cent for each 25 items produced? ⇒

34. You are standing near the side of a large wading pool of uniform depth when you see a child in trouble. You can run at a speed v_1 on land and at a slower speed v_2 in the water. Your perpendicular distance from the side of the pool is a, the child's perpendicular distance is b, and the distance along the side of the pool between the closest point to you and the closest point to the child is c (see the figure below). Without stopping to do any calculus, you instinctively choose the quickest route (shown in the figure) and save the child. Our purpose is to derive a relation between the angle θ_1 your path makes with the perpendicular to the side of the pool when you're on land, and the angle θ_2 your path makes with the perpendicular when you're in the water. To do this, let x be the distance between the closest point to you at the side of the pool and the point where you enter the water. Write the total time you run (on land and in the water) in terms of x (and also the constants a, b, c, v_1, v_2). Then set the derivative equal to zero. The result, called "Snell's law" or the "law of refraction," also governs the bending of light when it goes into water. ⇒

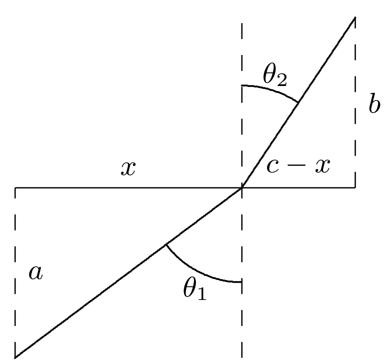


Figure 6.1.7 Wading pool rescue.

6.2 Related Rates

Suppose we have two variables x and y (in most problems the letters will be different, but for now let's use x and y) which are both changing with time. A "related rates" problem is a problem in which we know one of the rates of change at a given instant—say,

$\dot{x} = dx/dt$—and we want to find the other rate $\dot{y} = dy/dt$ at that instant. (The use of $\dot{x}$ to mean dx/dt goes back to Newton and is still used for this purpose, especially by physicists.)

If y is written in terms of x, i.e., $y = f(x)$, then this is easy to do using the chain rule:

$$\dot{y} = \frac{dy}{dt} = \frac{dy}{dx} \cdot \frac{dx}{dt} = \frac{dy}{dx}\dot{x}.$$

That is, find the derivative of $f(x)$, plug in the value of x at the instant in question, and multiply by the given value of $\dot{x} = dx/dt$ to get $\dot{y} = dy/dt$.

EXAMPLE 6.2.1 Suppose an object is moving along a path described by $y = x^2$, that is, it is moving on a parabolic path. At a particular time, say $t = 5$, the x coordinate is 6 and we measure the speed at which the x coordinate of the object is changing and find that $dx/dt = 3$. At the same time, how fast is the y coordinate changing?

Using the chain rule, $dy/dt = 2x \cdot dx/dt$. At $t = 5$ we know that $x = 6$ and $dx/dt = 3$, so $dy/dt = 2 \cdot 6 \cdot 3 = 36$. □

In many cases, particularly interesting ones, x and y will be related in some other way, for example $x = f(y)$, or $F(x, y) = k$, or perhaps $F(x, y) = G(x, y)$, where $F(x, y)$ and $G(x, y)$ are expressions involving both variables. In all cases, you can solve the related rates problem by taking the derivative of both sides, plugging in all the known values (namely, x, y, and $\dot{x}$), and then solving for $\dot{y}$.

To summarize, here are the steps in doing a related rates problem:

1. Decide what the two variables are.
2. Find an equation relating them.
3. Take d/dt of both sides.
4. Plug in all known values at the instant in question.
5. Solve for the unknown rate.

EXAMPLE 6.2.2 A plane is flying directly away from you at 500 mph at an altitude of 3 miles. How fast is the plane's distance from you increasing at the moment when the plane is flying over a point on the ground 4 miles from you?

To see what's going on, we first draw a schematic representation of the situation, as in figure 6.2.1.

Because the plane is in level flight directly away from you, the rate at which x changes is the speed of the plane, $dx/dt = 500$. The distance between you and the plane is y; it is dy/dt that we wish to know. By the Pythagorean Theorem we know that $x^2 + 9 = y^2$.

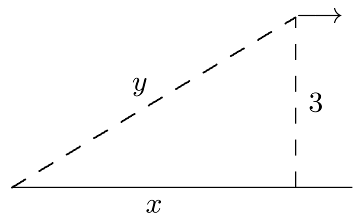


Figure 6.2.1 Receding airplane.

Taking the derivative:

$$2x\dot{x} = 2y\dot{y}.$$

We are interested in the time at which $x = 4$; at this time we know that $4^2 + 9 = y^2$, so $y = 5$. Putting together all the information we get

$$2(4)(500) = 2(5)\dot{y}.$$

Thus, $\dot{y} = 400$ mph. □

EXAMPLE 6.2.3 You are inflating a spherical balloon at the rate of 7 cm^3/sec. How fast is its radius increasing when the radius is 4 cm?

Here the variables are the radius r and the volume V. We know dV/dt, and we want dr/dt. The two variables are related by means of the equation $V = 4\pi r^3/3$. Taking the derivative of both sides gives $dV/dt = 4\pi r^2\dot{r}$. We now substitute the values we know at the instant in question: $7 = 4\pi 4^2\dot{r}$, so $\dot{r} = 7/(64\pi)$ cm/sec. □

EXAMPLE 6.2.4 Water is poured into a conical container at the rate of 10 cm^3/sec. The cone points directly down, and it has a height of 30 cm and a base radius of 10 cm; see figure 6.2.2. How fast is the water level rising when the water is 4 cm deep (at its deepest point)?

The water forms a conical shape within the big cone; its height and base radius and volume are all increasing as water is poured into the container. This means that we actually have three things varying with time: the water level h (the height of the cone of water), the radius r of the circular top surface of water (the base radius of the cone of water), and the volume of water V. The volume of a cone is given by $V = \pi r^2 h/3$. We know dV/dt, and we want dh/dt. At first something seems to be wrong: we have a third variable r whose rate we don't know.

But the dimensions of the cone of water must have the same proportions as those of the container. That is, because of similar triangles, $r/h = 10/30$ so $r = h/3$. Now we can eliminate r from the problem entirely: $V = \pi(h/3)^2 h/3 = \pi h^3/27$. We take the derivative of both sides and plug in $h = 4$ and $dV/dt = 10$, obtaining $10 = (3\pi \cdot 4^2/27)(dh/dt)$. Thus, $dh/dt = 90/(16\pi)$ cm/sec. □

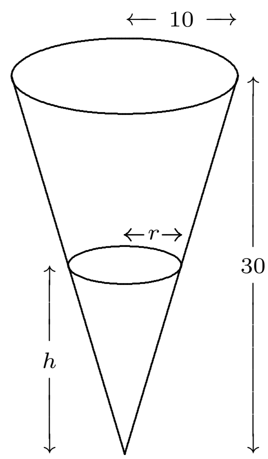


Figure 6.2.2 Conical water tank.

EXAMPLE 6.2.5 A swing consists of a board at the end of a 10 ft long rope. Think of the board as a point P at the end of the rope, and let Q be the point of attachment at the other end. Suppose that the swing is directly below Q at time $t = 0$, and is being pushed by someone who walks at 6 ft/sec from left to right. Find (a) how fast the swing is rising after 1 sec; (b) the angular speed of the rope in deg/sec after 1 sec.

We start out by asking: What is the geometric quantity whose rate of change we know, and what is the geometric quantity whose rate of change we're being asked about? Note that the person pushing the swing is moving horizontally at a rate we know. In other words, the horizontal coordinate of P is increasing at 6 ft/sec. In the xy-plane let us make the convenient choice of putting the origin at the location of P at time $t = 0$, i.e., a distance 10 directly below the point of attachment. Then the rate we know is dx/dt, and in part (a) the rate we want is dy/dt (the rate at which P is rising). In part (b) the rate we want is $\dot{\theta} = d\theta/dt$, where θ stands for the angle in radians through which the swing has swung from the vertical. (Actually, since we want our answer in deg/sec, at the end we must convert $d\theta/dt$ from rad/sec by multiplying by $180/\pi$.)

(a) From the diagram we see that we have a right triangle whose legs are x and $10 - y$, and whose hypotenuse is 10. Hence $x^2 + (10 - y)^2 = 100$. Taking the derivative of both sides we obtain: $2x\dot{x} + 2(10 - y)(0 - \dot{y}) = 0$. We now look at what we know after 1 second, namely $x = 6$ (because x started at 0 and has been increasing at the rate of 6 ft/sec for 1 sec), $y = 2$ (because we get $10 - y = 8$ from the Pythagorean theorem applied to the triangle with hypotenuse 10 and leg 6), and $\dot{x} = 6$. Putting in these values gives us $2 \cdot 6 \cdot 6 - 2 \cdot 8\dot{y} = 0$, from which we can easily solve for $\dot{y}$: $\dot{y} = 4.5$ ft/sec.

(b) Here our two variables are x and θ, so we want to use the same right triangle as in part (a), but this time relate θ to x. Since the hypotenuse is constant (equal to 10), the best way to do this is to use the sine: $\sin\theta = x/10$. Taking derivatives we obtain

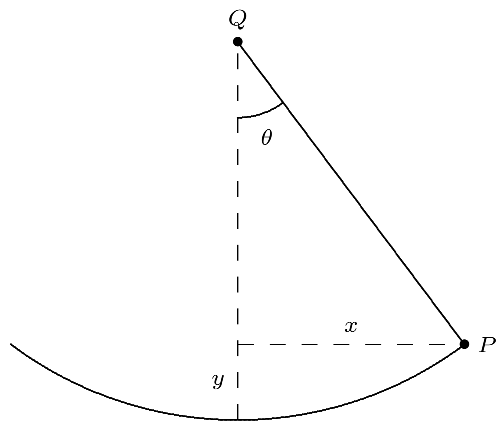


Figure 6.2.3 Swing.

$(\cos\theta)\dot{\theta} = 0.1\dot{x}$. At the instant in question ($t = 1$ sec), when we have a right triangle with sides 6–8–10, $\cos\theta = 8/10$ and $\dot{x} = 6$. Thus $(8/10)\dot{\theta} = 6/10$, i.e., $\dot{\theta} = 6/8 = 3/4$ rad/sec, or approximately 43 deg/sec. □

We have seen that sometimes there are apparently more than two variables that change with time, but in reality there are just two, as the others can be expressed in terms of just two. But sometimes there really are several variables that change with time; as long as you know the rates of change of all but one of them you can find the rate of change of the remaining one. As in the case when there are just two variables, take the derivative of both sides of the equation relating all of the variables, and then substitute all of the known values and solve for the unknown rate.

EXAMPLE 6.2.6 A road running north to south crosses a road going east to west at the point P. Car A is driving north along the first road, and car B is driving east along the second road. At a particular time car A is 10 kilometers to the north of P and traveling at 80 km/hr, while car B is 15 kilometers to the east of P and traveling at 100 km/hr. How fast is the distance between the two cars changing?

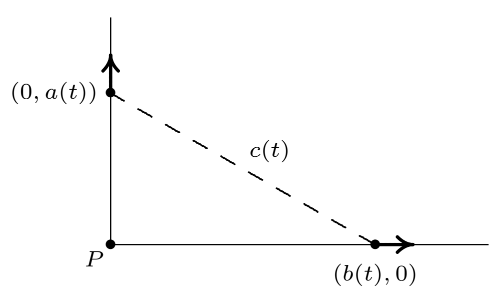


Figure 6.2.4 Cars moving apart.

Let $a(t)$ be the distance of car A north of P at time t, and $b(t)$ the distance of car B east of P at time t, and let $c(t)$ be the distance from car A to car B at time t. By the Pythagorean Theorem, $c(t)^2 = a(t)^2 + b(t)^2$. Taking derivatives we get $2c(t)c'(t) = 2a(t)a'(t) + 2b(t)b'(t)$, so

$$\dot{c} = \frac{a\dot{a} + b\dot{b}}{c} = \frac{a\dot{a} + b\dot{b}}{\sqrt{a^2 + b^2}}.$$

Substituting known values we get:

$$\dot{c} = \frac{10 \cdot 80 + 15 \cdot 100}{\sqrt{10^2 + 15^2}} = \frac{460}{\sqrt{13}} \approx 127.6 \text{km/hr}$$

at the time of interest. □

Notice how this problem differs from example 6.2.2. In both cases we started with the Pythagorean Theorem and took derivatives on both sides. However, in example 6.2.2 one of the sides was a constant (the altitude of the plane), and so the derivative of the square of that side of the triangle was simply zero. In this example, on the other hand, all three sides of the right triangle are variables, even though we are interested in a specific value of each side of the triangle (namely, when the sides have lengths 10 and 15). Make sure that you understand at the start of the problem what are the variables and what are the constants.

Exercises 6.2.

1. A cylindrical tank standing upright (with one circular base on the ground) has radius 20 cm. How fast does the water level in the tank drop when the water is being drained at 25 cm^3/sec? ⇒
2. A cylindrical tank standing upright (with one circular base on the ground) has radius 1 meter. How fast does the water level in the tank drop when the water is being drained at 3 liters per second? ⇒
3. A ladder 13 meters long rests on horizontal ground and leans against a vertical wall. The foot of the ladder is pulled away from the wall at the rate of 0.6 m/sec. How fast is the top sliding down the wall when the foot of the ladder is 5 m from the wall? ⇒
4. A ladder 13 meters long rests on horizontal ground and leans against a vertical wall. The top of the ladder is being pulled up the wall at 0.1 meters per second. How fast is the foot of the ladder approaching the wall when the foot of the ladder is 5 m from the wall? ⇒
5. A rotating beacon is located 2 miles out in the water. Let A be the point on the shore that is closest to the beacon. As the beacon rotates at 10 rev/min, the beam of light sweeps down the shore once each time it revolves. Assume that the shore is straight. How fast is the point where the beam hits the shore moving at an instant when the beam is lighting up a point 2 miles along the shore from the point A? ⇒
6. A baseball diamond is a square 90 ft on a side. A player runs from first base to second base at 15 ft/sec. At what rate is the player's distance from third base decreasing when she is half way from first to second base? ⇒

7. Sand is poured onto a surface at 15 cm^3/sec, forming a conical pile whose base diameter is always equal to its altitude. How fast is the altitude of the pile increasing when the pile is 3 cm high? ⇒

8. A boat is pulled in to a dock by a rope with one end attached to the front of the boat and the other end passing through a ring attached to the dock at a point 5 ft higher than the front of the boat. The rope is being pulled through the ring at the rate of 0.6 ft/sec. How fast is the boat approaching the dock when 13 ft of rope are out? ⇒

9. A balloon is at a height of 50 meters, and is rising at the constant rate of 5 m/sec. A bicyclist passes beneath it, traveling in a straight line at the constant speed of 10 m/sec. How fast is the distance between the bicyclist and the balloon increasing 2 seconds later? ⇒

10. A pyramid-shaped vat has square cross-section and stands on its tip. The dimensions at the top are 2 m × 2 m, and the depth is 5 m. If water is flowing into the vat at 3 m^3/min, how fast is the water level rising when the depth of water (at the deepest point) is 4 m? Note: the volume of any "conical" shape (including pyramids) is (1/3)(height)(area of base). ⇒

11. The sun is rising at the rate of 1/4 deg/min, and appears to be climbing into the sky perpendicular to the horizon, as depicted in figure 6.2.5. How fast is the shadow of a 200 meter building shrinking at the moment when the shadow is 500 meters long? ⇒

12. The sun is setting at the rate of 1/4 deg/min, and appears to be dropping perpendicular to the horizon, as depicted in figure 6.2.5. How fast is the shadow of a 25 meter wall lengthening at the moment when the shadow is 50 meters long? ⇒

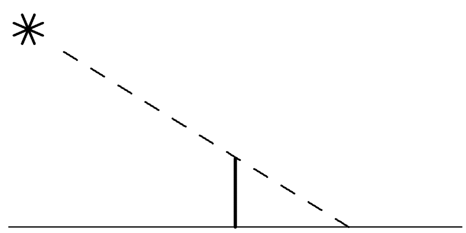


Figure 6.2.5 Sunrise or sunset.

13. The trough shown in figure 6.2.6 is constructed by fastening together three slabs of wood of dimensions 10 ft × 1 ft, and then attaching the construction to a wooden wall at each end. The angle θ was originally 30°, but because of poor construction the sides are collapsing. The trough is full of water. At what rate (in ft^3/sec) is the water spilling out over the top of the trough if the sides have each fallen to an angle of 45°, and are collapsing at the rate of 1° per second? ⇒

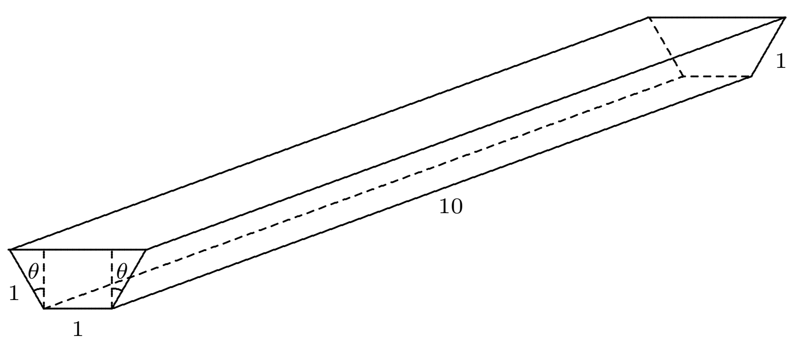


Figure 6.2.6 Trough.

14. A woman 5 ft tall walks at the rate of 3.5 ft/sec away from a streetlight that is 12 ft above the ground. At what rate is the tip of her shadow moving? At what rate is her shadow lengthening? ⇒

15. A man 1.8 meters tall walks at the rate of 1 meter per second toward a streetlight that is 4 meters above the ground. At what rate is the tip of his shadow moving? At what rate is his shadow shortening? ⇒

16. A police helicopter is flying at 150 mph at a constant altitude of 0.5 mile above a straight road. The pilot uses radar to determine that an oncoming car is at a distance of exactly 1 mile from the helicopter, and that this distance is decreasing at 190 mph. Find the speed of the car. ⇒

17. A police helicopter is flying at 200 kilometers per hour at a constant altitude of 1 km above a straight road. The pilot uses radar to determine that an oncoming car is at a distance of exactly 2 kilometers from the helicopter, and that this distance is decreasing at 250 kph. Find the speed of the car. ⇒

18. A light shines from the top of a pole 20 m high. A ball is falling 10 meters from the pole, casting a shadow on a building 30 meters away, as shown in figure 6.2.7. When the ball is 25 meters from the ground it is falling at 6 meters per second. How fast is its shadow moving? ⇒

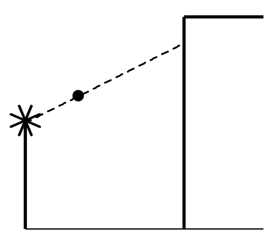


Figure 6.2.7 Falling ball.

19. Do example 6.2.6 assuming that the angle between the two roads is 120° instead of 90° (that is, the "north–south" road actually goes in a somewhat northwesterly direction from P). Recall the law of cosines: $c^2 = a^2 + b^2 - 2ab\cos\theta$. ⇒

20. Do example 6.2.6 assuming that car A is 300 meters north of P, car B is 400 meters east of P, both cars are going at constant speed toward P, and the two cars will collide in 10 seconds. ⇒

21. Do example 6.2.6 assuming that 8 seconds ago car A started from rest at P and has been picking up speed at the steady rate of 5 m/sec^2, and 6 seconds after car A started car B passed P moving east at constant speed 60 m/sec. ⇒

22. Referring again to example 6.2.6, suppose that instead of car B an airplane is flying at speed 200 km/hr to the east of P at an altitude of 2 km, as depicted in figure 6.2.8. How fast is the distance between car and airplane changing? ⇒

23. Referring again to example 6.2.6, suppose that instead of car B an airplane is flying at speed 200 km/hr to the east of P at an altitude of 2 km, and that it is gaining altitude at 10 km/hr. How fast is the distance between car and airplane changing? ⇒

24. A light shines from the top of a pole 20 m high. An object is dropped from the same height from a point 10 m away, so that its height at time t seconds is $h(t) = 20 - 9.8t^2/2$. How fast is the object's shadow moving on the ground one second later? ⇒

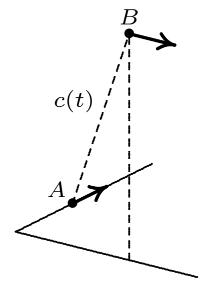


Figure 6.2.8 Car and airplane.

25. The two blades of a pair of scissors are fastened at the point A as shown in figure 6.2.9. Let a denote the distance from A to the tip of the blade (the point B). Let β denote the angle at the tip of the blade that is formed by the line $\overline{AB}$ and the bottom edge of the blade, line $\overline{BC}$, and let θ denote the angle between $\overline{AB}$ and the horizontal. Suppose that a piece of paper is cut in such a way that the center of the scissors at A is fixed, and the paper is also fixed. As the blades are closed (i.e., the angle θ in the diagram is decreased), the distance x between A and C increases, cutting the paper.

- **a.** Express x in terms of a, θ, and β.
- **b.** Express dx/dt in terms of a, θ, β, and $d\theta/dt$.
- **c.** Suppose that the distance a is 20 cm, and the angle β is 5°. Further suppose that θ is decreasing at 50 deg/sec. At the instant when $\theta = 30°$, find the rate (in cm/sec) at which the paper is being cut. ⇒

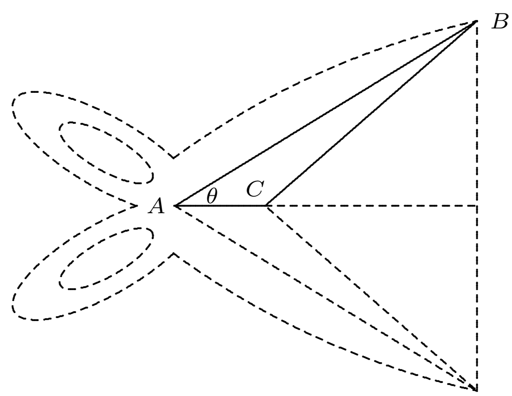


Figure 6.2.9 Scissors.

6.3 Newton's Method

Suppose you have a function $f(x)$, and you want to find as accurately as possible where it crosses the x-axis; in other words, you want to solve $f(x) = 0$. Suppose you know of no way to find an exact solution by any algebraic procedure, but you are able to use an approximation, provided it can be made quite close to the true value. Newton's method is a way to find a solution to the equation to as many decimal places as you want. It is what is called an "iterative procedure," meaning that it can be repeated again and again to get an answer of greater and greater accuracy. Iterative procedures like Newton's method are well suited to programming for a computer. Newton's method uses the fact that the tangent line to a curve is a good approximation to the curve near the point of tangency.

EXAMPLE 6.3.1 Approximate $\sqrt{3}$. Since $\sqrt{3}$ is a solution to $x^2 = 3$ or $x^2 - 3 = 0$, we use $f(x) = x^2 - 3$. We start by guessing something reasonably close to the true value; this is usually easy to do; let's use $\sqrt{3} \approx 2$. Now use the tangent line to the curve when $x = 2$ as an approximation to the curve, as shown in figure 6.3.1. Since $f'(x) = 2x$, the slope of this tangent line is 4 and its equation is $y = 4x - 7$. The tangent line is quite close to $f(x)$, so it crosses the x-axis near the point at which $f(x)$ crosses, that is, near $\sqrt{3}$. It is easy to find where the tangent line crosses the x-axis: solve $0 = 4x - 7$ to get $x = 7/4 = 1.75$. This is certainly a better approximation than 2, but let us say not close enough. We can improve it by doing the same thing again: find the tangent line at $x = 1.75$, find where this new tangent line crosses the x-axis, and use that value as a better approximation. We can continue this indefinitely, though it gets a bit tedious. Lets see if we can shortcut the process. Suppose the best approximation to the intercept we have so far is x_i. To find a better approximation we will always do the same thing: find the slope of the tangent line at x_i, find the equation of the tangent line, find the x-intercept. The slope is $2x_i$. The tangent line is $y = (2x_i)(x - x_i) + (x_i^2 - 3)$, using the point-slope formula for a line. Finally, the intercept is found by solving $0 = (2x_i)(x - x_i) + (x_i^2 - 3)$. With a little algebra this turns into $x = (x_i^2 + 3)/(2x_i)$; this is the next approximation, which we naturally call x_{i+1}. Instead of doing the whole tangent line computation every time we can simply use this formula to get as many approximations as we want. Starting with $x_0 = 2$, we get $x_1 = (x_0^2 + 3)/(2x_0) = (2^2 + 3)/4 = 7/4$ (the same approximation we got above, of course), $x_2 = (x_1^2 + 3)/(2x_1) = ((7/4)^2 + 3)/(7/2) = 97/56 \approx 1.73214$, $x_3 \approx 1.73205$, and so on. This is still a bit tedious by hand, but with a calculator or, even better, a good computer program, it is quite easy to get many, many approximations. We might guess already that 1.73205 is accurate to two decimal places, and in fact it turns out that it is accurate to 5 places. □

Let's think about this process in more general terms. We want to approximate a solution to $f(x) = 0$. We start with a rough guess, which we call x_0. We use the tangent line to $f(x)$ to get a new approximation that we hope will be closer to the true value. What is the equation of the tangent line when $x = x_0$? The slope is $f'(x_0)$ and the line goes through $(x_0, f(x_0))$, so the equation of the line is

$$y = f'(x_0)(x - x_0) + f(x_0).$$

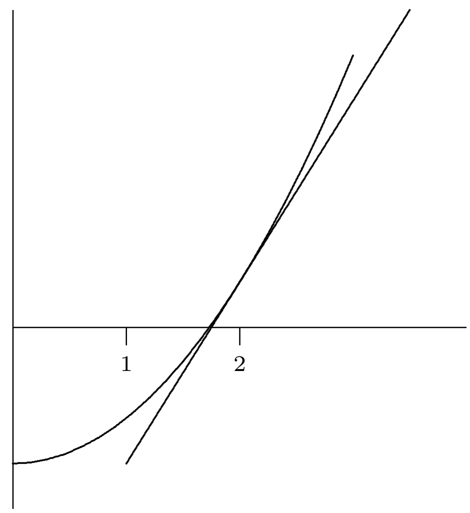


Figure 6.3.1 Newton's method. (AP)

Now we find where this crosses the x-axis by substituting $y = 0$ and solving for x:

$$x = \frac{x_0 f'(x_0) - f(x_0)}{f'(x_0)} = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

We will typically want to compute more than one of these improved approximations, so we number them consecutively; from x_0 we have computed x_1:

$$x_1 = \frac{x_0 f'(x_0) - f(x_0)}{f'(x_0)} = x_0 - \frac{f(x_0)}{f'(x_0)},$$

and in general from x_i we compute x_{i+1}:

$$x_{i+1} = \frac{x_i f'(x_i) - f(x_i)}{f'(x_i)} = x_i - \frac{f(x_i)}{f'(x_i)}.$$

EXAMPLE 6.3.2 Returning to the previous example, $f(x) = x^2 - 3$, $f'(x) = 2x$, and the formula becomes $x_{i+1} = x_i - (x_i^2 - 3)/(2x_i) = (x_i^2 + 3)/(2x_i)$, as before. □

In practice, which is to say, if you need to approximate a value in the course of designing a bridge or a building or an airframe, you will need to have some confidence that the approximation you settle on is accurate enough. As a rule of thumb, once a certain number of decimal places stop changing from one approximation to the next it is likely that those decimal places are correct. Still, this may not be enough assurance, in which case we can test the result for accuracy.

EXAMPLE 6.3.3 Find the x coordinate of the intersection of the curves $y = 2x$ and $y = \tan x$, accurate to three decimal places. To put this in the context of Newton's method, we note that we want to know where $2x = \tan x$ or $f(x) = \tan x - 2x = 0$. We compute $f'(x) = \sec^2 x - 2$ and set up the formula:

$$x_{i+1} = x_i - \frac{\tan x_i - 2x_i}{\sec^2 x_i - 2}.$$

From the graph in figure 6.3.2 we guess $x_0 = 1$ as a starting point, then using the formula we compute $x_1 = 1.310478030$, $x_2 = 1.223929096$, $x_3 = 1.176050900$, $x_4 = 1.165926508$, $x_5 = 1.165561636$. So we guess that the first three places are correct, but that is not the same as saying 1.165 is correct to three decimal places—1.166 might be the correct, rounded approximation. How can we tell? We can substitute 1.165, 1.1655 and 1.166 into $\tan x - 2x$; this gives -0.002483652, -0.000271247, 0.001948654. Since the first two are negative and the third is positive, $\tan x - 2x$ crosses the x axis between 1.1655 and 1.166, so the correct value to three places is 1.166. □

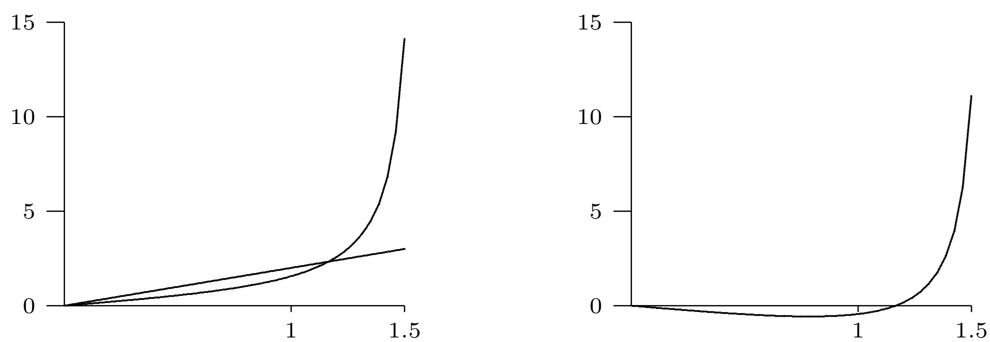


Figure 6.3.2 $y = \tan x$ and $y = 2x$ on the left, $y = \tan x - 2x$ on the right.

Exercises 6.3.

1. Approximate the fifth root of 7, using $x_0 = 1.5$ as a first guess. Use Newton's method to find x_3 as your approximation. ⇒
2. Use Newton's Method to approximate the cube root of 10 to two decimal places. ⇒
3. The function $f(x) = x^3 - 3x^2 - 3x + 6$ has a root between 3 and 4, because $f(3) = -3$ and $f(4) = 10$. Approximate the root to two decimal places. ⇒
4. A rectangular piece of cardboard of dimensions 8×17 is used to make an open-top box by cutting out a small square of side x from each corner and bending up the sides. (See exercise 20 in 6.1.) If $x = 2$, then the volume of the box is $2 \cdot 4 \cdot 13 = 104$. Use Newton's method to find a value of x for which the box has volume 100, accurate to 3 significant figures. ⇒

6.4 Linear Approximations

Newton's method is one example of the usefulness of the tangent line as an approximation to a curve. Here we explore another such application.

Recall that the tangent line to $f(x)$ at a point $x = a$ is given by $L(x) = f'(a)(x-a) + f(a)$. The tangent line in this context is also called the **linear approximation** to f at a.

If f is differentiable at a then L is a good approximation of f so long as x is "not too far" from a. Put another way, if f is differentiable at a then under a microscope f will look very much like a straight line. Figure 6.4.1 shows a tangent line to $y = x^2$ at three different magnifications.

If we want to approximate $f(b)$, because computing it exactly is difficult, we can approximate the value using a linear approximation, provided that we can compute the tangent line at some a close to b.

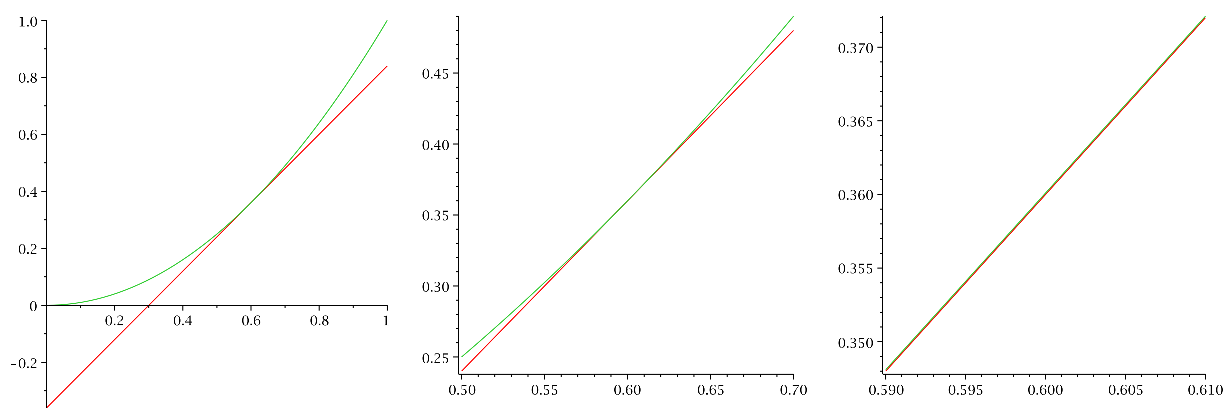


Figure 6.4.1 The linear approximation to $y = x^2$.

EXAMPLE 6.4.1 Let $f(x) = \sqrt{x+4}$. Then $f'(x) = 1/(2\sqrt{x+4})$. The linear approximation to f at $x = 5$ is $L(x) = 1/(2\sqrt{5+4})(x-5) + \sqrt{5+4} = (x-5)/6 + 3$. As an immediate application we can approximate square roots of numbers near 9 by hand. To estimate $\sqrt{10}$, we substitute 6 into the linear approximation instead of into $f(x)$, so $\sqrt{6+4} \approx (6-5)/6 + 3 = 19/6 \approx 3.1\overline{6}$. This rounds to 3.17 while the square root of 10 is actually 3.16 to two decimal places, so this estimate is only accurate to one decimal place. This is not too surprising, as 10 is really not very close to 9; on the other hand, for many calculations, 3.2 would be accurate enough. □

With modern calculators and computing software it may not appear necessary to use linear approximations. But in fact they are quite useful. In cases requiring an explicit numerical approximation, they allow us to get a quick rough estimate which can be used as a "reality check" on a more complex calculation. In some complex calculations involving functions, the linear approximation makes an otherwise intractable calculation possible, without serious loss of accuracy.

EXAMPLE 6.4.2 Consider the trigonometric function $\sin x$. Its linear approximation at $x = 0$ is simply $L(x) = x$. When x is small this is quite a good approximation and is used frequently by engineers and scientists to simplify some calculations. □

DEFINITION 6.4.3 Let $y = f(x)$ be a differentiable function. We define a new independent variable dx, and a new dependent variable $dy = f'(x)\,dx$. Notice that dy is a function both of x (since $f'(x)$ is a function of x) and of dx. We say that dx and dy are **differentials.** □

Let $\Delta x = x - a$ and $\Delta y = f(x) - f(a)$. If x is near a then Δx is small. If we set $dx = \Delta x$ then

$$dy = f'(a)\,dx \approx \frac{\Delta y}{\Delta x}\Delta x = \Delta y.$$

Thus, dy can be used to approximate Δy, the actual change in the function f between a and x. This is exactly the approximation given by the tangent line:

$$dy = f'(a)(x - a) = f'(a)(x - a) + f(a) - f(a) = L(x) - f(a).$$

While $L(x)$ approximates $f(x)$, dy approximates how $f(x)$ has changed from $f(a)$. Figure 6.4.2 illustrates the relationships.

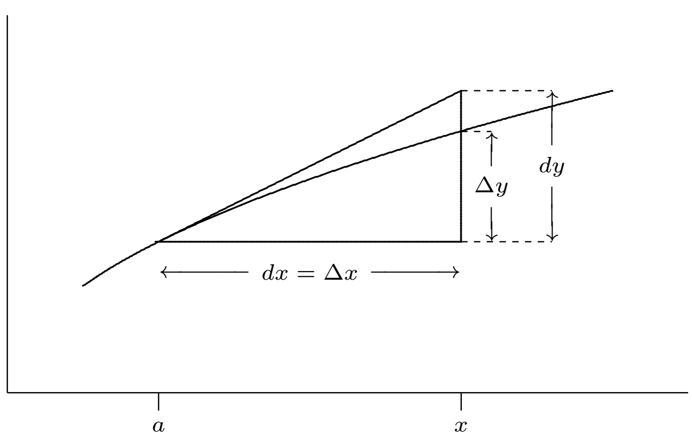


Figure 6.4.2 Differentials.

Exercises 6.4.

1. Let $f(x) = x^4$. If $a = 1$ and $dx = \Delta x = 1/2$, what are Δy and dy? ⇒
2. Let $f(x) = \sqrt{x}$. If $a = 1$ and $dx = \Delta x = 1/10$, what are Δy and dy? ⇒
3. Let $f(x) = \sin(2x)$. If $a = \pi$ and $dx = \Delta x = \pi/100$, what are Δy and dy? ⇒
4. Use differentials to estimate the amount of paint needed to apply a coat of paint 0.02 cm thick to a sphere with diameter 40 meters. (Recall that the volume of a sphere of radius r is $V = (4/3)\pi r^3$. Notice that you are given that $dr = 0.02$.) ⇒
5. Show in detail that the linear approximation of $\sin x$ at $x = 0$ is $L(x) = x$ and the linear approximation of $\cos x$ at $x = 0$ is $L(x) = 1$.

6.5 The Mean Value Theorem

Here are two interesting questions involving derivatives:

1. Suppose two different functions have the same derivative; what can you say about the relationship between the two functions?
2. Suppose you drive a car from toll booth on a toll road to another toll booth at an average speed of 70 miles per hour. What can be concluded about your actual speed during the trip? In particular, did you exceed the 65 mile per hour speed limit?

While these sound very different, it turns out that the two problems are very closely related. We know that "speed" is really the derivative by a different name; let's start by translating the second question into something that may be easier to visualize. Suppose that the function $f(t)$ gives the position of your car on the toll road at time t. Your change in position between one toll booth and the next is given by $f(t_1) - f(t_0)$, assuming that at time t_0 you were at the first booth and at time t_1 you arrived at the second booth. Your average speed for the trip is $(f(t_1) - f(t_0))/(t_1 - t_0)$. If we think about the graph of $f(t)$, the average speed is the slope of the line that connects the two points $(t_0, f(t_0))$ and $(t_1, f(t_1))$. Your speed at any particular time t between t_0 and t_1 is $f'(t)$, the slope of the curve. Now question (2) becomes a question about slope. In particular, if the slope between endpoints is 70, what can be said of the slopes at points between the endpoints?

As a general rule, when faced with a new problem it is often a good idea to examine one or more simplified versions of the problem, in the hope that this will lead to an understanding of the original problem. In this case, the problem in its "slope" form is somewhat easier to simplify than the original, but equivalent, problem.

Here is a special instance of the problem. Suppose that $f(t_0) = f(t_1)$. Then the two endpoints have the same height and the slope of the line connecting the endpoints is zero. What can we say about the slope between the endpoints? It shouldn't take much experimentation before you are convinced of the truth of this statement: Somewhere between t_0 and t_1 the slope is exactly zero, that is, somewhere between t_0 and t_1 the slope is equal to the slope of the line between the endpoints. This suggests that perhaps the same is true even if the endpoints are at different heights, and again a bit of experimentation will probably convince you that this is so. But we can do better than "experimentation"—we can prove that this is so.

We start with the simplified version:

THEOREM 6.5.1 Rolle's Theorem Suppose that $f(x)$ has a derivative on the interval (a,b), is continuous on the interval $[a,b]$, and $f(a) = f(b)$. Then at some value $c \in (a,b)$, $f'(c) = 0$.

Proof. We know that $f(x)$ has a maximum and minimum value on $[a,b]$ (because it is continuous), and we also know that the maximum and minimum must occur at an endpoint, at a point at which the derivative is zero, or at a point where the derivative is undefined. Since the derivative is never undefined, that possibility is removed.

If the maximum or minimum occurs at a point c, other than an endpoint, where $f'(c) = 0$, then we have found the point we seek. Otherwise, the maximum and minimum both occur at an endpoint, and since the endpoints have the same height, the maximum and minimum are the same. This means that $f(x) = f(a) = f(b)$ at every $x \in [a,b]$, so the function is a horizontal line, and it has derivative zero everywhere in (a,b). Then we may choose any c at all to get $f'(c) = 0$. ∎

Perhaps remarkably, this special case is all we need to prove the more general one as well.

THEOREM 6.5.2 Mean Value Theorem Suppose that $f(x)$ has a derivative on the interval (a,b) and is continuous on the interval $[a,b]$. Then at some value $c \in (a,b)$, $f'(c) = \dfrac{f(b) - f(a)}{b - a}$.

Proof. Let $m = \dfrac{f(b) - f(a)}{b - a}$, and consider a new function $g(x) = f(x) - m(x-a) - f(a)$. We know that $g(x)$ has a derivative everywhere, since $g'(x) = f'(x) - m$. We can compute $g(a) = f(a) - m(a-a) - f(a) = 0$ and

$$\begin{aligned} g(b) = f(b) - m(b-a) - f(a) &= f(b) - \frac{f(b) - f(a)}{b - a}(b-a) - f(a) \\ &= f(b) - (f(b) - f(a)) - f(a) = 0. \end{aligned}$$

So the height of $g(x)$ is the same at both endpoints. This means, by Rolle's Theorem, that at some c, $g'(c) = 0$. But we know that $g'(c) = f'(c) - m$, so

$$0 = f'(c) - m = f'(c) - \frac{f(b) - f(a)}{b - a},$$

which turns into

$$f'(c) = \frac{f(b) - f(a)}{b - a},$$

exactly what we want. ■

Returning to the original formulation of question (2), we see that if $f(t)$ gives the position of your car at time t, then the Mean Value Theorem says that at some time c, $f'(c) = 70$, that is, at some time you must have been traveling at exactly your average speed for the trip, and that indeed you exceeded the speed limit.

Now let's return to question (1). Suppose, for example, that two functions are known to have derivative equal to 5 everywhere, $f'(x) = g'(x) = 5$. It is easy to find such functions: $5x$, $5x + 47$, $5x - 132$, etc. Are there other, more complicated, examples? No—the only functions that work are the "obvious" ones, namely, $5x$ plus some constant. How can we see that this is true?

Although "5" is a very simple derivative, let's look at an even simpler one. Suppose that $f'(x) = g'(x) = 0$. Again we can find examples: $f(x) = 0$, $f(x) = 47$, $f(x) = -511$ all have $f'(x) = 0$. Are there non-constant functions f with derivative 0? No, and here's why: Suppose that $f(x)$ is not a constant function. This means that there are two points on the function with different heights, say $f(a) \neq f(b)$. The Mean Value Theorem tells us that at some point c, $f'(c) = (f(b) - f(a))/(b - a) \neq 0$. So any non-constant function does not have a derivative that is zero everywhere; this is the same as saying that the only functions with zero derivative are the constant functions.

Let's go back to the slightly less easy example: suppose that $f'(x) = g'(x) = 5$. Then $(f(x) - g(x))' = f'(x) - g'(x) = 5 - 5 = 0$. So using what we discovered in the previous paragraph, we know that $f(x) - g(x) = k$, for some constant k. So any two functions with derivative 5 must differ by a constant; since $5x$ is known to work, the only other examples must look like $5x + k$.

Now we can extend this to more complicated functions, without any extra work. Suppose that $f'(x) = g'(x)$. Then as before $(f(x) - g(x))' = f'(x) - g'(x) = 0$, so $f(x) - g(x) = k$. Again this means that if we find just a single function $g(x)$ with a certain derivative, then every other function with the same derivative must be of the form $g(x) + k$.

EXAMPLE 6.5.3 Describe all functions that have derivative $5x - 3$. It's easy to find one: $g(x) = (5/2)x^2 - 3x$ has $g'(x) = 5x - 3$. The only other functions with the same derivative are therefore of the form $f(x) = (5/2)x^2 - 3x + k$.

Alternately, though not obviously, you might have first noticed that $g(x) = (5/2)x^2 - 3x + 47$ has $g'(x) = 5x - 3$. Then every other function with the same derivative must have the form $f(x) = (5/2)x^2 - 3x + 47 + k$. This looks different, but it really isn't. The functions of the form $f(x) = (5/2)x^2 - 3x + k$ are exactly the same as the ones of the form $f(x) = (5/2)x^2 - 3x + 47 + k$. For example, $(5/2)x^2 - 3x + 10$ is the same as $(5/2)x^2 - 3x + 47 + (-37)$, and the first is of the first form while the second has the second form. □

This is worth calling a theorem:

THEOREM 6.5.4 If $f'(x) = g'(x)$ for every $x \in (a, b)$, then for some constant k, $f(x) = g(x) + k$ on the interval (a, b). ■

EXAMPLE 6.5.5 Describe all functions with derivative $\sin x + e^x$. One such function is $-\cos x + e^x$, so all such functions have the form $-\cos x + e^x + k$. □

Exercises 6.5.

1. Let $f(x) = x^2$. Find a value $c \in (-1, 2)$ so that $f'(c)$ equals the slope between the endpoints of $f(x)$ on $[-1, 2]$. ⇒

2. Verify that $f(x) = x/(x + 2)$ satisfies the hypotheses of the Mean Value Theorem on the interval $[1, 4]$ and then find all of the values, c, that satisfy the conclusion of the theorem. ⇒

3. Verify that $f(x) = 3x/(x + 7)$ satisfies the hypotheses of the Mean Value Theorem on the interval $[-2, 6]$ and then find all of the values, c, that satisfy the conclusion of the theorem.

4. Let $f(x) = \tan x$. Show that $f(\pi) = f(2\pi) = 0$ but there is no number $c \in (\pi, 2\pi)$ such that $f'(c) = 0$. Why does this not contradict Rolle's theorem?

5. Let $f(x) = (x - 3)^{-2}$. Show that there is no value $c \in (1, 4)$ such that $f'(c) = (f(4) - f(1))/(4 - 1)$. Why is this not a contradiction of the Mean Value Theorem?

6. Describe all functions with derivative $x^2 + 47x - 5$. ⇒

7. Describe all functions with derivative $\dfrac{1}{1 + x^2}$. ⇒

8. Describe all functions with derivative $x^3 - \dfrac{1}{x}$. ⇒

9. Describe all functions with derivative $\sin(2x)$. ⇒

10. Show that the equation $6x^4 - 7x + 1 = 0$ does not have more than two distinct real roots.

11. Let f be differentiable on $\mathbb{R}$. Suppose that $f'(x) \neq 0$ for every x. Prove that f has at most one real root.

12. Prove that for all real x and y $|\cos x - \cos y| \le |x - y|$. State and prove an analogous result involving sine.

13. Show that $\sqrt{1 + x} \le 1 + (x/2)$ if $-1 < x < 1$.

7

Integration

7.1 Two examples

Up to now we have been concerned with extracting information about how a function changes from the function itself. Given knowledge about an object's position, for example, we want to know the object's speed. Given information about the height of a curve we want to know its slope. We now consider problems that are, whether obviously or not, the reverse of such problems.

EXAMPLE 7.1.1 An object moves in a straight line so that its speed at time t is given by $v(t) = 3t$ in, say, cm/sec. If the object is at position 10 on the straight line when $t = 0$, where is the object at any time t?

There are two reasonable ways to approach this problem. If $s(t)$ is the position of the object at time t, we know that $s'(t) = v(t)$. Because of our knowledge of derivatives, we know therefore that $s(t) = 3t^2/2 + k$, and because $s(0) = 10$ we easily discover that $k = 10$, so $s(t) = 3t^2/2 + 10$. For example, at $t = 1$ the object is at position $3/2 + 10 = 11.5$. This is certainly the easiest way to deal with this problem. Not all similar problems are so easy, as we will see; the second approach to the problem is more difficult but also more general.

We start by considering how we might approximate a solution. We know that at $t = 0$ the object is at position 10. How might we approximate its position at, say, $t = 1$? We know that the speed of the object at time $t = 0$ is 0; if its speed were constant then in the first second the object would not move and its position would still be 10 when $t = 1$. In fact, the object will not be too far from 10 at $t = 1$, but certainly we can do better. Let's look at the times 0.1, 0.2, 0.3, …, 1.0, and try approximating the location of the object at each, by supposing that during each tenth of a second the object is going at a constant speed. Since the object initially has speed 0, we again suppose it maintains this speed, but only for a tenth of second; during that time the object would not move. During the tenth of a second from $t = 0.1$ to $t = 0.2$, we suppose that the object is traveling at 0.3 cm/sec, namely, its actual speed at $t = 0.1$. In this case the object would travel $(0.3)(0.1) = 0.03$ centimeters: 0.3 cm/sec times 0.1 seconds. Similarly, between $t = 0.2$ and $t = 0.3$ the object would travel $(0.6)(0.1) = 0.06$ centimeters. Continuing, we get as an approximation that the object travels

$$(0.0)(0.1) + (0.3)(0.1) + (0.6)(0.1) + \cdots + (2.7)(0.1) = 1.35$$

centimeters, ending up at position 11.35. This is a better approximation than 10, certainly, but is still just an approximation. (We know in fact that the object ends up at position 11.5, because we've already done the problem using the first approach.) Presumably, we will get a better approximation if we divide the time into one hundred intervals of a hundredth of a second each, and repeat the process:

$$(0.0)(0.01) + (0.03)(0.01) + (0.06)(0.01) + \cdots + (2.97)(0.01) = 1.485.$$

We thus approximate the position as 11.485. Since we know the exact answer, we can see that this is much closer, but if we did not already know the answer, we wouldn't really know how close.

We can keep this up, but we'll never really know the exact answer if we simply compute more and more examples. Let's instead look at a "typical" approximation. Suppose we divide the time into n equal intervals, and imagine that on each of these the object travels at a constant speed. Over the first time interval we approximate the distance traveled as $(0.0)(1/n) = 0$, as before. During the second time interval, from $t = 1/n$ to $t = 2/n$, the object travels approximately $3(1/n)(1/n) = 3/n^2$ centimeters. During time interval number i, the object travels approximately $(3(i-1)/n)(1/n) = 3(i-1)/n^2$ centimeters, that is, its speed at time $(i-1)/n$, $3(i-1)/n$, times the length of time interval number i, $1/n$. Adding these up as before, we approximate the distance traveled as

$$(0)\frac{1}{n} + 3\frac{1}{n^2} + 3(2)\frac{1}{n^2} + 3(3)\frac{1}{n^2} + \cdots + 3(n-1)\frac{1}{n^2}$$

centimeters. What can we say about this? At first it looks rather less useful than the concrete calculations we've already done. But in fact a bit of algebra reveals it to be much more useful. We can factor out a 3 and $1/n^2$ to get

$$\frac{3}{n^2}(0+1+2+3+\cdots+(n-1)),$$

that is, $3/n^2$ times the sum of the first $n-1$ positive integers. Now we make use of a fact you may have run across before:

$$1+2+3+\cdots+k=\frac{k(k+1)}{2}.$$

In our case we're interested in $k=n-1$, so

$$1+2+3+\cdots+(n-1)=\frac{(n-1)(n)}{2}=\frac{n^2-n}{2}.$$

This simplifies the approximate distance traveled to

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

Now this is quite easy to understand: as n gets larger and larger this approximation gets closer and closer to $(3/2)(1-0)=3/2$, so that $3/2$ is the exact distance traveled during one second, and the final position is 11.5.

So for $t=1$, at least, this rather cumbersome approach gives the same answer as the first approach. But really there's nothing special about $t=1$; let's just call it t instead. In this case the approximate distance traveled during time interval number i is $3(i-1)(t/n)(t/n)=3(i-1)t^2/n^2$, that is, speed $3(i-1)(t/n)$ times time t/n, and the total distance traveled is approximately

$$(0)\frac{t}{n}+3(1)\frac{t^2}{n^2}+3(2)\frac{t^2}{n^2}+3(3)\frac{t^2}{n^2}+\cdots+3(n-1)\frac{t^2}{n^2}.$$

As before we can simplify this to

$$\frac{3t^2}{n^2}(0+1+2+\cdots+(n-1))=\frac{3t^2}{n^2}\frac{n^2-n}{2}=\frac{3}{2}t^2\left(1-\frac{1}{n}\right).$$

In the limit, as n gets larger, this gets closer and closer to $(3/2)t^2$ and the approximated position of the object gets closer and closer to $(3/2)t^2+10$, so the actual position is $(3/2)t^2+10$, exactly the answer given by the first approach to the problem. □

EXAMPLE 7.1.2 Find the area under the curve $y=3x$ between $x=0$ and any positive value x. There is here no obvious analogue to the first approach in the previous example, but the second approach works fine. (Because the function $y = 3x$ is so simple, there is another approach that works here, but it is even more limited in potential application than is approach number one.) How might we approximate the desired area? We know how to compute areas of rectangles, so we approximate the area by rectangles. Jumping straight to the general case, suppose we divide the interval between 0 and x into n equal subintervals, and use a rectangle above each subinterval to approximate the area under the curve. There are many ways we might do this, but let's use the height of the curve at the left endpoint of the subinterval as the height of the rectangle, as in figure 7.1.1. The height of rectangle number i is then $3(i-1)(x/n)$, the width is x/n, and the area is $3(i-1)(x^2/n^2)$. The total area of the rectangles is

$$(0)\frac{x}{n} + 3(1)\frac{x^2}{n^2} + 3(2)\frac{x^2}{n^2} + 3(3)\frac{x^2}{n^2} + \cdots + 3(n-1)\frac{x^2}{n^2}.$$

By factoring out $3x^2/n^2$ this simplifies to

$$\frac{3x^2}{n^2}(0 + 1 + 2 + \cdots + (n-1)) = \frac{3x^2}{n^2}\frac{n^2-n}{2} = \frac{3}{2}x^2\left(1 - \frac{1}{n}\right).$$

As n gets larger this gets closer and closer to $3x^2/2$, which must therefore be the true area under the curve. □

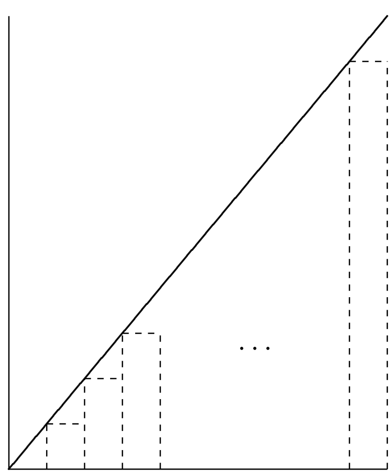


Figure 7.1.1 Approximating the area under $y = 3x$ with rectangles. Drag the slider to change the number of rectangles.

What you will have noticed, of course, is that while the problem in the second example appears to be much different than the problem in the first example, and while the easy approach to problem one does not appear to apply to problem two, the "approximation" approach works in both, and moreover the *calculations are identical.* As we will see, there are many, many problems that appear much different on the surface but that turn out to be the same as these problems, in the sense that when we try to approximate solutions we end up with mathematics that looks like the two examples, though of course the function involved will not always be so simple.

Even better, we now see that while the second problem did not appear to be amenable to approach one, it can in fact be solved in the same way. The reasoning is this: we know that problem one can be solved easily by finding a function whose derivative is $3t$. We also know that mathematically the two problems are the same, because both can be solved by taking a limit of a sum, and the sums are identical. Therefore, we don't really need to compute the limit of either sum because we know that we will get the same answer by computing a function with the derivative $3t$ or, which is the same thing, $3x$.

It's true that the first problem had the added complication of the "10", and we certainly need to be able to deal with such minor variations, but that turns out to be quite simple. The lesson then is this: whenever we can solve a problem by taking the limit of a sum of a certain form, we can instead of computing the (often nasty) limit find a new function with a certain derivative.

Exercises 7.1.

1. Suppose an object moves in a straight line so that its speed at time t is given by $v(t) = 2t+2$, and that at $t = 1$ the object is at position 5. Find the position of the object at $t = 2$. ⇒
2. Suppose an object moves in a straight line so that its speed at time t is given by $v(t) = t^2+2$, and that at $t = 0$ the object is at position 5. Find the position of the object at $t = 2$. ⇒
3. By a method similar to that in example 7.1.2, find the area under $y = 2x$ between $x = 0$ and any positive value for x. ⇒
4. By a method similar to that in example 7.1.2, find the area under $y = 4x$ between $x = 0$ and any positive value for x. ⇒
5. By a method similar to that in example 7.1.2, find the area under $y = 4x$ between $x = 2$ and any positive value for x bigger than 2. ⇒
6. By a method similar to that in example 7.1.2, find the area under $y = 4x$ between any two positive values for x, say $a < b$. ⇒
7. Let $f(x) = x^2 + 3x + 2$. Approximate the area under the curve between $x = 0$ and $x = 2$ using 4 rectangles and also using 8 rectangles. ⇒
8. Let $f(x) = x^2 - 2x + 3$. Approximate the area under the curve between $x = 1$ and $x = 3$ using 4 rectangles. ⇒

7.2 The Fundamental Theorem of Calculus

Let's recast the first example from the previous section. Suppose that the speed of the object is $3t$ at time t. How far does the object travel between time $t = a$ and time $t = b$? We are no longer assuming that we know where the object is at time $t = 0$ or at any other time. It is certainly true that it is *somewhere,* so let's suppose that at $t = 0$ the position is k. Then just as in the example, we know that the position of the object at any time is $3t^2/2+k$. This means that at time $t = a$ the position is $3a^2/2 + k$ and at time $t = b$ the position is $3b^2/2 + k$. Therefore the change in position is $3b^2/2 + k - (3a^2/2 + k) = 3b^2/2 - 3a^2/2$. Notice that the k drops out; this means that it doesn't matter that we don't know k, it doesn't even matter if we use the wrong k, we get the correct answer. In other words, to find the change in position between time a and time b we can use *any* antiderivative of the speed function $3t$—it need not be the one antiderivative that actually gives the location of the object.

What about the second approach to this problem, in the new form? We now want to approximate the change in position between time a and time b. We take the interval of time between a and b, divide it into n subintervals, and approximate the distance traveled during each. The starting time of subinterval number i is now $a + (i - 1)(b - a)/n$, which we abbreviate as t_{i-1}, so that $t_0 = a$, $t_1 = a + (b - a)/n$, and so on. The speed of the object is $f(t) = 3t$, and each subinterval is $(b - a)/n = \Delta t$ seconds long. The distance traveled during subinterval number i is approximately $f(t_{i-1})\Delta t$, and the total change in distance is approximately

$$f(t_0)\Delta t + f(t_1)\Delta t + \cdots + f(t_{n-1})\Delta t.$$

The exact change in position is the limit of this sum as n goes to infinity. We abbreviate this sum using **sigma notation**:

$$\sum_{i=0}^{n-1} f(t_i)\Delta t = f(t_0)\Delta t + f(t_1)\Delta t + \cdots + f(t_{n-1})\Delta t.$$

The notation on the left side of the equal sign uses a large capital sigma, a Greek letter, and the left side is an abbreviation for the right side. The answer we seek is

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} f(t_i)\Delta t.$$

Since this must be the same as the answer we have already obtained, we know that

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} f(t_i)\Delta t = \frac{3b^2}{2} - \frac{3a^2}{2}.$$

The significance of $3t^2/2$, into which we substitute $t = b$ and $t = a$, is of course that it is a function whose derivative is $f(t)$. As we have discussed, by the time we know that we want to compute

$$\lim_{n\to\infty}\sum_{i=0}^{n-1} f(t_i)\Delta t,$$

it no longer matters what $f(t)$ stands for—it could be a speed, or the height of a curve, or something else entirely. We know that the limit can be computed by finding any function with derivative $f(t)$, substituting a and b, and subtracting. We summarize this in a theorem. First, we introduce some new notation and terms.

We write

$$\int_a^b f(t)\,dt = \lim_{n\to\infty}\sum_{i=0}^{n-1} f(t_i)\Delta t$$

if the limit exists. That is, the left hand side means, or is an abbreviation for, the right hand side. The symbol $\int$ is called an **integral sign**, and the whole expression is read as "the integral of $f(t)$ from a to b." What we have learned is that this integral can be computed by finding a function, say $F(t)$, with the property that $F'(t) = f(t)$, and then computing $F(b) - F(a)$. The function $F(t)$ is called an **antiderivative** of $f(t)$. Now the theorem:

THEOREM 7.2.1 Fundamental Theorem of Calculus Suppose that $f(x)$ is continuous on the interval $[a, b]$. If $F(x)$ is any antiderivative of $f(x)$, then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

□

Let's rewrite this slightly:

$$\int_a^x f(t)\,dt = F(x) - F(a).$$

We've replaced the variable x by t and b by x. These are just different names for quantities, so the substitution doesn't change the meaning. It does make it easier to think of the two sides of the equation as functions. The expression

$$\int_a^x f(t)\,dt$$

is a function: plug in a value for x, get out some other value. The expression $F(x) - F(a)$ is of course also a function, and it has a nice property:

$$\frac{d}{dx}(F(x) - F(a)) = F'(x) = f(x),$$

since $F(a)$ is a constant and has derivative zero. In other words, by shifting our point of view slightly, we see that the odd looking function

$$G(x) = \int_a^x f(t)\,dt$$

has a derivative, and that in fact $G'(x) = f(x)$. This is really just a restatement of the Fundamental Theorem of Calculus, and indeed is often called the Fundamental Theorem of Calculus. To avoid confusion, some people call the two versions of the theorem "The Fundamental Theorem of Calculus, part I" and "The Fundamental Theorem of Calculus, part II", although unfortunately there is no universal agreement as to which is part I and which part II. Since it really is the same theorem, differently stated, some people simply call them both "The Fundamental Theorem of Calculus."

THEOREM 7.2.2 Fundamental Theorem of Calculus Suppose that $f(x)$ is continuous on the interval $[a,b]$ and let

$$G(x) = \int_a^x f(t)\,dt.$$

Then $G'(x) = f(x)$. □

We have not really proved the Fundamental Theorem. In a nutshell, we gave the following argument to justify it: Suppose we want to know the value of

$$\int_a^b f(t)\,dt = \lim_{n\to\infty} \sum_{i=0}^{n-1} f(t_i)\Delta t.$$

We can interpret the right hand side as the distance traveled by an object whose speed is given by $f(t)$. We know another way to compute the answer to such a problem: find the position of the object by finding an antiderivative of $f(t)$, then substitute $t = a$ and $t = b$ and subtract to find the distance traveled. This must be the answer to the original problem as well, even if $f(t)$ does not represent a speed.

What's wrong with this? In some sense, nothing. As a practical matter it is a very convincing argument, because our understanding of the relationship between speed and distance seems to be quite solid. From the point of view of mathematics, however, it is unsatisfactory to justify a purely mathematical relationship by appealing to our understanding of the physical universe, which could, however unlikely it is in this case, be wrong.

A complete proof is a bit too involved to include here, but we will indicate how it goes. First, if we can prove the second version of the Fundamental Theorem, theorem 7.2.2, then we can prove the first version from that:

Proof of Theorem 7.2.1. We know from theorem 7.2.2 that

$$G(x) = \int_a^x f(t)\,dt$$

is an antiderivative of $f(x)$, and therefore any antiderivative $F(x)$ of $f(x)$ is of the form $F(x) = G(x) + k$. Then

$$\begin{aligned} F(b) - F(a) = G(b) + k - (G(a) + k) &= G(b) - G(a) \\ &= \int_a^b f(t)\,dt - \int_a^a f(t)\,dt. \end{aligned}$$

It is not hard to see that $\int_a^a f(t)\,dt = 0$, so this means that

$$F(b) - F(a) = \int_a^b f(t)\,dt,$$

which is exactly what theorem 7.2.1 says. ∎

So the real job is to prove theorem 7.2.2. We will sketch the proof, using some facts that we do not prove. First, the following identity is true of integrals:

$$\int_a^b f(t)\,dt = \int_a^c f(t)\,dt + \int_c^b f(t)\,dt.$$

This can be proved directly from the definition of the integral, that is, using the limits of sums. It is quite easy to see that it must be true by thinking of either of the two applications of integrals that we have seen. It turns out that the identity is true no matter what c is, but it is easiest to think about the meaning when $a \le c \le b$.

First, if $f(t)$ represents a speed, then we know that the three integrals represent the distance traveled between time a and time b; the distance traveled between time a and time c; and the distance traveled between time c and time b. Clearly the sum of the latter two is equal to the first of these.

Second, if $f(t)$ represents the height of a curve, the three integrals represent the area under the curve between a and b; the area under the curve between a and c; and the area under the curve between c and b. Again it is clear from the geometry that the first is equal to the sum of the second and third.

Proof sketch for Theorem 7.2.2. We want to compute $G'(x)$, so we start with the definition of the derivative in terms of a limit:

$$\begin{aligned} G'(x) &= \lim_{\Delta x\to 0} \frac{G(x+\Delta x)-G(x)}{\Delta x} \\ &= \lim_{\Delta x\to 0} \frac{1}{\Delta x}\left(\int_a^{x+\Delta x} f(t)\,dt - \int_a^x f(t)\,dt\right) \\ &= \lim_{\Delta x\to 0} \frac{1}{\Delta x}\left(\int_a^{x} f(t)\,dt + \int_x^{x+\Delta x} f(t)\,dt - \int_a^x f(t)\,dt\right) \\ &= \lim_{\Delta x\to 0} \frac{1}{\Delta x}\int_x^{x+\Delta x} f(t)\,dt. \end{aligned}$$

Now we need to know something about

$$\int_x^{x+\Delta x} f(t)\,dt$$

when Δx is small; in fact, it is very close to $\Delta x f(x)$, but we will not prove this. Once again, it is easy to believe this is true by thinking of our two applications: The integral

$$\int_x^{x+\Delta x} f(t)\,dt$$

can be interpreted as the distance traveled by an object over a very short interval of time. Over a sufficiently short period of time, the speed of the object will not change very much, so the distance traveled will be approximately the length of time multiplied by the speed at the beginning of the interval, namely, $\Delta x f(x)$. Alternately, the integral may be interpreted as the area under the curve between x and $x+\Delta x$. When Δx is very small, this will be very close to the area of the rectangle with base Δx and height $f(x)$; again this is $\Delta x f(x)$. If we accept this, we may proceed:

$$\lim_{\Delta x\to 0} \frac{1}{\Delta x}\int_x^{x+\Delta x} f(t)\,dt = \lim_{\Delta x\to 0} \frac{\Delta x f(x)}{\Delta x} = f(x),$$

which is what we wanted to show. ∎

It is still true that we are depending on an interpretation of the integral to justify the argument, but we have isolated this part of the argument into two facts that are not too hard to prove. Once the last reference to interpretation has been removed from the proofs of these facts, we will have a real proof of the Fundamental Theorem.

Now we know that to solve certain kinds of problems, those that lead to a sum of a certain form, we "merely" find an antiderivative and substitute two values and subtract. Unfortunately, finding antiderivatives can be quite difficult. While there are a small number of rules that allow us to compute the derivative of any common function, there are no such rules for antiderivatives. There are some techniques that frequently prove useful, but we will never be able to reduce the problem to a completely mechanical process.

Because of the close relationship between an integral and an antiderivative, the integral sign is also used to mean "antiderivative". You can tell which is intended by whether the limits of integration are included:

$$\int_1^2 x^2\,dx$$

is an ordinary integral, also called a **definite integral**, because it has a definite value, namely

$$\int_1^2 x^2\,dx = \frac{2^3}{3} - \frac{1^3}{3} = \frac{7}{3}.$$

We use

$$\int x^2\,dx$$

to denote the antiderivative of x^2, also called an **indefinite integral**. So this is evaluated as

$$\int x^2\,dx = \frac{x^3}{3} + C.$$

It is customary to include the constant C to indicate that there are really an infinite number of antiderivatives. We do not need this C to compute definite integrals, but in other circumstances we will need to remember that the C is there, so it is best to get into the habit of writing the C. When we compute a definite integral, we first find an antiderivative and then substitute. It is convenient to first display the antiderivative and then do the substitution; we need a notation indicating that the substitution is yet to be done. A typical solution would look like this:

$$\int_1^2 x^2\,dx = \left.\frac{x^3}{3}\right|_1^2 = \frac{2^3}{3} - \frac{1^3}{3} = \frac{7}{3}.$$

The vertical line with subscript and superscript is used to indicate the operation "substitute and subtract" that is needed to finish the evaluation.

Exercises 7.2.

Find the antiderivatives of the functions:

1. $8\sqrt{x}$ ⇒

2. $3t^2 + 1$ ⇒

3. $4/\sqrt{x}$ ⇒

4. $2/z^2$ ⇒

5. $7s^{-1}$ ⇒

6. $(5x + 1)^2$ ⇒

7. $(x - 6)^2$ ⇒

8. $x^{3/2}$ ⇒

9. $\dfrac{2}{x\sqrt{x}}$ ⇒

10. $|2t - 4|$ ⇒

Compute the values of the integrals:

11. $\displaystyle\int_1^4 t^2 + 3t\,dt$ ⇒

12. $\displaystyle\int_0^\pi \sin t\,dt$ ⇒

13. $\displaystyle\int_1^{10} \frac{1}{x}\,dx$ ⇒

14. $\displaystyle\int_0^5 e^x\,dx$ ⇒

15. $\displaystyle\int_0^3 x^3\,dx$ ⇒

16. $\displaystyle\int_1^2 x^5\,dx$ ⇒

17. Find the derivative of $G(x) = \displaystyle\int_1^x t^2 - 3t\,dt$ ⇒

18. Find the derivative of $G(x) = \displaystyle\int_1^{x^2} t^2 - 3t\,dt$ ⇒

19. Find the derivative of $G(x) = \displaystyle\int_1^x e^{t^2}\,dt$ ⇒

20. Find the derivative of $G(x) = \displaystyle\int_1^{x^2} e^{t^2}\,dt$ ⇒

21. Find the derivative of $G(x) = \displaystyle\int_1^x \tan(t^2)\,dt$ ⇒

22. Find the derivative of $G(x) = \displaystyle\int_1^{x^2} \tan(t^2)\,dt$ ⇒

7.3 Some Properties of Integrals

Suppose an object moves so that its speed, or more properly velocity, is given by $v(t) = -t^2 + 5t$, as shown in figure 7.3.1. Let's examine the motion of this object carefully. We know that the velocity is the derivative of position, so position is given by $s(t) = -t^3/3 + 5t^2/2 + C$. Let's suppose that at time $t = 0$ the object is at position 0, so $s(t) = -t^3/3 + 5t^2/2$; this function is also pictured in figure 7.3.1.

Between $t = 0$ and $t = 5$ the velocity is positive, so the object moves away from the starting point, until it is a bit past position 20. Then the velocity becomes negative and the object moves back toward its starting point. The position of the object at $t = 5$ is

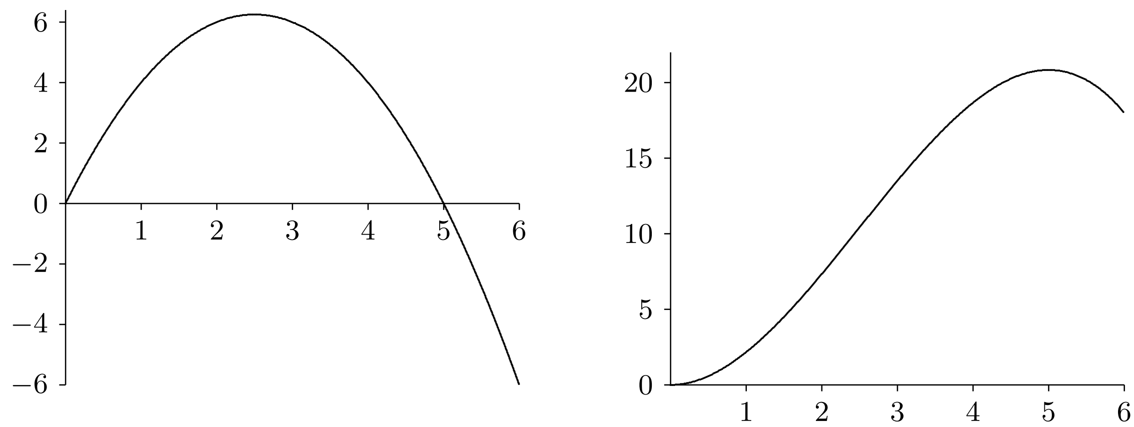


Figure 7.3.1 The velocity of an object and its position.

exactly $s(5) = 125/6$, and at $t = 6$ it is $s(6) = 18$. The total distance traveled by the object is therefore $125/6 + (125/6 - 18) = 71/3 \approx 23.7$.

As we have seen, we can also compute distance traveled with an integral; let's try it.

$$\int_0^6 v(t)\,dt = \int_0^6 -t^2 + 5t\,dt = \left.\frac{-t^3}{3} + \frac{5}{2}t^2\right|_0^6 = 18.$$

What went wrong? Well, nothing really, except that it's not really true after all that "we can also compute distance traveled with an integral". Instead, as you might guess from this example, the integral actually computes the *net* distance traveled, that is, the difference between the starting and ending point.

As we have already seen,

$$\int_0^6 v(t)\,dt = \int_0^5 v(t)\,dt + \int_5^6 v(t)\,dt.$$

Computing the two integrals on the right (do it!) gives $125/6$ and $-17/6$, and the sum of these is indeed 18. But what does that negative sign mean? It means precisely what you might think: it means that the object moves backwards. To get the total distance traveled we can add $125/6 + 17/6 = 71/3$, the same answer we got before.

Remember that we can also interpret an integral as measuring an area, but now we see that this too is a little more complicated that we have suspected. The area under the curve $v(t)$ from 0 to 5 is given by

$$\int_0^5 v(t)\,dt = \frac{125}{6},$$

and the "area" from 5 to 6 is

$$\int_5^6 v(t)\,dt = -\frac{17}{6}.$$

In other words, the area between the x-axis and the curve, but under the x-axis, "counts as negative area". So the integral

$$\int_0^6 v(t)\,dt = 18$$

measures "net area", the area above the axis minus the (positive) area below the axis.

If we recall that the integral is the limit of a certain kind of sum, this behavior is not surprising. Recall the sort of sum involved:

$$\sum_{i=0}^{n-1} v(t_i)\Delta t.$$

In each term $v(t)\Delta t$ the Δt is positive, but if $v(t_i)$ is negative then the term is negative. If over an entire interval, like 5 to 6, the function is always negative, then the entire sum is negative. In terms of area, $v(t)\Delta t$ is then a negative height times a positive width, giving a negative rectangle "area".

So now we see that when evaluating

$$\int_5^6 v(t)\,dt = -\frac{17}{6}$$

by finding an antiderivative, substituting, and subtracting, we get a surprising answer, but one that turns out to make sense.

Let's now try something a bit different:

$$\int_6^5 v(t)\,dt = \left.\frac{-t^3}{3} + \frac{5}{2}t^2\right|_6^5 = \frac{-5^3}{3} + \frac{5}{2}5^2 - \frac{-6^3}{3} - \frac{5}{2}6^2 = \frac{17}{6}.$$

Here we simply interchanged the limits 5 and 6, so of course when we substitute and subtract we're subtracting in the opposite order and we end up multiplying the answer by -1. This too makes sense in terms of the underlying sum, though it takes a bit more thought. Recall that in the sum

$$\sum_{i=0}^{n-1} v(t_i)\Delta t,$$

the Δt is the "length" of each little subinterval, but more precisely we could say that $\Delta t = t_{i+1} - t_i$, the difference between two endpoints of a subinterval. We have until now assumed that we were working left to right, but could as well number the subintervals from right to left, so that $t_0 = b$ and $t_n = a$. Then $\Delta t = t_{i+1} - t_i$ is negative and in

$$\int_6^5 v(t)\,dt = \sum_{i=0}^{n-1} v(t_i)\Delta t,$$

the values $v(t_i)$ are negative but also Δt is negative, so all terms are positive again. On the other hand, in

$$\int_5^0 v(t)\,dt = \sum_{i=0}^{n-1} v(t_i)\Delta t,$$

the values $v(t_i)$ are positive but Δt is negative,and we get a negative result:

$$\int_5^0 v(t)\,dt = \left.\frac{-t^3}{3} + \frac{5}{2}t^2\right|_5^0 = 0 - \frac{-5^3}{3} - \frac{5}{2}5^2 = -\frac{125}{6}.$$

Finally we note one simple property of integrals:

$$\int_a^b f(x) + g(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$$

This is easy to understand once you recall that $(F(x)+G(x))' = F'(x)+G'(x)$. Hence, if $F'(x) = f(x)$ and $G'(x) = g(x)$, then

$$\begin{aligned}\int_a^b f(x) + g(x)\,dx &= \left.(F(x)+G(x))\right|_a^b \\ &= F(b) + G(b) - F(a) - G(a) \\ &= F(b) - F(a) + G(b) - G(a) \\ &= \left.F(x)\right|_a^b + \left.G(x)\right|_a^b \\ &= \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.\end{aligned}$$

In summary, we will frequently use these properties of integrals:

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

$$\int_a^b f(x) + g(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$$

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$$

and if $a < b$ and $f(x) \le 0$ on $[a,b]$ then

$$\int_a^b f(x)\,dx \le 0$$

and in fact

$$\int_a^b f(x)\,dx = -\int_a^b |f(x)|\,dx.$$

Exercises 7.3.

1. An object moves so that its velocity at time t is $v(t) = -9.8t + 20$ m/s. Describe the motion of the object between $t = 0$ and $t = 5$, find the total distance traveled by the object during that time, and find the net distance traveled. ⇒

2. An object moves so that its velocity at time t is $v(t) = \sin t$. Set up and evaluate a single definite integral to compute the net distance traveled between $t = 0$ and $t = 2\pi$. ⇒

3. An object moves so that its velocity at time t is $v(t) = 1 + 2\sin t$ m/s. Find the net distance traveled by the object between $t = 0$ and $t = 2\pi$, and find the total distance traveled during the same period. ⇒

4. Consider the function $f(x) = (x+2)(x+1)(x-1)(x-2)$ on $[-2, 2]$. Find the total area between the curve and the x-axis (measuring all area as positive). ⇒

5. Consider the function $f(x) = x^2 - 3x + 2$ on $[0, 4]$. Find the total area between the curve and the x-axis (measuring all area as positive). ⇒

6. Evaluate the three integrals:

$$A = \int_0^3 (-x^2+9)\,dx \qquad B = \int_0^4 (-x^2+9)\,dx \qquad C = \int_4^3 (-x^2+9)\,dx,$$

and verify that $A = B + C$. ⇒

8

Techniques of Integration

Over the next few sections we examine some techniques that are frequently successful when seeking antiderivatives of functions. Sometimes this is a simple problem, since it will be apparent that the function you wish to integrate is a derivative in some straightforward way. For example, faced with

$$\int x^{10}\,dx$$

we realize immediately that the derivative of x^{11} will supply an x^{10}: $(x^{11})' = 11x^{10}$. We don't want the "11", but constants are easy to alter, because differentiation "ignores" them in certain circumstances, so

$$\frac{d}{dx}\frac{1}{11}x^{11} = \frac{1}{11}11x^{10} = x^{10}.$$

From our knowledge of derivatives, we can immediately write down a number of antiderivatives. Here is a list of those most often used:

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C, \quad \text{if } n \neq -1$$

$$\int x^{-1}\,dx = \ln|x| + C$$

$$\int e^x\,dx = e^x + C$$

$$\int \sin x\,dx = -\cos x + C$$

$$\int \cos x\,dx = \sin x + C$$

$$\int \sec^2 x\,dx = \tan x + C$$

$$\int \sec x \tan x\,dx = \sec x + C$$

$$\int \frac{1}{1+x^2}\,dx = \arctan x + C$$

$$\int \frac{1}{\sqrt{1-x^2}}\,dx = \arcsin x + C$$

8.1 Substitution

Needless to say, most problems we encounter will not be so simple. Here's a slightly more complicated example: find

$$\int 2x\cos(x^2)\,dx.$$

This is not a "simple" derivative, but a little thought reveals that it must have come from an application of the chain rule. Multiplied on the "outside" is $2x$, which is the derivative of the "inside" function x^2. Checking:

$$\frac{d}{dx}\sin(x^2) = \cos(x^2)\frac{d}{dx}x^2 = 2x\cos(x^2),$$

so

$$\int 2x\cos(x^2)\,dx = \sin(x^2) + C.$$

Even when the chain rule has "produced" a certain derivative, it is not always easy to see. Consider this problem:

$$\int x^3\sqrt{1-x^2}\,dx.$$

There are two factors in this expression, x^3 and $\sqrt{1-x^2}$, but it is not apparent that the chain rule is involved. Some clever rearrangement reveals that it is:

$$\int x^3\sqrt{1-x^2}\,dx = \int (-2x)\left(-\frac{1}{2}\right)(1-(1-x^2))\sqrt{1-x^2}\,dx.$$

This looks messy, but we do now have something that looks like the result of the chain rule: the function $1-x^2$ has been substituted into $-(1/2)(1-x)\sqrt{x}$, and the derivative of $1-x^2$, $-2x$, multiplied on the outside. If we can find a function $F(x)$ whose derivative is $-(1/2)(1-x)\sqrt{x}$ we'll be done, since then

$$\begin{aligned}\frac{d}{dx}F(1-x^2) = -2xF'(1-x^2) &= (-2x)\left(-\frac{1}{2}\right)(1-(1-x^2))\sqrt{1-x^2}\\ &= x^3\sqrt{1-x^2}\end{aligned}$$

But this isn't hard:

$$\begin{aligned}\int -\frac{1}{2}(1-x)\sqrt{x}\,dx &= \int -\frac{1}{2}(x^{1/2}-x^{3/2})\,dx \qquad (8.1.1)\\ &= -\frac{1}{2}\left(\frac{2}{3}x^{3/2}-\frac{2}{5}x^{5/2}\right)+C\\ &= \left(\frac{1}{5}x-\frac{1}{3}\right)x^{3/2}+C.\end{aligned}$$

So finally we have

$$\int x^3\sqrt{1-x^2}\,dx = \left(\frac{1}{5}(1-x^2)-\frac{1}{3}\right)(1-x^2)^{3/2}+C.$$

So we succeeded, but it required a clever first step, rewriting the original function so that it looked like the result of using the chain rule. Fortunately, there is a technique that makes such problems simpler, without requiring cleverness to rewrite a function in just the right way. It does sometimes not work, or may require more than one attempt, but the idea is simple: guess at the most likely candidate for the "inside function", then do some algebra to see what this requires the rest of the function to look like.

One frequently good guess is any complicated expression inside a square root, so we start by trying $u = 1-x^2$, using a new variable, u, for convenience in the manipulations that follow. Now we know that the chain rule will multiply by the derivative of this inner function:

$$\frac{du}{dx} = -2x,$$

so we need to rewrite the original function to include this:

$$\int x^3\sqrt{1-x^2} = \int x^3\sqrt{u}\frac{-2x}{-2x}\,dx = \int \frac{x^2}{-2}\sqrt{u}\,\frac{du}{dx}\,dx.$$

Recall that one benefit of the Leibniz notation is that it often turns out that what looks like ordinary arithmetic gives the correct answer, even if something more complicated is going on. For example, in Leibniz notation the chain rule is

$$\frac{dy}{dx} = \frac{dy}{dt}\frac{dt}{dx}.$$

The same is true of our current expression:

$$\int \frac{x^2}{-2}\sqrt{u}\,\frac{du}{dx}\,dx = \int \frac{x^2}{-2}\sqrt{u}\,du.$$

Now we're almost there: since $u = 1 - x^2$, $x^2 = 1 - u$ and the integral is

$$\int -\frac{1}{2}(1-u)\sqrt{u}\,du.$$

It's no coincidence that this is exactly the integral we computed in (8.1.1), we have simply renamed the variable u to make the calculations less confusing. Just as before:

$$\int -\frac{1}{2}(1-u)\sqrt{u}\,du = \left(\frac{1}{5}u - \frac{1}{3}\right)u^{3/2} + C.$$

Then since $u = 1 - x^2$:

$$\int x^3\sqrt{1-x^2}\,dx = \left(\frac{1}{5}(1-x^2) - \frac{1}{3}\right)(1-x^2)^{3/2} + C.$$

To summarize: if we suspect that a given function is the derivative of another via the chain rule, we let u denote a likely candidate for the inner function, then translate the given function so that it is written entirely in terms of u, with no x remaining in the expression. If we can integrate this new function of u, then the antiderivative of the original function is obtained by replacing u by the equivalent expression in x.

Even in simple cases you may prefer to use this mechanical procedure, since it often helps to avoid silly mistakes. For example, consider again this simple problem:

$$\int 2x\cos(x^2)\,dx.$$

Let $u = x^2$, then $du/dx = 2x$ or $du = 2x\,dx$. Since we have exactly $2x\,dx$ in the original integral, we can replace it by du:

$$\int 2x\cos(x^2)\,dx = \int \cos u\,du = \sin u + C = \sin(x^2) + C.$$

This is not the only way to do the algebra, and typically there are many paths to the correct answer. Another possibility, for example, is: Since $du/dx = 2x$, $dx = du/2x$, and then the integral becomes

$$\int 2x\cos(x^2)\,dx = \int 2x\cos u\,\frac{du}{2x} = \int \cos u\,du.$$

The important thing to remember is that you must eliminate all instances of the original variable x.

EXAMPLE 8.1.1 Evaluate $\int (ax+b)^n\,dx$, assuming that a and b are constants, $a \neq 0$, and n is a positive integer. We let $u = ax+b$ so $du = a\,dx$ or $dx = du/a$. Then

$$\int (ax+b)^n\,dx = \int \frac{1}{a}u^n\,du = \frac{1}{a(n+1)}u^{n+1} + C = \frac{1}{a(n+1)}(ax+b)^{n+1} + C.$$

□

EXAMPLE 8.1.2 Evaluate $\int \sin(ax+b)\,dx$, assuming that a and b are constants and $a \neq 0$. Again we let $u = ax+b$ so $du = a\,dx$ or $dx = du/a$. Then

$$\int \sin(ax+b)\,dx = \int \frac{1}{a}\sin u\,du = \frac{1}{a}(-\cos u) + C = -\frac{1}{a}\cos(ax+b) + C.$$

□

EXAMPLE 8.1.3 Evaluate $\int_2^4 x\sin(x^2)\,dx$. First we compute the antiderivative, then evaluate the definite integral. Let $u = x^2$ so $du = 2x\,dx$ or $x\,dx = du/2$. Then

$$\int x\sin(x^2)\,dx = \int \frac{1}{2}\sin u\,du = \frac{1}{2}(-\cos u) + C = -\frac{1}{2}\cos(x^2) + C.$$

Now

$$\int_2^4 x\sin(x^2)\,dx = -\frac{1}{2}\cos(x^2)\Big|_2^4 = -\frac{1}{2}\cos(16) + \frac{1}{2}\cos(4).$$

A somewhat neater alternative to this method is to change the original limits to match the variable u. Since $u = x^2$, when $x = 2$, $u = 4$, and when $x = 4$, $u = 16$. So we can do this:

$$\int_2^4 x\sin(x^2)\,dx = \int_4^{16} \frac{1}{2}\sin u\,du = -\frac{1}{2}(\cos u)\Big|_4^{16} = -\frac{1}{2}\cos(16) + \frac{1}{2}\cos(4).$$

An incorrect, and dangerous, alternative is something like this:

$$\int_2^4 x\sin(x^2)\,dx = \int_2^4 \frac{1}{2}\sin u\,du = -\frac{1}{2}\cos(u)\Big|_2^4 = -\frac{1}{2}\cos(x^2)\Big|_2^4 = -\frac{1}{2}\cos(16) + \frac{1}{2}\cos(4).$$

This is incorrect because $\int_2^4 \frac{1}{2}\sin u\,du$ means that u takes on values between 2 and 4, which is wrong. It is dangerous, because it is very easy to get to the point $-\frac{1}{2}\cos(u)\Big|_2^4$ and forget to substitute x^2 back in for u, thus getting the incorrect answer $-\frac{1}{2}\cos(4)+\frac{1}{2}\cos(2)$. A somewhat clumsy, but acceptable, alternative is something like this:

$$\int_2^4 x\sin(x^2)\,dx = \int_{x=2}^{x=4}\frac{1}{2}\sin u\,du = -\frac{1}{2}\cos(u)\Big|_{x=2}^{x=4} = -\frac{1}{2}\cos(x^2)\Big|_2^4 = -\frac{\cos(16)}{2}+\frac{\cos(4)}{2}.$$

□

EXAMPLE 8.1.4 Evaluate $\displaystyle\int_{1/4}^{1/2}\frac{\cos(\pi t)}{\sin^2(\pi t)}\,dt$. Let $u=\sin(\pi t)$ so $du=\pi\cos(\pi t)\,dt$ or $du/\pi=\cos(\pi t)\,dt$. We change the limits to $\sin(\pi/4)=\sqrt{2}/2$ and $\sin(\pi/2)=1$. Then

$$\int_{1/4}^{1/2}\frac{\cos(\pi t)}{\sin^2(\pi t)}\,dt = \int_{\sqrt{2}/2}^{1}\frac{1}{\pi}\frac{1}{u^2}\,du = \int_{\sqrt{2}/2}^{1}\frac{1}{\pi}u^{-2}\,du = \frac{1}{\pi}\frac{u^{-1}}{-1}\Big|_{\sqrt{2}/2}^{1} = -\frac{1}{\pi}+\frac{\sqrt{2}}{\pi}.$$

□

Exercises 8.1.

Find the antiderivatives or evaluate the definite integral in each problem.

1. $\displaystyle\int (1-t)^9\,dt \Rightarrow$

2. $\displaystyle\int (x^2+1)^2\,dx \Rightarrow$

3. $\displaystyle\int x(x^2+1)^{100}\,dx \Rightarrow$

4. $\displaystyle\int \frac{1}{\sqrt[3]{1-5t}}\,dt \Rightarrow$

5. $\displaystyle\int \sin^3 x\cos x\,dx \Rightarrow$

6. $\displaystyle\int x\sqrt{100-x^2}\,dx \Rightarrow$

7. $\displaystyle\int \frac{x^2}{\sqrt{1-x^3}}\,dx \Rightarrow$

8. $\displaystyle\int \cos(\pi t)\cos\big(\sin(\pi t)\big)\,dt \Rightarrow$

9. $\displaystyle\int \frac{\sin x}{\cos^3 x}\,dx \Rightarrow$

10. $\displaystyle\int \tan x\,dx \Rightarrow$

11. $\displaystyle\int_0^{\pi} \sin^5(3x)\cos(3x)\,dx \Rightarrow$

12. $\displaystyle\int \sec^2 x\tan x\,dx \Rightarrow$

13. $\displaystyle\int_0^{\sqrt{\pi}/2} x\sec^2(x^2)\tan(x^2)\,dx \Rightarrow$

14. $\displaystyle\int \frac{\sin(\tan x)}{\cos^2 x}\,dx \Rightarrow$

15. $\displaystyle\int_3^4 \frac{1}{(3x-7)^2}\,dx \Rightarrow$

16. $\displaystyle\int_0^{\pi/6} (\cos^2 x-\sin^2 x)\,dx \Rightarrow$

17. $\displaystyle\int \frac{6x}{(x^2-7)^{1/9}}\,dx \Rightarrow$

18. $\displaystyle\int_{-1}^{1} (2x^3-1)(x^4-2x)^6\,dx \Rightarrow$

19. $\displaystyle\int_{-1}^{1} \sin^7 x\,dx \Rightarrow$

20. $\displaystyle\int f(x)f'(x)\,dx \Rightarrow$

8.2 Powers of sine and cosine

Functions consisting of products of the sine and cosine can be integrated by using substitution and trigonometric identities. These can sometimes be tedious, but the technique is straightforward. Some examples will suffice to explain the approach.

EXAMPLE 8.2.1 Evaluate $\int \sin^5 x\,dx$. Rewrite the function:

$$\int \sin^5 x\,dx = \int \sin x \sin^4 x\,dx = \int \sin x(\sin^2 x)^2\,dx = \int \sin x(1-\cos^2 x)^2\,dx.$$

Now use $u = \cos x$, $du = -\sin x\,dx$:

$$\begin{aligned}\int \sin x(1-\cos^2 x)^2\,dx &= \int -(1-u^2)^2\,du\\ &= \int -(1-2u^2+u^4)\,du\\ &= -u + \frac{2}{3}u^3 - \frac{1}{5}u^5 + C\\ &= -\cos x + \frac{2}{3}\cos^3 x - \frac{1}{5}\cos^5 x + C.\end{aligned}$$

□

EXAMPLE 8.2.2 Evaluate $\int \sin^6 x\,dx$. Use $\sin^2 x = (1-\cos(2x))/2$ to rewrite the function:

$$\begin{aligned}\int \sin^6 x\,dx = \int (\sin^2 x)^3\,dx &= \int \frac{(1-\cos 2x)^3}{8}\,dx\\ &= \frac{1}{8}\int 1 - 3\cos 2x + 3\cos^2 2x - \cos^3 2x\,dx.\end{aligned}$$

Now we have four integrals to evaluate:

$$\int 1\,dx = x$$

and

$$\int -3\cos 2x\,dx = -\frac{3}{2}\sin 2x$$

are easy. The $\cos^3 2x$ integral is like the previous example:

$$\begin{aligned}
\int -\cos^3 2x\,dx &= \int -\cos 2x\cos^2 2x\,dx \\
&= \int -\cos 2x(1-\sin^2 2x)\,dx \\
&= \int -\frac{1}{2}(1-u^2)\,du \\
&= -\frac{1}{2}\left(u-\frac{u^3}{3}\right) \\
&= -\frac{1}{2}\left(\sin 2x-\frac{\sin^3 2x}{3}\right).
\end{aligned}$$

And finally we use another trigonometric identity, $\cos^2 x = (1+\cos(2x))/2$:

$$\int 3\cos^2 2x\,dx = 3\int \frac{1+\cos 4x}{2}\,dx = \frac{3}{2}\left(x+\frac{\sin 4x}{4}\right).$$

So at long last we get

$$\int \sin^6 x\,dx = \frac{x}{8} - \frac{3}{16}\sin 2x - \frac{1}{16}\left(\sin 2x - \frac{\sin^3 2x}{3}\right) + \frac{3}{16}\left(x+\frac{\sin 4x}{4}\right) + C.$$

□

EXAMPLE 8.2.3 Evaluate $\int \sin^2 x\cos^2 x\,dx$. Use the formulas $\sin^2 x = (1-\cos(2x))/2$ and $\cos^2 x = (1+\cos(2x))/2$ to get:

$$\int \sin^2 x\cos^2 x\,dx = \int \frac{1-\cos(2x)}{2}\cdot\frac{1+\cos(2x)}{2}\,dx.$$

The remainder is left as an exercise. □

Exercises 8.2.

Find the antiderivatives.

1. $\int \sin^2 x\,dx \Rightarrow$

2. $\int \sin^3 x\,dx \Rightarrow$

3. $\int \sin^4 x\,dx \Rightarrow$

4. $\int \cos^2 x\sin^3 x\,dx \Rightarrow$

5. $\int \cos^3 x\,dx \Rightarrow$

6. $\int \sin^2 x\cos^2 x\,dx \Rightarrow$

7. $\int \cos^3 x\sin^2 x\,dx \Rightarrow$

8. $\int \sin x(\cos x)^{3/2}\,dx \Rightarrow$

9. $\int \sec^2 x\csc^2 x\,dx \Rightarrow$

10. $\int \tan^3 x\sec x\,dx \Rightarrow$

8.3 Trigonometric Substitutions

So far we have seen that it sometimes helps to replace a subexpression of a function by a single variable. Occasionally it can help to replace the original variable by something more complicated. This seems like a "reverse" substitution, but it is really no different in principle than ordinary substitution.

EXAMPLE 8.3.1 Evaluate $\int \sqrt{1-x^2}\,dx$. Let $x = \sin u$ so $dx = \cos u\,du$. Then

$$\int \sqrt{1-x^2}\,dx = \int \sqrt{1-\sin^2 u}\cos u\,du = \int \sqrt{\cos^2 u}\cos u\,du.$$

We would like to replace $\sqrt{\cos^2 u}$ by $\cos u$, but this is valid only if $\cos u$ is positive, since $\sqrt{\cos^2 u}$ is positive. Consider again the substitution $x = \sin u$. We could just as well think of this as $u = \arcsin x$. If we do, then by the definition of the arcsine, $-\pi/2 \le u \le \pi/2$, so $\cos u \ge 0$. Then we continue:

$$\begin{aligned}\int \sqrt{\cos^2 u}\cos u\,du = \int \cos^2 u\,du = \int \frac{1+\cos 2u}{2}\,du &= \frac{u}{2} + \frac{\sin 2u}{4} + C \\ &= \frac{\arcsin x}{2} + \frac{\sin(2\arcsin x)}{4} + C.\end{aligned}$$

This is a perfectly good answer, though the term $\sin(2\arcsin x)$ is a bit unpleasant. It is possible to simplify this. Using the identity $\sin 2x = 2\sin x\cos x$, we can write $\sin 2u = 2\sin u\cos u = 2\sin(\arcsin x)\sqrt{1-\sin^2 u} = 2x\sqrt{1-\sin^2(\arcsin x)} = 2x\sqrt{1-x^2}$. Then the full antiderivative is

$$\frac{\arcsin x}{2} + \frac{2x\sqrt{1-x^2}}{4} = \frac{\arcsin x}{2} + \frac{x\sqrt{1-x^2}}{2} + C.$$

□

This type of substitution is usually indicated when the function you wish to integrate contains a polynomial expression that might allow you to use the fundamental identity $\sin^2 x + \cos^2 x = 1$ in one of three forms:

$$\cos^2 x = 1 - \sin^2 x \qquad \sec^2 x = 1 + \tan^2 x \qquad \tan^2 x = \sec^2 x - 1.$$

If your function contains $1-x^2$, as in the example above, try $x = \sin u$; if it contains $1+x^2$ try $x = \tan u$; and if it contains $x^2 - 1$, try $x = \sec u$. Sometimes you will need to try something a bit different to handle constants other than one.

EXAMPLE 8.3.2 Evaluate $\int \sqrt{4-9x^2}\,dx$. We start by rewriting this so that it looks more like the previous example:

$$\int \sqrt{4-9x^2}\,dx = \int \sqrt{4(1-(3x/2)^2)}\,dx = \int 2\sqrt{1-(3x/2)^2}\,dx.$$

Now let $3x/2 = \sin u$ so $(3/2)\,dx = \cos u\,du$ or $dx = (2/3)\cos u\,du$. Then

$$\begin{aligned}\int 2\sqrt{1-(3x/2)^2}\,dx = \int 2\sqrt{1-\sin^2 u}\,(2/3)\cos u\,du &= \frac{4}{3}\int \cos^2 u\,du\\ &= \frac{4u}{6} + \frac{4\sin 2u}{12} + C\\ &= \frac{2\arcsin(3x/2)}{3} + \frac{2\sin u\cos u}{3} + C\\ &= \frac{2\arcsin(3x/2)}{3} + \frac{2\sin(\arcsin(3x/2))\cos(\arcsin(3x/2))}{3} + C\\ &= \frac{2\arcsin(3x/2)}{3} + \frac{2(3x/2)\sqrt{1-(3x/2)^2}}{3} + C\\ &= \frac{2\arcsin(3x/2)}{3} + \frac{x\sqrt{4-9x^2}}{2} + C,\end{aligned}$$

using some of the work from example 8.3.1. □

EXAMPLE 8.3.3 Evaluate $\int \sqrt{1+x^2}\,dx$. Let $x = \tan u$, $dx = \sec^2 u\,du$, so

$$\int \sqrt{1+x^2}\,dx = \int \sqrt{1+\tan^2 u}\sec^2 u\,du = \int \sqrt{\sec^2 u}\sec^2 u\,du.$$

Since $u = \arctan(x)$, $-\pi/2 \le u \le \pi/2$ and $\sec u \ge 0$, so $\sqrt{\sec^2 u} = \sec u$. Then

$$\int \sqrt{\sec^2 u}\sec^2 u\,du = \int \sec^3 u\,du.$$

In problems of this type, two integrals come up frequently: $\int \sec^3 u\,du$ and $\int \sec u\,du$. Both have relatively nice expressions but they are a bit tricky to discover.

First we do $\int \sec u\,du$, which we will need to compute $\displaystyle\int \sec^3 u\,du$:

$$\begin{aligned}\int \sec u\,du &= \int \sec u\,\frac{\sec u + \tan u}{\sec u + \tan u}\,du\\ &= \int \frac{\sec^2 u + \sec u \tan u}{\sec u + \tan u}\,du.\end{aligned}$$

Now let $w = \sec u + \tan u$, $dw = \sec u \tan u + \sec^2 u\,du$, exactly the numerator of the function we are integrating. Thus

$$\begin{aligned}\int \sec u\,du = \int \frac{\sec^2 u + \sec u \tan u}{\sec u + \tan u}\,du &= \int \frac{1}{w}\,dw = \ln|w| + C\\ &= \ln|\sec u + \tan u| + C.\end{aligned}$$

Now for $\displaystyle\int \sec^3 u\,du$:

$$\begin{aligned}\sec^3 u &= \frac{\sec^3 u}{2} + \frac{\sec^3 u}{2} = \frac{\sec^3 u}{2} + \frac{(\tan^2 u + 1)\sec u}{2}\\ &= \frac{\sec^3 u}{2} + \frac{\sec u \tan^2 u}{2} + \frac{\sec u}{2} = \frac{\sec^3 u + \sec u \tan^2 u}{2} + \frac{\sec u}{2}.\end{aligned}$$

We already know how to integrate $\sec u$, so we just need the first quotient. This is "simply" a matter of recognizing the product rule in action:

$$\int \sec^3 u + \sec u \tan^2 u\,du = \sec u \tan u.$$

So putting these together we get

$$\int \sec^3 u\,du = \frac{\sec u \tan u}{2} + \frac{\ln|\sec u + \tan u|}{2} + C,$$

and reverting to the original variable x:

$$\begin{aligned}\int \sqrt{1+x^2}\,dx &= \frac{\sec u \tan u}{2} + \frac{\ln|\sec u + \tan u|}{2} + C\\ &= \frac{\sec(\arctan x)\tan(\arctan x)}{2} + \frac{\ln|\sec(\arctan x) + \tan(\arctan x)|}{2} + C\\ &= \frac{x\sqrt{1+x^2}}{2} + \frac{\ln|\sqrt{1+x^2} + x|}{2} + C,\end{aligned}$$

using $\tan(\arctan x) = x$ and $\sec(\arctan x) = \sqrt{1 + \tan^2(\arctan x)} = \sqrt{1+x^2}$. □

Exercises 8.3.

Find the antiderivatives.

1. $\int \csc x\,dx \Rightarrow$

2. $\int \csc^3 x\,dx \Rightarrow$

3. $\int \sqrt{x^2-1}\,dx \Rightarrow$

4. $\int \sqrt{9+4x^2}\,dx \Rightarrow$

5. $\int x\sqrt{1-x^2}\,dx \Rightarrow$

6. $\int x^2\sqrt{1-x^2}\,dx \Rightarrow$

7. $\int \frac{1}{\sqrt{1+x^2}}\,dx \Rightarrow$

8. $\int \sqrt{x^2+2x}\,dx \Rightarrow$

9. $\int \frac{1}{x^2(1+x^2)}\,dx \Rightarrow$

10. $\int \frac{x^2}{\sqrt{4-x^2}}\,dx \Rightarrow$

11. $\int \frac{\sqrt{x}}{\sqrt{1-x}}\,dx \Rightarrow$

12. $\int \frac{x^3}{\sqrt{4x^2-1}}\,dx \Rightarrow$

8.4 Integration by Parts

We have already seen that recognizing the product rule can be useful, when we noticed that

$$\int \sec^3 u + \sec u \tan^2 u\,du = \sec u \tan u.$$

As with substitution, we do not have to rely on insight or cleverness to discover such antiderivatives; there is a technique that will often help to uncover the product rule.

Start with the product rule:

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

We can rewrite this as

$$f(x)g(x) = \int f'(x)g(x)\,dx + \int f(x)g'(x)\,dx,$$

and then

$$\int f(x)g'(x)\,dx = f(x)g(x) - \int f'(x)g(x)\,dx.$$

This may not seem particularly useful at first glance, but it turns out that in many cases we have an integral of the form

$$\int f(x)g'(x)\,dx$$

but that

$$\int f'(x)g(x)\,dx$$

is easier. This technique for turning one integral into another is called **integration by parts**, and is usually written in more compact form. If we let $u = f(x)$ and $v = g(x)$ then

$du = f'(x)\,dx$ and $dv = g'(x)\,dx$ and

$$\int u\,dv = uv - \int v\,du.$$

To use this technique we need to identify likely candidates for $u = f(x)$ and $dv = g'(x)\,dx$.

EXAMPLE 8.4.1 Evaluate $\int x \ln x\,dx$. Let $u = \ln x$ so $du = 1/x\,dx$. Then we must let $dv = x\,dx$ so $v = x^2/2$ and

$$\int x \ln x\,dx = \frac{x^2 \ln x}{2} - \int \frac{x^2}{2}\frac{1}{x}\,dx = \frac{x^2 \ln x}{2} - \int \frac{x}{2}\,dx = \frac{x^2 \ln x}{2} - \frac{x^2}{4} + C.$$

□

EXAMPLE 8.4.2 Evaluate $\int x \sin x\,dx$. Let $u = x$ so $du = dx$. Then we must let $dv = \sin x\,dx$ so $v = -\cos x$ and

$$\int x \sin x\,dx = -x \cos x - \int -\cos x\,dx = -x\cos x + \int \cos x\,dx = -x\cos x + \sin x + C.$$

□

EXAMPLE 8.4.3 Evaluate $\int \sec^3 x\,dx$. Of course we already know the answer to this, but we needed to be clever to discover it. Here we'll use the new technique to discover the antiderivative. Let $u = \sec x$ and $dv = \sec^2 x\,dx$. Then $du = \sec x \tan x\,dx$ and $v = \tan x$ and

$$\begin{aligned}\int \sec^3 x\,dx &= \sec x \tan x - \int \tan^2 x \sec x\,dx\\ &= \sec x \tan x - \int (\sec^2 x - 1)\sec x\,dx\\ &= \sec x \tan x - \int \sec^3 x\,dx + \int \sec x\,dx.\end{aligned}$$

At first this looks useless—we're right back to $\int \sec^3 x\,dx$. But looking more closely:

$$\begin{aligned}
\int \sec^3 x\,dx &= \sec x\tan x - \int \sec^3 x\,dx + \int \sec x\,dx\\
\int \sec^3 x\,dx + \int \sec^3 x\,dx &= \sec x\tan x + \int \sec x\,dx\\
2\int \sec^3 x\,dx &= \sec x\tan x + \int \sec x\,dx\\
\int \sec^3 x\,dx &= \frac{\sec x\tan x}{2} + \frac{1}{2}\int \sec x\,dx\\
&= \frac{\sec x\tan x}{2} + \frac{\ln|\sec x+\tan x|}{2} + C.
\end{aligned}$$

□

EXAMPLE 8.4.4 Evaluate $\int x^2\sin x\,dx$. Let $u = x^2$, $dv = \sin x\,dx$; then $du = 2x\,dx$ and $v = -\cos x$. Now $\int x^2\sin x\,dx = -x^2\cos x + \int 2x\cos x\,dx$. This is better than the original integral, but we need to do integration by parts again. Let $u = 2x$, $dv = \cos x\,dx$; then $du = 2$ and $v = \sin x$, and

$$\begin{aligned}
\int x^2\sin x\,dx &= -x^2\cos x + \int 2x\cos x\,dx\\
&= -x^2\cos x + 2x\sin x - \int 2\sin x\,dx\\
&= -x^2\cos x + 2x\sin x + 2\cos x + C.
\end{aligned}$$

□

Such repeated use of integration by parts is fairly common, but it can be a bit tedious to accomplish, and it is easy to make errors, especially sign errors involving the subtraction in the formula. There is a nice tabular method to accomplish the calculation that minimizes the chance for error and speeds up the whole process. We illustrate with the previous example. Here is the table:

sign	u	dv
	x^2	$\sin x$
$-$	$2x$	$-\cos x$
	2	$-\sin x$
$-$	0	$\cos x$

or

u	dv
x^2	$\sin x$
$-2x$	$-\cos x$
2	$-\sin x$
0	$\cos x$

To form the first table, we start with u at the top of the second column and repeatedly compute the derivative; starting with dv at the top of the third column, we repeatedly compute the antiderivative. In the first column, we place a "$-$" in every second row. To form the second table we combine the first and second columns by ignoring the boundary; if you do this by hand, you may simply start with two columns and add a "$-$" to every second row.

To compute with this second table we begin at the top. Multiply the first entry in column u by the second entry in column dv to get $-x^2 \cos x$, and add this to the integral of the product of the second entry in column u and second entry in column dv. This gives:

$$-x^2 \cos x + \int 2x \cos x\, dx,$$

or exactly the result of the first application of integration by parts. Since this integral is not yet easy, we return to the table. Now we multiply twice on the diagonal, $(x^2)(-\cos x)$ and $(-2x)(-\sin x)$ and then once straight across, $(2)(-\sin x)$, and combine these as

$$-x^2 \cos x + 2x \sin x - \int 2 \sin x\, dx,$$

giving the same result as the second application of integration by parts. While this integral is easy, we may return yet once more to the table. Now multiply three times on the diagonal to get $(x^2)(-\cos x)$, $(-2x)(-\sin x)$, and $(2)(\cos x)$, and once straight across, $(0)(\cos x)$. We combine these as before to get

$$-x^2 \cos x + 2x \sin x + 2 \cos x + \int 0\, dx = -x^2 \cos x + 2x \sin x + 2 \cos x + C.$$

Typically we would fill in the table one line at a time, until the "straight across" multiplication gives an easy integral. If we can see that the u column will eventually become zero, we can instead fill in the whole table; computing the products as indicated will then give the entire integral, including the "$+C$", as above.

Exercises 8.4.

Find the antiderivatives.

1. $\int x \cos x\, dx$ ⇒

2. $\int x^2 \cos x\, dx$ ⇒

3. $\int x e^x\, dx$ ⇒

4. $\int x e^{x^2}\, dx$ ⇒

5. $\int \sin^2 x\, dx$ ⇒

6. $\int \ln x\, dx$ ⇒

7. $\int x \arctan x\,dx \Rightarrow$

8. $\int x^3 \sin x\,dx \Rightarrow$

9. $\int x^3 \cos x\,dx \Rightarrow$

10. $\int x \sin^2 x\,dx \Rightarrow$

11. $\int x \sin x \cos x\,dx \Rightarrow$

12. $\int \arctan(\sqrt{x})\,dx \Rightarrow$

13. $\int \sin(\sqrt{x})\,dx \Rightarrow$

14. $\int \sec^2 x \csc^2 x\,dx \Rightarrow$

8.5 Rational Functions

A **rational function** is a fraction with polynomials in the numerator and denominator. For example,

$$\frac{x^3}{x^2+x-6}, \qquad \frac{1}{(x-3)^2}, \qquad \frac{x^2+1}{x^2-1},$$

are all rational functions of x. There is a general technique called "partial fractions" that, in principle, allows us to integrate any rational function. The algebraic steps in the technique are rather cumbersome if the polynomial in the denominator has degree more than 2, and the technique requires that we factor the denominator, something that is not always possible. However, in practice one does not often run across rational functions with high degree polynomials in the denominator for which one has to find the antiderivative function. So we shall explain how to find the antiderivative of a rational function only when the denominator is a quadratic polynomial ax^2+bx+c.

We should mention a special type of rational function that we already know how to integrate: If the denominator has the form $(ax+b)^n$, the substitution $u = ax+b$ will always work. The denominator becomes u^n, and each x in the numerator is replaced by $(u-b)/a$, and $dx = du/a$. While it may be tedious to complete the integration if the numerator has high degree, it is merely a matter of algebra.

EXAMPLE 8.5.1 Find $\displaystyle\int \frac{x^3}{(3-2x)^5}\,dx$. Using the substitution $u = 3 - 2x$ we get

$$\begin{aligned}
\int \frac{x^3}{(3-2x)^5}\,dx &= \frac{1}{-2}\int \frac{\left(\frac{u-3}{-2}\right)^3}{u^5}\,du = \frac{1}{16}\int \frac{u^3 - 9u^2 + 27u - 27}{u^5}\,du \\
&= \frac{1}{16}\int u^{-2} - 9u^{-3} + 27u^{-4} - 27u^{-5}\,du \\
&= \frac{1}{16}\left(\frac{u^{-1}}{-1} - \frac{9u^{-2}}{-2} + \frac{27u^{-3}}{-3} - \frac{27u^{-4}}{-4}\right) + C \\
&= \frac{1}{16}\left(\frac{(3-2x)^{-1}}{-1} - \frac{9(3-2x)^{-2}}{-2} + \frac{27(3-2x)^{-3}}{-3} - \frac{27(3-2x)^{-4}}{-4}\right) + C \\
&= -\frac{1}{16(3-2x)} + \frac{9}{32(3-2x)^2} - \frac{9}{16(3-2x)^3} + \frac{27}{64(3-2x)^4} + C
\end{aligned}$$

□

We now proceed to the case in which the denominator is a quadratic polynomial. We can always factor out the coefficient of x^2 and put it outside the integral, so we can assume that the denominator has the form $x^2 + bx + c$. There are three possible cases, depending on how the quadratic factors: either $x^2 + bx + c = (x - r)(x - s)$, $x^2 + bx + c = (x - r)^2$, or it doesn't factor. We can use the quadratic formula to decide which of these we have, and to factor the quadratic if it is possible.

EXAMPLE 8.5.2 Determine whether $x^2 + x + 1$ factors, and factor it if possible. The quadratic formula tells us that $x^2 + x + 1 = 0$ when

$$x = \frac{-1 \pm \sqrt{1-4}}{2}.$$

Since there is no square root of -3, this quadratic does not factor. □

EXAMPLE 8.5.3 Determine whether $x^2 - x - 1$ factors, and factor it if possible. The quadratic formula tells us that $x^2 - x - 1 = 0$ when

$$x = \frac{1 \pm \sqrt{1+4}}{2} = \frac{1 \pm \sqrt{5}}{2}.$$

Therefore

$$x^2 - x - 1 = \left(x - \frac{1+\sqrt{5}}{2}\right)\left(x - \frac{1-\sqrt{5}}{2}\right).$$

□

If $x^2+bx+c=(x-r)^2$ then we have the special case we have already seen, that can be handled with a substitution. The other two cases require different approaches.

If $x^2+bx+c=(x-r)(x-s)$, we have an integral of the form

$$\int \frac{p(x)}{(x-r)(x-s)}\,dx$$

where $p(x)$ is a polynomial. The first step is to make sure that $p(x)$ has degree less than 2.

EXAMPLE 8.5.4 Rewrite $\displaystyle\int \frac{x^3}{(x-2)(x+3)}\,dx$ in terms of an integral with a numerator that has degree less than 2. To do this we use long division of polynomials to discover that

$$\frac{x^3}{(x-2)(x+3)} = \frac{x^3}{x^2+x-6} = x-1+\frac{7x-6}{x^2+x-6} = x-1+\frac{7x-6}{(x-2)(x+3)},$$

so

$$\int \frac{x^3}{(x-2)(x+3)}\,dx = \int x-1\,dx + \int \frac{7x-6}{(x-2)(x+3)}\,dx.$$

The first integral is easy, so only the second requires some work. □

Now consider the following simple algebra of fractions:

$$\frac{A}{x-r}+\frac{B}{x-s} = \frac{A(x-s)+B(x-r)}{(x-r)(x-s)} = \frac{(A+B)x-As-Br}{(x-r)(x-s)}.$$

That is, adding two fractions with constant numerator and denominators $(x-r)$ and $(x-s)$ produces a fraction with denominator $(x-r)(x-s)$ and a polynomial of degree less than 2 for the numerator. We want to reverse this process: starting with a single fraction, we want to write it as a sum of two simpler fractions. An example should make it clear how to proceed.

EXAMPLE 8.5.5 Evaluate $\displaystyle\int \frac{x^3}{(x-2)(x+3)}\,dx$. We start by writing $\displaystyle\frac{7x-6}{(x-2)(x+3)}$ as the sum of two fractions. We want to end up with

$$\frac{7x-6}{(x-2)(x+3)} = \frac{A}{x-2}+\frac{B}{x+3}.$$

If we go ahead and add the fractions on the right hand side we get

$$\frac{7x-6}{(x-2)(x+3)} = \frac{(A+B)x+3A-2B}{(x-2)(x+3)}.$$

So all we need to do is find A and B so that $7x-6=(A+B)x+3A-2B$, which is to say, we need $7=A+B$ and $-6=3A-2B$. This is a problem you've seen before: solve a system of two equations in two unknowns. There are many ways to proceed; here's one: If $7 = A+B$ then $B = 7-A$ and so $-6 = 3A-2B = 3A-2(7-A) = 3A-14+2A = 5A-14$. This is easy to solve for A: $A = 8/5$, and then $B = 7 - A = 7 - 8/5 = 27/5$. Thus

$$\int \frac{7x-6}{(x-2)(x+3)}\,dx = \int \frac{8}{5}\frac{1}{x-2} + \frac{27}{5}\frac{1}{x+3}\,dx = \frac{8}{5}\ln|x-2| + \frac{27}{5}\ln|x+3| + C.$$

The answer to the original problem is now

$$\begin{aligned}\int \frac{x^3}{(x-2)(x+3)}\,dx &= \int x-1\,dx + \int \frac{7x-6}{(x-2)(x+3)}\,dx \\ &= \frac{x^2}{2} - x + \frac{8}{5}\ln|x-2| + \frac{27}{5}\ln|x+3| + C.\end{aligned}$$

□

Now suppose that x^2+bx+c doesn't factor. Again we can use long division to ensure that the numerator has degree less than 2, then we complete the square.

EXAMPLE 8.5.6 Evaluate $\displaystyle\int \frac{x+1}{x^2+4x+8}\,dx$. The quadratic denominator does not factor. We could complete the square and use a trigonometric substitution, but it is simpler to rearrange the integrand:

$$\int \frac{x+1}{x^2+4x+8}\,dx = \int \frac{x+2}{x^2+4x+8}\,dx - \int \frac{1}{x^2+4x+8}\,dx.$$

The first integral is an easy substitution problem, using $u = x^2+4x+8$:

$$\int \frac{x+2}{x^2+4x+8}\,dx = \frac{1}{2}\int \frac{du}{u} = \frac{1}{2}\ln|x^2+4x+8|.$$

For the second integral we complete the square:

$$x^2+4x+8 = (x+2)^2 + 4 = 4\left(\left(\frac{x+2}{2}\right)^2 + 1\right),$$

making the integral

$$\frac{1}{4}\int \frac{1}{\left(\frac{x+2}{2}\right)^2 + 1}\,dx.$$

Using $u = \dfrac{x+2}{2}$ we get

$$\frac{1}{4}\int \frac{1}{\left(\frac{x+2}{2}\right)^2 + 1}\,dx = \frac{1}{4}\int \frac{2}{u^2+1}\,du = \frac{1}{2}\arctan\left(\frac{x+2}{2}\right).$$

The final answer is now

$$\int \frac{x+1}{x^2+4x+8}\,dx = \frac{1}{2}\ln|x^2+4x+8| - \frac{1}{2}\arctan\left(\frac{x+2}{2}\right) + C.$$

□

Exercises 8.5.

Find the antiderivatives.

1. $\displaystyle\int \frac{1}{4-x^2}\,dx \Rightarrow$

2. $\displaystyle\int \frac{x^4}{4-x^2}\,dx \Rightarrow$

3. $\displaystyle\int \frac{1}{x^2+10x+25}\,dx \Rightarrow$

4. $\displaystyle\int \frac{x^2}{4-x^2}\,dx \Rightarrow$

5. $\displaystyle\int \frac{x^4}{4+x^2}\,dx \Rightarrow$

6. $\displaystyle\int \frac{1}{x^2+10x+29}\,dx \Rightarrow$

7. $\displaystyle\int \frac{x^3}{4+x^2}\,dx \Rightarrow$

8. $\displaystyle\int \frac{1}{x^2+10x+21}\,dx \Rightarrow$

9. $\displaystyle\int \frac{1}{2x^2-x-3}\,dx \Rightarrow$

10. $\displaystyle\int \frac{1}{x^2+3x}\,dx \Rightarrow$

8.6 Numerical Integration

We have now seen some of the most generally useful methods for discovering antiderivatives, and there are others. Unfortunately, some functions have no simple antiderivatives; in such cases if the value of a definite integral is needed it will have to be approximated. We will see two methods that work reasonably well and yet are fairly simple; in some cases more sophisticated techniques will be needed.

Of course, we already know one way to approximate an integral: if we think of the integral as computing an area, we can add up the areas of some rectangles. While this is quite simple, it is usually the case that a large number of rectangles is needed to get acceptable accuracy. A similar approach is much better: we approximate the area under a curve over a small interval as the area of a trapezoid. In figure 8.6.1 we see an area under a curve approximated by rectangles and by trapezoids; it is apparent that the trapezoids give a substantially better approximation on each subinterval.

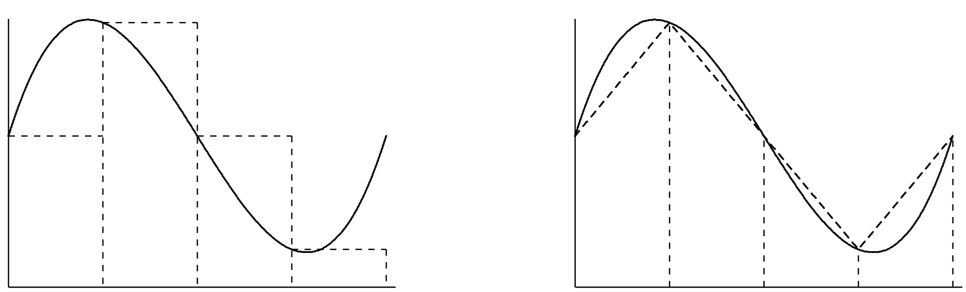


Figure 8.6.1 Approximating an area with rectangles and with trapezoids.

As with rectangles, we divide the interval into n equal subintervals of length Δx. A typical trapezoid is pictured in figure 8.6.2; it has area $\dfrac{f(x_i)+f(x_{i+1})}{2}\Delta x$. If we add up the areas of all trapezoids we get

$$\frac{f(x_0)+f(x_1)}{2}\Delta x+\frac{f(x_1)+f(x_2)}{2}\Delta x+\cdots+\frac{f(x_{n-1})+f(x_n)}{2}\Delta x=$$
$$\left(\frac{f(x_0)}{2}+f(x_1)+f(x_2)+\cdots+f(x_{n-1})+\frac{f(x_n)}{2}\right)\Delta x.$$

This is usually known as the **Trapezoid Rule**. For a modest number of subintervals this is not too difficult to do with a calculator; a computer can easily do many subintervals.

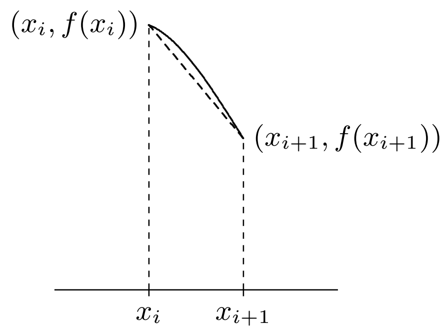


Figure 8.6.2 A single trapezoid.

In practice, an approximation is useful only if we know how accurate it is; for example, we might need a particular value accurate to three decimal places. When we compute a particular approximation to an integral, the error is the difference between the approximation and the true value of the integral. For any approximation technique, we need an **error estimate**, a value that is guaranteed to be larger than the actual error. If A is an approximation and E is the associated error estimate, then we know that the true value of the integral is between $A-E$ and $A+E$. In the case of our approximation of the integral, we want $E=E(\Delta x)$ to be a function of Δx that gets small rapidly as Δx gets small. Fortunately, for many functions, there is such an error estimate associated with the trapezoid approximation.

THEOREM 8.6.1 Suppose f has a second derivative f'' everywhere on the interval $[a,b]$, and $|f''(x)|\le M$ for all x in the interval. With $\Delta x=(b-a)/n$, an error estimate for the trapezoid approximation is

$$E(\Delta x)=\frac{b-a}{12}M(\Delta x)^2=\frac{(b-a)^3}{12n^2}M.$$

■

Let's see how we can use this.

EXAMPLE 8.6.2 Approximate $\int_0^1 e^{-x^2}\,dx$ to two decimal places. The second derivative of $f = e^{-x^2}$ is $(4x^2-2)e^{-x^2}$, and it is not hard to see that on $[0,1]$, $|(4x^2-2)e^{-x^2}| \le 2$. We begin by estimating the number of subintervals we are likely to need. To get two decimal places of accuracy, we will certainly need $E(\Delta x) < 0.005$ or

$$\begin{aligned}\frac{1}{12}(2)\frac{1}{n^2} &< 0.005\\ \frac{1}{6}(200) &< n^2\\ 5.77 \approx \sqrt{\frac{100}{3}} &< n\end{aligned}$$

With $n = 6$, the error estimate is thus $1/6^3 < 0.0047$. We compute the trapezoid approximation for six intervals:

$$\left(\frac{f(0)}{2} + f(1/6) + f(2/6) + \cdots + f(5/6) + \frac{f(1)}{2}\right)\frac{1}{6} \approx 0.74512.$$

So the true value of the integral is between $0.74512 - 0.0047 = 0.74042$ and $0.74512 + 0.0047 = 0.74982$. Unfortunately, the first rounds to 0.74 and the second rounds to 0.75, so we can't be sure of the correct value in the second decimal place; we need to pick a larger n. As it turns out, we need to go to $n = 12$ to get two bounds that both round to the same value, which turns out to be 0.75. For comparison, using 12 rectangles to approximate the area gives 0.7727, which is considerably less accurate than the approximation using six trapezoids.

In practice it generally pays to start by requiring better than the maximum possible error; for example, we might have initially required $E(\Delta x) < 0.001$, or

$$\begin{aligned}\frac{1}{12}(2)\frac{1}{n^2} &< 0.001\\ \frac{1}{6}(1000) &< n^2\\ 12.91 \approx \sqrt{\frac{500}{3}} &< n\end{aligned}$$

Had we immediately tried $n = 13$ this would have given us the desired answer. □

The trapezoid approximation works well, especially compared to rectangles, because the tops of the trapezoids form a reasonably good approximation to the curve when Δx is fairly small. We can extend this idea: what if we try to approximate the curve more closely, by using something other than a straight line? The obvious candidate is a parabola: if we can approximate a short piece of the curve with a parabola with equation $y = ax^2 + bx + c$, we can easily compute the area under the parabola.

There are an infinite number of parabolas through any two given points, but only one through three given points. If we find a parabola through three consecutive points $(x_i, f(x_i))$, $(x_{i+1}, f(x_{i+1}))$, $(x_{i+2}, f(x_{i+2}))$ on the curve, it should be quite close to the curve over the whole interval $[x_i, x_{i+2}]$, as in figure 8.6.3. If we divide the interval $[a, b]$ into an even number of subintervals, we can then approximate the curve by a sequence of parabolas, each covering two of the subintervals. For this to be practical, we would like a simple formula for the area under one parabola, namely, the parabola through $(x_i, f(x_i))$, $(x_{i+1}, f(x_{i+1}))$, and $(x_{i+2}, f(x_{i+2}))$. That is, we should attempt to write down the parabola $y = ax^2 + bx + c$ through these points and then integrate it, and hope that the result is fairly simple. Although the algebra involved is messy, this turns out to be possible. The algebra is well within the capability of a good computer algebra system like Sage, so we will present the result without all of the algebra; you can see how to do it in this Sage worksheet.

To find the parabola, we solve these three equations for a, b, and c:

$$\begin{aligned}
f(x_i) &= a(x_{i+1} - \Delta x)^2 + b(x_{i+1} - \Delta x) + c \\
f(x_{i+1}) &= a(x_{i+1})^2 + b(x_{i+1}) + c \\
f(x_{i+2}) &= a(x_{i+1} + \Delta x)^2 + b(x_{i+1} + \Delta x) + c
\end{aligned}$$

Not surprisingly, the solutions turn out to be quite messy. Nevertheless, Sage can easily compute and simplify the integral to get

$$\int_{x_{i+1}-\Delta x}^{x_{i+1}+\Delta x} ax^2 + bx + c\,dx = \frac{\Delta x}{3}(f(x_i) + 4f(x_{i+1}) + f(x_{i+2})).$$

Now the sum of the areas under all parabolas is

$$\frac{\Delta x}{3}(f(x_0) + 4f(x_1) + f(x_2) + f(x_2) + 4f(x_3) + f(x_4) + \cdots + f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)) =$$
$$\frac{\Delta x}{3}(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \cdots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)).$$

This is just slightly more complicated than the formula for trapezoids; we need to remember the alternating 2 and 4 coefficients; note that n must be even for this to make sense. This approximation technique is referred to as **Simpson's Rule**.

As with the trapezoid method, this is useful only with an error estimate:

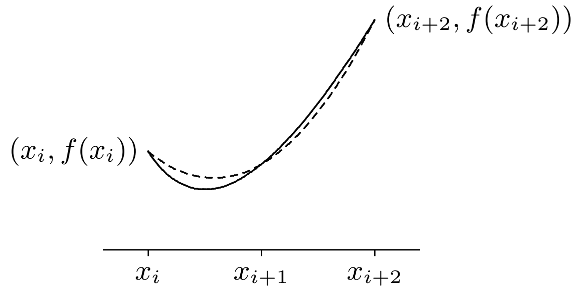


Figure 8.6.3 A parabola (dashed) approximating a curve (solid).

THEOREM 8.6.3 Suppose f has a fourth derivative $f^{(4)}$ everywhere on the interval $[a, b]$, and $|f^{(4)}(x)| \le M$ for all x in the interval. With $\Delta x = (b-a)/n$, an error estimate for Simpson's approximation is

$$E(\Delta x) = \frac{b-a}{180} M(\Delta x)^4 = \frac{(b-a)^5}{180n^4} M.$$

■

EXAMPLE 8.6.4 Let us again approximate $\int_0^1 e^{-x^2}\,dx$ to two decimal places. The fourth derivative of $f = e^{-x^2}$ is $(16x^2 - 48x^2 + 12)e^{-x^2}$; on $[0, 1]$ this is at most 12 in absolute value. We begin by estimating the number of subintervals we are likely to need. To get two decimal places of accuracy, we will certainly need $E(\Delta x) < 0.005$, but taking a cue from our earlier example, let's require $E(\Delta x) < 0.001$:

$$\begin{aligned} \frac{1}{180}(12)\frac{1}{n^4} &< 0.001 \\ \frac{200}{3} &< n^4 \\ 2.86 \approx \sqrt[\text{[4]}]{\frac{200}{3}} &< n \end{aligned}$$

So we try $n = 4$, since we need an even number of subintervals. Then the error estimate is $12/180/4^4 < 0.0003$ and the approximation is

$$(f(0) + 4f(1/4) + 2f(1/2) + 4f(3/4) + f(1))\frac{1}{3 \cdot 4} \approx 0.746855.$$

So the true value of the integral is between $0.746855 - 0.0003 = 0.746555$ and $0.746855 + 0.0003 = 0.7471555$, both of which round to 0.75. □

Exercises 8.6.

In the following problems, compute the trapezoid and Simpson approximations using 4 subintervals, and compute the error estimate for each. (Finding the maximum values of the second and fourth derivatives can be challenging for some of these; you may use a graphing calculator or computer software to estimate the maximum values.) If you have access to Sage or similar software, approximate each integral to two decimal places. You can use this Sage worksheet to get started.

1. $\int_1^3 x\,dx \Rightarrow$

2. $\int_0^3 x^2\,dx \Rightarrow$

3. $\int_2^4 x^3\,dx \Rightarrow$

4. $\int_1^3 \frac{1}{x}\,dx \Rightarrow$

5. $\int_1^2 \frac{1}{1+x^2}\,dx \Rightarrow$

6. $\int_0^1 x\sqrt{1+x}\,dx \Rightarrow$

7. $\int_1^5 \frac{x}{1+x}\,dx \Rightarrow$

8. $\int_0^1 \sqrt{x^3+1}\,dx \Rightarrow$

9. $\int_0^1 \sqrt{x^4+1}\,dx \Rightarrow$

10. $\int_1^4 \sqrt{1+1/x}\,dx \Rightarrow$

11. Using Simpson's rule on a parabola $f(x)$, even with just two subintervals, gives the exact value of the integral, because the parabolas used to approximate f will be f itself. Remarkably, Simpson's rule also computes the integral of a cubic function $f(x) = ax^3 + bx^2 + cx + d$ exactly. Show this is true by showing that

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{x_2 - x_0}{3 \cdot 2}(f(x_0) + 4f((x_0 + x_2)/2) + f(x_2)).$$

This does require a bit of messy algebra, so you may prefer to use Sage.

8.7 Additional exercises

These problems require the techniques of this chapter, and are in no particular order. Some problems may be done in more than one way.

1. $\int (t+4)^3\,dt \Rightarrow$

2. $\int t(t^2-9)^{3/2}\,dt \Rightarrow$

3. $\int (e^{t^2}+16)te^{t^2}\,dt \Rightarrow$

4. $\int \sin t \cos 2t\,dt \Rightarrow$

5. $\int \tan t \sec^2 t\,dt \Rightarrow$

6. $\int \frac{2t+1}{t^2+t+3}\,dt \Rightarrow$

7. $\int \frac{1}{t(t^2-4)}\,dt \Rightarrow$

8. $\int \frac{1}{(25-t^2)^{3/2}}\,dt \Rightarrow$

9. $\int \frac{\cos 3t}{\sqrt{\sin 3t}}\,dt \Rightarrow$

10. $\int t \sec^2 t\,dt \Rightarrow$

11. $\int \frac{e^t}{\sqrt{e^t+1}}\,dt \Rightarrow$

12. $\int \cos^4 t\,dt \Rightarrow$

13. $\displaystyle\int \frac{1}{t^2+3t}\,dt \Rightarrow$

14. $\displaystyle\int \frac{1}{t^2\sqrt{1+t^2}}\,dt \Rightarrow$

15. $\displaystyle\int \frac{\sec^2 t}{(1+\tan t)^3}\,dt \Rightarrow$

16. $\displaystyle\int t^3\sqrt{t^2+1}\,dt \Rightarrow$

17. $\displaystyle\int e^t \sin t\,dt \Rightarrow$

18. $\displaystyle\int (t^{3/2}+47)^3\sqrt{t}\,dt \Rightarrow$

19. $\displaystyle\int \frac{t^3}{(2-t^2)^{5/2}}\,dt \Rightarrow$

20. $\displaystyle\int \frac{1}{t(9+4t^2)}\,dt \Rightarrow$

21. $\displaystyle\int \frac{\arctan 2t}{1+4t^2}\,dt \Rightarrow$

22. $\displaystyle\int \frac{t}{t^2+2t-3}\,dt \Rightarrow$

23. $\displaystyle\int \sin^3 t\cos^4 t\,dt \Rightarrow$

24. $\displaystyle\int \frac{1}{t^2-6t+9}\,dt \Rightarrow$

25. $\displaystyle\int \frac{1}{t(\ln t)^2}\,dt \Rightarrow$

26. $\displaystyle\int t(\ln t)^2\,dt \Rightarrow$

27. $\displaystyle\int t^3 e^t\,dt \Rightarrow$

28. $\displaystyle\int \frac{t+1}{t^2+t-1}\,dt \Rightarrow$

9

Applications of Integration

9.1 Area between curves

We have seen how integration can be used to find an area between a curve and the x-axis. With very little change we can find some areas between curves; indeed, the area between a curve and the x-axis may be interpreted as the area between the curve and a second "curve" with equation $y = 0$. In the simplest of cases, the idea is quite easy to understand.

EXAMPLE 9.1.1 Find the area below $f(x) = -x^2 + 4x + 3$ and above $g(x) = -x^3 + 7x^2 - 10x + 5$ over the interval $1 \le x \le 2$. In figure 9.1.1 we show the two curves together, with the desired area shaded, then f alone with the area under f shaded, and then g alone with the area under g shaded.

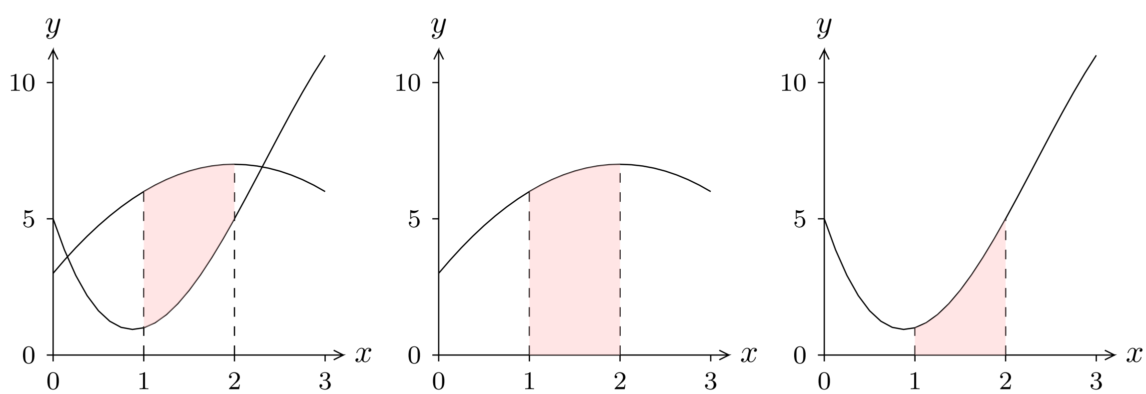


Figure 9.1.1 Area between curves as a difference of areas.

It is clear from the figure that the area we want is the area under f minus the area under g, which is to say

$$\int_1^2 f(x)\,dx - \int_1^2 g(x)\,dx = \int_1^2 f(x) - g(x)\,dx.$$

It doesn't matter whether we compute the two integrals on the left and then subtract or compute the single integral on the right. In this case, the latter is perhaps a bit easier:

$$\begin{aligned}
\int_1^2 f(x) - g(x)\,dx &= \int_1^2 -x^2 + 4x + 3 - (-x^3 + 7x^2 - 10x + 5)\,dx \\
&= \int_1^2 x^3 - 8x^2 + 14x - 2\,dx \\
&= \left.\frac{x^4}{4} - \frac{8x^3}{3} + 7x^2 - 2x\right|_1^2 \\
&= \frac{16}{4} - \frac{64}{3} + 28 - 4 - (\frac{1}{4} - \frac{8}{3} + 7 - 2) \\
&= 23 - \frac{56}{3} - \frac{1}{4} = \frac{49}{12}.
\end{aligned}$$

□

It is worth examining this problem a bit more. We have seen one way to look at it, by viewing the desired area as a big area minus a small area, which leads naturally to the difference between two integrals. But it is instructive to consider how we might find the desired area directly. We can approximate the area by dividing the area into thin sections and approximating the area of each section by a rectangle, as indicated in figure 9.1.2. The area of a typical rectangle is $\Delta x(f(x_i) - g(x_i))$, so the total area is approximately

$$\sum_{i=0}^{n-1}(f(x_i) - g(x_i))\Delta x.$$

This is exactly the sort of sum that turns into an integral in the limit, namely the integral

$$\int_1^2 f(x) - g(x)\,dx.$$

Of course, this is the integral we actually computed above, but we have now arrived at it directly rather than as a modification of the difference between two other integrals. In that example it really doesn't matter which approach we take, but in some cases this second approach is better.

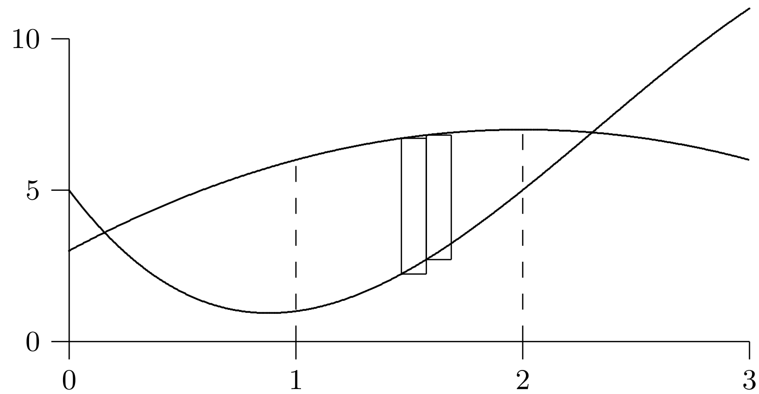


Figure 9.1.2 Approximating area between curves with rectangles.

EXAMPLE 9.1.2 Find the area below $f(x) = -x^2 + 4x + 1$ and above $g(x) = -x^3 + 7x^2 - 10x + 3$ over the interval $1 \le x \le 2$; these are the same curves as before but lowered by 2. In figure 9.1.3 we show the two curves together. Note that the lower curve now dips below the x-axis. This makes it somewhat tricky to view the desired area as a big area minus a smaller area, but it is just as easy as before to think of approximating the area by rectangles. The height of a typical rectangle will still be $f(x_i) - g(x_i)$, even if $g(x_i)$ is negative. Thus the area is

$$\int_1^2 -x^2 + 4x + 1 - (-x^3 + 7x^2 - 10x + 3)\,dx = \int_1^2 x^3 - 8x^2 + 14x - 2\,dx.$$

This is of course the same integral as before, because the region between the curves is identical to the former region—it has just been moved down by 2. □

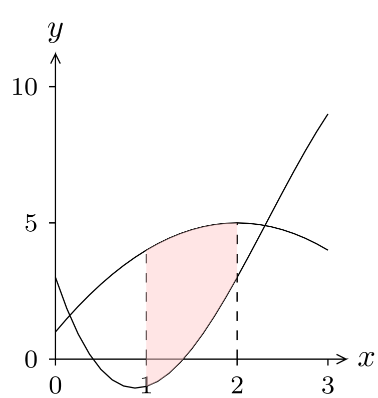


Figure 9.1.3 Area between curves.

EXAMPLE 9.1.3 Find the area between $f(x) = -x^2 + 4x$ and $g(x) = x^2 - 6x + 5$ over the interval $0 \le x \le 1$; the curves are shown in figure 9.1.4. Generally we should interpret "area" in the usual sense, as a necessarily positive quantity. Since the two curves cross, we need to compute two areas and add them. First we find the intersection point of the curves:

$$\begin{aligned} -x^2+4x &= x^2-6x+5 \\ 0 &= 2x^2-10x+5 \\ x &= \frac{10\pm\sqrt{100-40}}{4} = \frac{5\pm\sqrt{15}}{2}. \end{aligned}$$

The intersection point we want is $x = a = (5-\sqrt{15})/2$. Then the total area is

$$\begin{aligned} \int_0^a x^2-6x+5-(-x^2+4x)\,dx + \int_a^1 -x^2+4x-(x^2-6x+5)\,dx \\ = \int_0^a 2x^2-10x+5\,dx + \int_a^1 -2x^2+10x-5\,dx \\ = \left.\frac{2x^3}{3}-5x^2+5x\right|_0^a + \left.-\frac{2x^3}{3}+5x^2-5x\right|_a^1 \\ = -\frac{52}{3}+5\sqrt{15}, \end{aligned}$$

after a bit of simplification. □

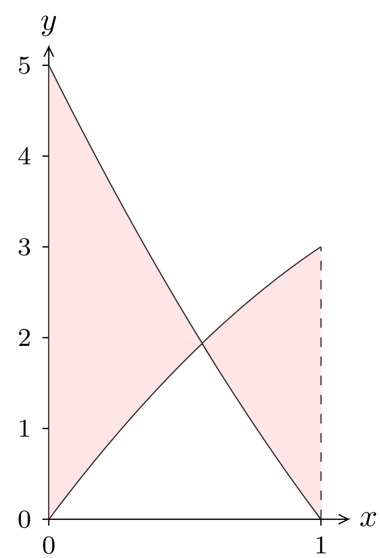


Figure 9.1.4 Area between curves that cross.

EXAMPLE 9.1.4 Find the area between $f(x) = -x^2+4x$ and $g(x) = x^2-6x+5$; the curves are shown in figure 9.1.5. Here we are not given a specific interval, so it must be the case that there is a "natural" region involved. Since the curves are both parabolas, the only reasonable interpretation is the region between the two intersection points, which we found in the previous example:

$$\frac{5 \pm \sqrt{15}}{2}.$$

If we let $a = (5 - \sqrt{15})/2$ and $b = (5 + \sqrt{15})/2$, the total area is

$$\begin{aligned}\int_a^b -x^2 + 4x - (x^2 - 6x + 5)\, dx &= \int_a^b -2x^2 + 10x - 5\, dx \\ &= -\frac{2x^3}{3} + 5x^2 - 5x\bigg|_a^b \\ &= 5\sqrt{15}.\end{aligned}$$

after a bit of simplification. □

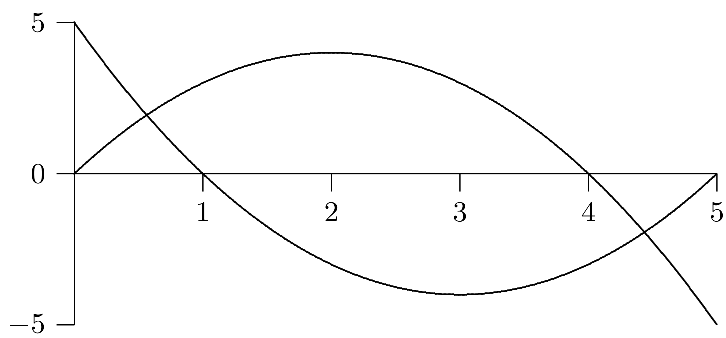


Figure 9.1.5 Area bounded by two curves.

Exercises 9.1.

Find the area bounded by the curves.

1. $y = x^4 - x^2$ and $y = x^2$ (the part to the right of the y-axis) ⇒

2. $x = y^3$ and $x = y^2$ ⇒

3. $x = 1 - y^2$ and $y = -x - 1$ ⇒

4. $x = 3y - y^2$ and $x + y = 3$ ⇒

5. $y = \cos(\pi x/2)$ and $y = 1 - x^2$ (in the first quadrant) ⇒

6. $y = \sin(\pi x/3)$ and $y = x$ (in the first quadrant) ⇒

7. $y = \sqrt{x}$ and $y = x^2$ ⇒

8. $y = \sqrt{x}$ and $y = \sqrt{x+1}$, $0 \le x \le 4$ ⇒

9. $x = 0$ and $x = 25 - y^2$ ⇒

10. $y = \sin x \cos x$ and $y = \sin x$, $0 \le x \le \pi$ ⇒

11. $y = x^{3/2}$ and $y = x^{2/3}$ ⇒

12. $y = x^2 - 2x$ and $y = x - 2$ ⇒

The following three exercises expand on the geometric interpretation of the hyperbolic functions. Refer to section 4.11 and particularly to figure 4.11.2 and exercise 6 in section 4.11.

13. Compute $\int \sqrt{x^2 - 1}\,dx$ using the substitution $u = \operatorname{arccosh} x$, or $x = \cosh u$; use exercise 6 in section 4.11.

14. Fix $t > 0$. Sketch the region R in the right half plane bounded by the curves $y = x \tanh t$, $y = -x \tanh t$, and $x^2 - y^2 = 1$. Note well: t is fixed, the plane is the x-y plane.

15. Prove that the area of R is t.

9.2 Distance, Velocity, Acceleration

We next recall a general principle that will later be applied to distance-velocity-acceleration problems, among other things. If $F(u)$ is an anti-derivative of $f(u)$, then $\int_a^b f(u)\,du = F(b) - F(a)$. Suppose that we want to let the upper limit of integration vary, i.e., we replace b by some variable x. We think of a as a fixed starting value x_0. In this new notation the last equation (after adding $F(a)$ to both sides) becomes:

$$F(x) = F(x_0) + \int_{x_0}^{x} f(u)\,du.$$

(Here u is the variable of integration, called a "dummy variable," since it is not the variable in the function $F(x)$. In general, it is not a good idea to use the same letter as a variable of integration and as a limit of integration. That is, $\int_{x_0}^{x} f(x)dx$ is bad notation, and can lead to errors and confusion.)

An important application of this principle occurs when we are interested in the position of an object at time t (say, on the x-axis) and we know its position at time t_0. Let $s(t)$ denote the position of the object at time t (its distance from a reference point, such as the origin on the x-axis). Then the net change in position between t_0 and t is $s(t) - s(t_0)$. Since $s(t)$ is an anti-derivative of the velocity function $v(t)$, we can write

$$s(t) = s(t_0) + \int_{t_0}^{t} v(u)du.$$

Similarly, since the velocity is an anti-derivative of the acceleration function $a(t)$, we have

$$v(t) = v(t_0) + \int_{t_0}^{t} a(u)du.$$

EXAMPLE 9.2.1 Suppose an object is acted upon by a constant force F. Find $v(t)$ and $s(t)$. By Newton's law $F = ma$, so the acceleration is F/m, where m is the mass of the object. Then we first have

$$v(t) = v(t_0) + \int_{t_0}^{t} \frac{F}{m}\, du = v_0 + \frac{F}{m}u\Big|_{t_0}^{t} = v_0 + \frac{F}{m}(t - t_0),$$

using the usual convention $v_0 = v(t_0)$. Then

$$\begin{aligned} s(t) = s(t_0) + \int_{t_0}^{t} \left(v_0 + \frac{F}{m}(u - t_0)\right) du &= s_0 + (v_0 u + \frac{F}{2m}(u - t_0)^2)\Big|_{t_0}^{t} \\ &= s_0 + v_0(t - t_0) + \frac{F}{2m}(t - t_0)^2. \end{aligned}$$

For instance, when $F/m = -g$ is the constant of gravitational acceleration, then this is the falling body formula (if we neglect air resistance) familiar from elementary physics:

$$s_0 + v_0(t - t_0) - \frac{g}{2}(t - t_0)^2,$$

or in the common case that $t_0 = 0$,

$$s_0 + v_0 t - \frac{g}{2}t^2.$$

□

Recall that the integral of the velocity function gives the *net* distance traveled. If you want to know the *total* distance traveled, you must find out where the velocity function crosses the t-axis, integrate separately over the time intervals when $v(t)$ is positive and when $v(t)$ is negative, and add up the absolute values of the different integrals. For example, if an object is thrown straight upward at 19.6 m/sec, its velocity function is $v(t) = -9.8t + 19.6$, using $g = 9.8$ m/sec for the force of gravity. This is a straight line which is positive for $t < 2$ and negative for $t > 2$. The net distance traveled in the first 4 seconds is thus

$$\int_0^4 (-9.8t + 19.6)dt = 0,$$

while the total distance traveled in the first 4 seconds is

$$\int_0^2 (-9.8t + 19.6)dt + \left|\int_2^4 (-9.8t + 19.6)dt\right| = 19.6 + |-19.6| = 39.2$$

meters, 19.6 meters up and 19.6 meters down.

EXAMPLE 9.2.2 The acceleration of an object is given by $a(t) = \cos(\pi t)$, and its velocity at time $t = 0$ is $1/(2\pi)$. Find both the net and the total distance traveled in the first 1.5 seconds.

We compute

$$v(t) = v(0) + \int_0^t \cos(\pi u)du = \frac{1}{2\pi} + \frac{1}{\pi}\sin(\pi u)\Big|_0^t = \frac{1}{\pi}\left(\frac{1}{2} + \sin(\pi t)\right).$$

The *net* distance traveled is then

$$\begin{aligned} s(3/2) - s(0) &= \int_0^{3/2} \frac{1}{\pi}\left(\frac{1}{2} + \sin(\pi t)\right) dt \\ &= \frac{1}{\pi}\left(\frac{t}{2} - \frac{1}{\pi}\cos(\pi t)\right)\Big|_0^{3/2} = \frac{3}{4\pi} + \frac{1}{\pi^2} \approx 0.340 \text{ meters.} \end{aligned}$$

To find the *total* distance traveled, we need to know when $(0.5 + \sin(\pi t))$ is positive and when it is negative. This function is 0 when $\sin(\pi t)$ is -0.5, i.e., when $\pi t = 7\pi/6$, $11\pi/6$, etc. The value $\pi t = 7\pi/6$, i.e., $t = 7/6$, is the only value in the range $0 \le t \le 1.5$. Since $v(t) > 0$ for $t < 7/6$ and $v(t) < 0$ for $t > 7/6$, the total distance traveled is

$$\begin{aligned} &\int_0^{7/6} \frac{1}{\pi}\left(\frac{1}{2} + \sin(\pi t)\right) dt + \left|\int_{7/6}^{3/2} \frac{1}{\pi}\left(\frac{1}{2} + \sin(\pi t)\right) dt\right| \\ &\quad = \frac{1}{\pi}\left(\frac{7}{12} + \frac{1}{\pi}\cos(7\pi/6) + \frac{1}{\pi}\right) + \frac{1}{\pi}\left|\frac{3}{4} - \frac{7}{12} + \frac{1}{\pi}\cos(7\pi/6)\right| \\ &\quad = \frac{1}{\pi}\left(\frac{7}{12} + \frac{1}{\pi}\frac{\sqrt{3}}{2} + \frac{1}{\pi}\right) + \frac{1}{\pi}\left|\frac{3}{4} - \frac{7}{12} + \frac{1}{\pi}\frac{\sqrt{3}}{2}.\right| \approx 0.409 \text{ meters.} \end{aligned}$$

□

Exercises 9.2.

For each velocity function find both the net distance and the total distance traveled during the indicated time interval (graph $v(t)$ to determine when it's positive and when it's negative):

1. $v = \cos(\pi t)$, $0 \le t \le 2.5$ ⇒

2. $v = -9.8t + 49$, $0 \le t \le 10$ ⇒

3. $v = 3(t - 3)(t - 1)$, $0 \le t \le 5$ ⇒

4. $v = \sin(\pi t/3) - t$, $0 \le t \le 1$ ⇒

5. An object is shot upwards from ground level with an initial velocity of 2 meters per second; it is subject only to the force of gravity (no air resistance). Find its maximum altitude and the time at which it hits the ground. ⇒

6. An object is shot upwards from ground level with an initial velocity of 3 meters per second; it is subject only to the force of gravity (no air resistance). Find its maximum altitude and the time at which it hits the ground. ⇒

7. An object is shot upwards from ground level with an initial velocity of 100 meters per second; it is subject only to the force of gravity (no air resistance). Find its maximum altitude and the time at which it hits the ground. ⇒

8. An object moves along a straight line with acceleration given by $a(t) = -\cos(t)$, and $s(0) = 1$ and $v(0) = 0$. Find the maximum distance the object travels from zero, and find its maximum speed. Describe the motion of the object. ⇒

9. An object moves along a straight line with acceleration given by $a(t) = \sin(\pi t)$. Assume that when $t = 0$, $s(t) = v(t) = 0$. Find $s(t)$, $v(t)$, and the maximum speed of the object. Describe the motion of the object. ⇒

10. An object moves along a straight line with acceleration given by $a(t) = 1 + \sin(\pi t)$. Assume that when $t = 0$, $s(t) = v(t) = 0$. Find $s(t)$ and $v(t)$. ⇒

11. An object moves along a straight line with acceleration given by $a(t) = 1 - \sin(\pi t)$. Assume that when $t = 0$, $s(t) = v(t) = 0$. Find $s(t)$ and $v(t)$. ⇒

9.3 Volume

We have seen how to compute certain areas by using integration; some volumes may also be computed by evaluating an integral. Generally, the volumes that we can compute this way have cross-sections that are easy to describe.

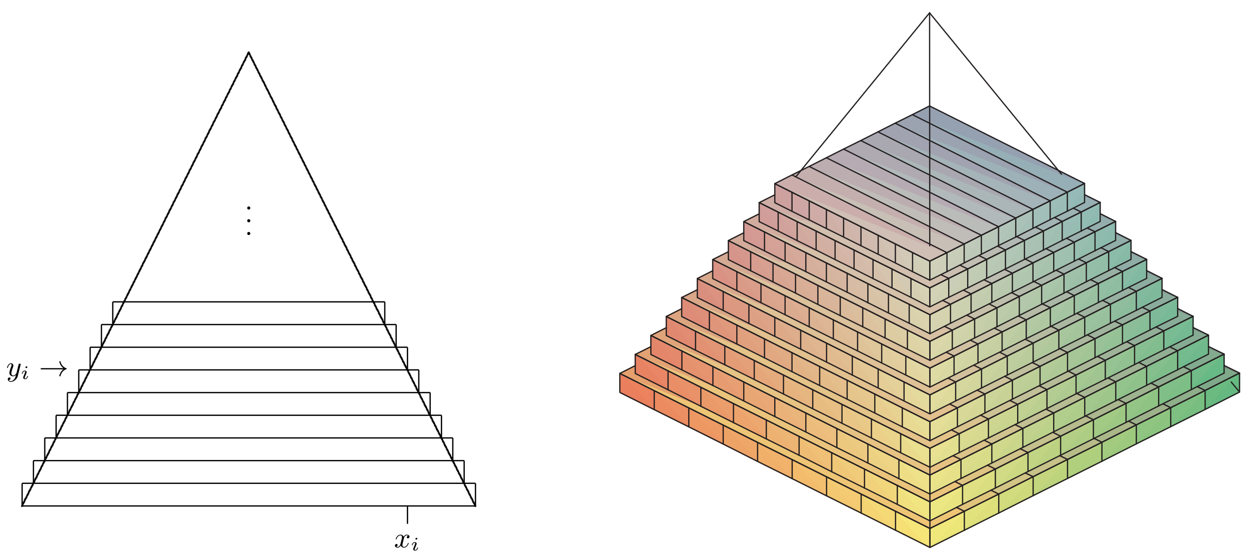


Figure 9.3.1 Volume of a pyramid approximated by rectangular prisms. (AP)

EXAMPLE 9.3.1 Find the volume of a pyramid with a square base that is 20 meters tall and 20 meters on a side at the base. As with most of our applications of integration, we begin by asking how we might approximate the volume. Since we can easily compute the volume of a rectangular prism (that is, a "box"), we will use some boxes to approximate the volume of the pyramid, as shown in figure 9.3.1: on the left is a cross-sectional view, on the right is a 3D view of part of the pyramid with some of the boxes used to approximate the volume.

Each box has volume of the form $(2x_i)(2x_i)\Delta y$. Unfortunately, there are two variables here; fortunately, we can write x in terms of y: $x = 10 - y/2$ or $x_i = 10 - y_i/2$. Then the total volume is approximately

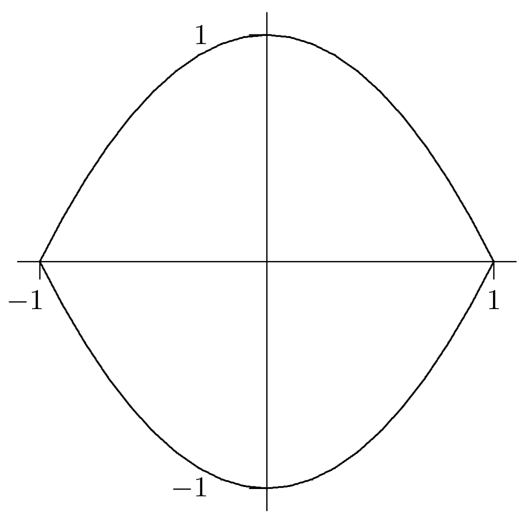
$$\sum_{i=0}^{n-1} 4(10 - y_i/2)^2 \Delta y$$

and in the limit we get the volume as the value of an integral:

$$\int_0^{20} 4(10 - y/2)^2\, dy = \int_0^{20} (20 - y)^2\, dy = -\left.\frac{(20-y)^3}{3}\right|_0^{20} = -\frac{0^3}{3} - -\frac{20^3}{3} = \frac{8000}{3}.$$

As you may know, the volume of a pyramid is (1/3)(height)(area of base) = (1/3)(20)(400), which agrees with our answer. □

EXAMPLE 9.3.2 The base of a solid is the region between $f(x) = x^2 - 1$ and $g(x) = -x^2 + 1$, and its cross-sections perpendicular to the x-axis are equilateral triangles, as indicated in figure 9.3.2. The solid has been truncated to show a triangular cross-section above $x = 1/2$. Find the volume of the solid.



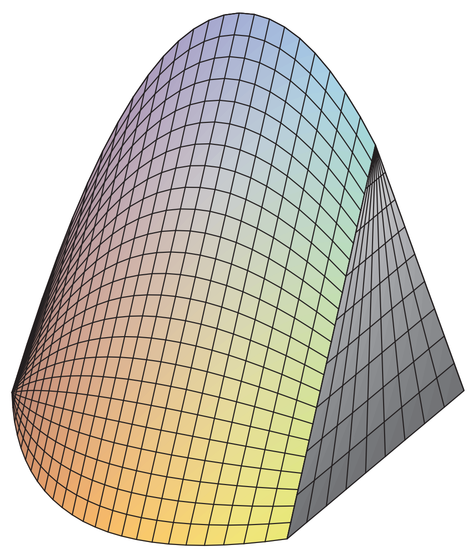


Figure 9.3.2 Solid with equilateral triangles as cross-sections. (AP)

A cross-section at a value x_i on the x-axis is a triangle with base $2(1 - x_i^2)$ and height $\sqrt{3}(1 - x_i^2)$, so the area of the cross-section is

$$\frac{1}{2}(\text{base})(\text{height}) = (1 - x_i^2)\sqrt{3}(1 - x_i^2),$$

and the volume of a thin "slab" is then

$$(1 - x_i^2)\sqrt{3}(1 - x_i^2)\Delta x.$$

Thus the total volume is

$$\int_{-1}^{1} \sqrt{3}(1 - x^2)^2\, dx = \frac{16}{15}\sqrt{3}.$$

□

One easy way to get "nice" cross-sections is by rotating a plane figure around a line. For example, in figure 9.3.3 we see a plane region under a curve and between two vertical lines; then the result of rotating this around the x-axis, and a typical circular cross-section.

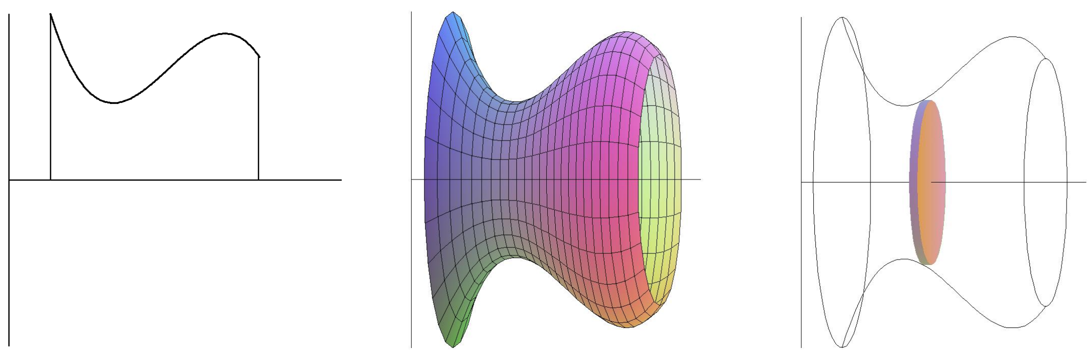


Figure 9.3.3 A solid of rotation. (AP)

Of course a real "slice" of this figure will not have straight sides, but we can approximate the volume of the slice by a cylinder or disk with circular top and bottom and straight sides; the volume of this disk will have the form $\pi r^2 \Delta x$. As long as we can write r in terms of x we can compute the volume by an integral.

EXAMPLE 9.3.3 Find the volume of a right circular cone with base radius 10 and height 20. (A right circular cone is one with a circular base and with the tip of the cone directly over the center of the base.) We can view this cone as produced by the rotation of the line $y = x/2$ rotated about the x-axis, as indicated in figure 9.3.4.

At a particular point on the x-axis, say x_i, the radius of the resulting cone is the y-coordinate of the corresponding point on the line, namely $y_i = x_i/2$. Thus the total volume is approximately

$$\sum_{i=0}^{n-1} \pi (x_i/2)^2\, dx$$

and the exact volume is

$$\int_0^{20} \pi \frac{x^2}{4}\, dx = \frac{\pi}{4}\frac{20^3}{3} = \frac{2000\pi}{3}.$$

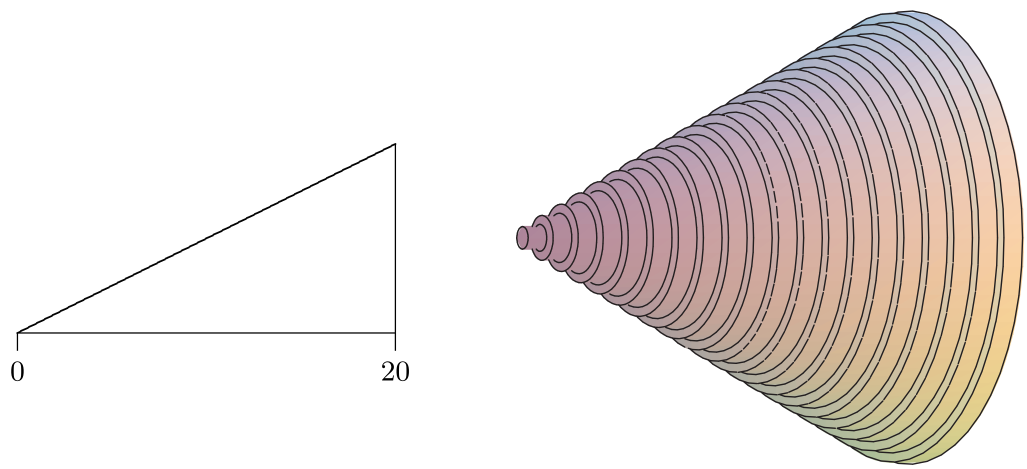


Figure 9.3.4 A region that generates a cone; approximating the volume by circular disks. (AP)

Note that we can instead do the calculation with a generic height and radius:

$$\int_0^h \pi \frac{r^2}{h^2} x^2\,dx = \frac{\pi r^2}{h^2}\frac{h^3}{3} = \frac{\pi r^2 h}{3},$$

giving us the usual formula for the volume of a cone. □

EXAMPLE 9.3.4 Find the volume of the object generated when the area between $y = x^2$ and $y = x$ is rotated around the x-axis. This solid has a "hole" in the middle; we can compute the volume by subtracting the volume of the hole from the volume enclosed by the outer surface of the solid. In figure 9.3.5 we show the region that is rotated, the resulting solid with the front half cut away, the cone that forms the outer surface, the horn-shaped hole, and a cross-section perpendicular to the x-axis.

We have already computed the volume of a cone; in this case it is $\pi/3$. At a particular value of x, say x_i, the cross-section of the horn is a circle with radius x_i^2, so the volume of the horn is

$$\int_0^1 \pi (x^2)^2\,dx = \int_0^1 \pi x^4\,dx = \pi\frac{1}{5},$$

so the desired volume is $\pi/3 - \pi/5 = 2\pi/15$.

As with the area between curves, there is an alternate approach that computes the desired volume "all at once" by approximating the volume of the actual solid. We can approximate the volume of a slice of the solid with a washer-shaped volume, as indicated in figure 9.3.5.

The volume of such a washer is the area of the face times the thickness. The thickness, as usual, is Δx, while the area of the face is the area of the outer circle minus the area of

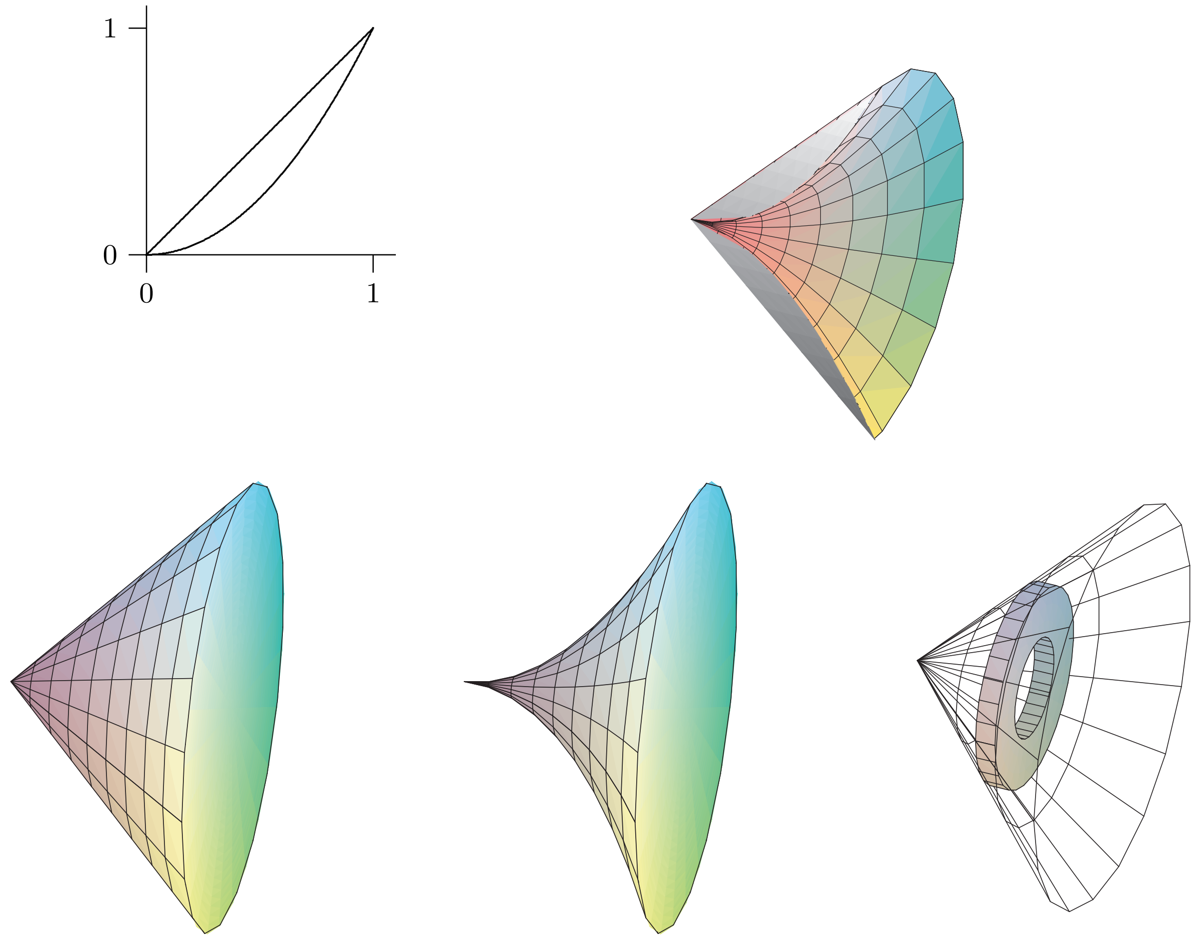


Figure 9.3.5 Solid with a hole, showing the outer cone and the shape to be removed to form the hole. (AP)

the inner circle, say $\pi R^2 - \pi r^2$. In the present example, at a particular x_i, the radius R is x_i and r is x_i^2. Hence, the whole volume is

$$\int_0^1 \pi x^2 - \pi x^4 \, dx = \pi \left(\frac{x^3}{3} - \frac{x^5}{5} \right) \bigg|_0^1 = \pi \left(\frac{1}{3} - \frac{1}{5} \right) = \frac{2\pi}{15}.$$

Of course, what we have done here is exactly the same calculation as before, except we have in effect recomputed the volume of the outer cone. □

Suppose the region between $f(x) = x + 1$ and $g(x) = (x - 1)^2$ is rotated around the y-axis; see figure 9.3.6. It is possible, but inconvenient, to compute the volume of the resulting solid by the method we have used so far. The problem is that there are two "kinds" of typical rectangles: those that go from the line to the parabola and those that touch the parabola on both ends. To compute the volume using this approach, we need to break the problem into two parts and compute two integrals:

$$\pi \int_0^1 (1+\sqrt{y})^2 - (1-\sqrt{y})^2\,dy + \pi \int_1^4 (1+\sqrt{y})^2 - (y-1)^2\,dy = \frac{8}{3}\pi + \frac{65}{6}\pi = \frac{27}{2}\pi.$$

If instead we consider a typical vertical rectangle, but still rotate around the y-axis, we get a thin "shell" instead of a thin "washer". If we add up the volume of such thin shells we will get an approximation to the true volume. What is the volume of such a shell? Consider the shell at x_i. Imagine that we cut the shell vertically in one place and "unroll" it into a thin, flat sheet. This sheet will be almost a rectangular prism that is Δx thick, $f(x_i) - g(x_i)$ tall, and $2\pi x_i$ wide (namely, the circumference of the shell before it was unrolled). The volume will then be approximately the volume of a rectangular prism with these dimensions: $2\pi x_i(f(x_i) - g(x_i))\Delta x$. If we add these up and take the limit as usual, we get the integral

$$\int_0^3 2\pi x(f(x) - g(x))\,dx = \int_0^3 2\pi x(x + 1 - (x-1)^2)\,dx = \frac{27}{2}\pi.$$

Not only does this accomplish the task with only one integral, the integral is somewhat easier than those in the previous calculation. Things are not always so neat, but it is often the case that one of the two methods will be simpler than the other, so it is worth considering both before starting to do calculations.

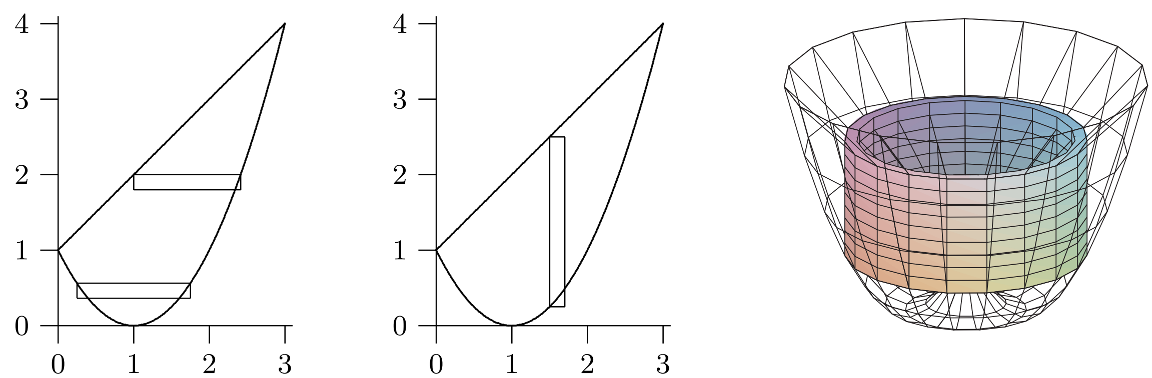


Figure 9.3.6 Computing volumes with "shells". (AP)

EXAMPLE 9.3.5 Suppose the area under $y = -x^2 + 1$ between $x = 0$ and $x = 1$ is rotated around the x-axis. Find the volume by both methods.

Disk method: $\displaystyle\int_0^1 \pi(1-x^2)^2\,dx = \frac{8}{15}\pi.$

Shell method: $\displaystyle\int_0^1 2\pi y\sqrt{1-y}\,dy = \frac{8}{15}\pi.$ □

Exercises 9.3.

1. Verify that $\pi\int_0^1 (1+\sqrt{y})^2-(1-\sqrt{y})^2\,dy+\pi\int_1^4 (1+\sqrt{y})^2-(y-1)^2 = \frac{8}{3}\pi+\frac{65}{6}\pi=\frac{27}{2}\pi$.

2. Verify that $\int_0^3 2\pi x(x+1-(x-1)^2)\,dx=\frac{27}{2}\pi$.

3. Verify that $\int_0^1 \pi(1-x^2)^2\,dx=\frac{8}{15}\pi$.

4. Verify that $\int_0^1 2\pi y\sqrt{1-y}\,dy=\frac{8}{15}\pi$.

5. Use integration to find the volume of the solid obtained by revolving the region bounded by $x+y=2$ and the x and y axes around the x-axis. ⇒

6. Find the volume of the solid obtained by revolving the region bounded by $y=x-x^2$ and the x-axis around the x-axis. ⇒

7. Find the volume of the solid obtained by revolving the region bounded by $y=\sqrt{\sin x}$ between $x=0$ and $x=\pi/2$, the y-axis, and the line $y=1$ around the x-axis. ⇒

8. Let S be the region of the xy-plane bounded above by the curve $x^3y=64$, below by the line $y=1$, on the left by the line $x=2$, and on the right by the line $x=4$. Find the volume of the solid obtained by rotating S around (a) the x-axis, (b) the line $y=1$, (c) the y-axis, (d) the line $x=2$. ⇒

9. The equation $x^2/9+y^2/4=1$ describes an ellipse. Find the volume of the solid obtained by rotating the ellipse around the x-axis and also around the y-axis. These solids are called **ellipsoids**; one is vaguely rugby-ball shaped, one is sort of flying-saucer shaped, or perhaps squished-beach-ball-shaped. ⇒

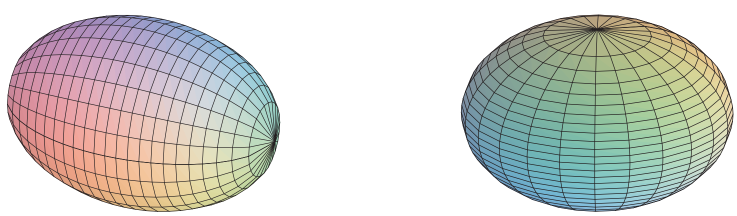


Figure 9.3.7 Ellipsoids.

10. Use integration to compute the volume of a sphere of radius r. You should of course get the well-known formula $4\pi r^3/3$.

11. A hemispheric bowl of radius r contains water to a depth h. Find the volume of water in the bowl. ⇒

12. The base of a tetrahedron (a triangular pyramid) of height h is an equilateral triangle of side s. Its cross-sections perpendicular to an altitude are equilateral triangles. Express its volume V as an integral, and find a formula for V in terms of h and s. Verify that your answer is (1/3)(area of base)(height).

13. The base of a solid is the region between $f(x)=\cos x$ and $g(x)=-\cos x$, $-\pi/2\le x\le\pi/2$, and its cross-sections perpendicular to the x-axis are squares. Find the volume of the solid. ⇒

9.4 Average value of a function

The average of some finite set of values is a familiar concept. If, for example, the class scores on a quiz are 10, 9, 10, 8, 7, 5, 7, 6, 3, 2, 7, 8, then the average score is the sum of these numbers divided by the size of the class:

$$\text{average score} = \frac{10+9+10+8+7+5+7+6+3+2+7+8}{12} = \frac{82}{12} \approx 6.83.$$

Suppose that between $t=0$ and $t=1$ the speed of an object is $\sin(\pi t)$. What is the average speed of the object over that time? The question sounds as if it must make sense, yet we can't merely add up some number of speeds and divide, since the speed is changing continuously over the time interval.

To make sense of "average" in this context, we fall back on the idea of approximation. Consider the speed of the object at tenth of a second intervals: $\sin 0$, $\sin(0.1\pi)$, $\sin(0.2\pi)$, $\sin(0.3\pi)$,…, $\sin(0.9\pi)$. The average speed "should" be fairly close to the average of these ten speeds:

$$\frac{1}{10}\sum_{i=0}^{9}\sin(\pi i/10) \approx \frac{1}{10}6.3 = 0.63.$$

Of course, if we compute more speeds at more times, the average of these speeds should be closer to the "real" average. If we take the average of n speeds at evenly spaced times, we get:

$$\frac{1}{n}\sum_{i=0}^{n-1}\sin(\pi i/n).$$

Here the individual times are $t_i = i/n$, so rewriting slightly we have

$$\frac{1}{n}\sum_{i=0}^{n-1}\sin(\pi t_i).$$

This is almost the sort of sum that we know turns into an integral; what's apparently missing is Δt—but in fact, $\Delta t = 1/n$, the length of each subinterval. So rewriting again:

$$\sum_{i=0}^{n-1}\sin(\pi t_i)\frac{1}{n} = \sum_{i=0}^{n-1}\sin(\pi t_i)\Delta t.$$

Now this has exactly the right form, so that in the limit we get

$$\text{average speed} = \int_0^1 \sin(\pi t)\,dt = -\frac{\cos(\pi t)}{\pi}\bigg|_0^1 = -\frac{\cos(\pi)}{\pi} + \frac{\cos(0)}{\pi} = \frac{2}{\pi} \approx 0.6366 \approx 0.64.$$

It's not entirely obvious from this one simple example how to compute such an average in general. Let's look at a somewhat more complicated case. Suppose that the velocity of an object is $16t^2 + 5$ feet per second. What is the average velocity between $t = 1$ and $t = 3$? Again we set up an approximation to the average:

$$\frac{1}{n}\sum_{i=0}^{n-1} 16t_i^2 + 5,$$

where the values t_i are evenly spaced times between 1 and 3. Once again we are "missing" Δt, and this time $1/n$ is not the correct value. What is Δt in general? It is the length of a subinterval; in this case we take the interval $[1, 3]$ and divide it into n subintervals, so each has length $(3 - 1)/n = 2/n = \Delta t$. Now with the usual "multiply and divide by the same thing" trick we can rewrite the sum:

$$\frac{1}{n}\sum_{i=0}^{n-1} 16t_i^2 + 5 = \frac{1}{3-1}\sum_{i=0}^{n-1}(16t_i^2 + 5)\frac{3-1}{n} = \frac{1}{2}\sum_{i=0}^{n-1}(16t_i^2 + 5)\frac{2}{n} = \frac{1}{2}\sum_{i=0}^{n-1}(16t_i^2 + 5)\Delta t.$$

In the limit this becomes

$$\frac{1}{2}\int_1^3 16t^2 + 5\,dt = \frac{1}{2}\frac{446}{3} = \frac{223}{3}.$$

Does this seem reasonable? Let's picture it: in figure 9.4.1 is the velocity function together with the horizontal line $y = 223/3 \approx 74.3$. Certainly the height of the horizontal line looks at least plausible for the average height of the curve.

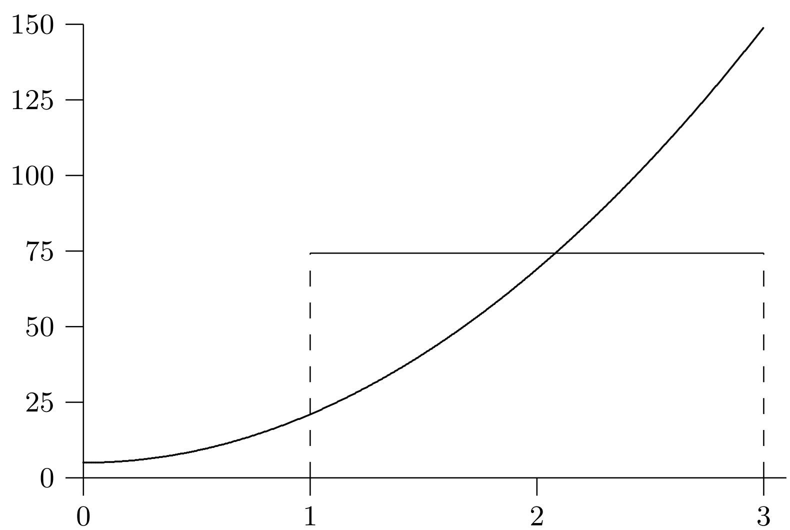


Figure 9.4.1 Average velocity.

Here's another way to interpret "average" that may make our computation appear even more reasonable. The object of our example goes a certain distance between $t = 1$

and $t = 3$. If instead the object were to travel at the average speed over the same time, it should go the same distance. At an average speed of 223/3 feet per second for two seconds the object would go 446/3 feet. How far does it actually go? We know how to compute this:

$$\int_1^3 v(t)\,dt = \int_1^3 16t^2 + 5\,dt = \frac{446}{3}.$$

So now we see that another interpretation of the calculation

$$\frac{1}{2}\int_1^3 16t^2 + 5\,dt = \frac{1}{2}\frac{446}{3} = \frac{223}{3}$$

is: total distance traveled divided by the time in transit, namely, the usual interpretation of average speed.

In the case of speed, or more properly velocity, we can always interpret "average" as total (net) distance divided by time. But in the case of a different sort of quantity this interpretation does not obviously apply, while the approximation approach always does. We might interpret the same problem geometrically: what is the average height of $16x^2+5$ on the interval $[1,3]$? We approximate this in exactly the same way, by adding up many sample heights and dividing by the number of samples. In the limit we get the same result:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} 16x_i^2 + 5 = \frac{1}{2}\int_1^3 16x^2 + 5\,dx = \frac{1}{2}\frac{446}{3} = \frac{223}{3}.$$

We can interpret this result in a slightly different way. The area under $y = 16x^2 + 5$ above $[1,3]$ is

$$\int_1^3 16t^2 + 5\,dt = \frac{446}{3}.$$

The area under $y = 223/3$ over the same interval $[1,3]$ is simply the area of a rectangle that is 2 by 223/3 with area 446/3. So the average height of a function is the height of the horizontal line that produces the same area over the given interval.

Exercises 9.4.

1. Find the average height of $\cos x$ over the intervals $[0,\pi/2]$, $[-\pi/2,\pi/2]$, and $[0,2\pi]$. ⇒

2. Find the average height of x^2 over the interval $[-2,2]$. ⇒

3. Find the average height of $1/x^2$ over the interval $[1,A]$. ⇒

4. Find the average height of $\sqrt{1-x^2}$ over the interval $[-1,1]$. ⇒

5. An object moves with velocity $v(t) = -t^2+1$ feet per second between $t = 0$ and $t = 2$. Find the average velocity and the average speed of the object between $t = 0$ and $t = 2$. ⇒

6. The observation deck on the 102nd floor of the Empire State Building is 1,224 feet above the ground. If a steel ball is dropped from the observation deck its velocity at time t is approximately $v(t) = -32t$ feet per second. Find the average speed between the time it is dropped and the time it hits the ground, and find its speed when it hits the ground. ⇒

9.5 Work

A fundamental concept in classical physics is **work**: If an object is moved in a straight line against a force F for a distance s the work done is $W = Fs$.

EXAMPLE 9.5.1 How much work is done in lifting a 10 pound weight vertically a distance of 5 feet? The force due to gravity on a 10 pound weight is 10 pounds at the surface of the earth, and it does not change appreciably over 5 feet. The work done is $W = 10 \cdot 5 = 50$ foot-pounds. □

In reality few situations are so simple. The force might not be constant over the range of motion, as in the next example.

EXAMPLE 9.5.2 How much work is done in lifting a 10 pound weight from the surface of the earth to an orbit 100 miles above the surface? Over 100 miles the force due to gravity does change significantly, so we need to take this into account. The force exerted on a 10 pound weight at a distance r from the center of the earth is $F = k/r^2$ and by definition it is 10 when r is the radius of the earth (we assume the earth is a sphere). How can we approximate the work done? We divide the path from the surface to orbit into n small subpaths. On each subpath the force due to gravity is roughly constant, with value k/r_i^2 at distance r_i. The work to raise the object from r_i to r_{i+1} is thus approximately $k/r_i^2 \Delta r$ and the total work is approximately

$$\sum_{i=0}^{n-1} \frac{k}{r_i^2} \Delta r,$$

or in the limit

$$W = \int_{r_0}^{r_1} \frac{k}{r^2}\, dr,$$

where r_0 is the radius of the earth and r_1 is r_0 plus 100 miles. The work is

$$W = \int_{r_0}^{r_1} \frac{k}{r^2}\, dr = -\left.\frac{k}{r}\right|_{r_0}^{r_1} = -\frac{k}{r_1} + \frac{k}{r_0}.$$

Using $r_0 = 20925525$ feet we have $r_1 = 21453525$. The force on the 10 pound weight at the surface of the earth is 10 pounds, so $10 = k/20925525^2$, giving $k = 4378775965256250$.

Then

$$-\frac{k}{r_1}+\frac{k}{r_0}=\frac{491052320000}{95349}\approx 5150052 \quad \text{foot-pounds.}$$

Note that if we assume the force due to gravity is 10 pounds over the whole distance we would calculate the work as $10(r_1-r_0)=10\cdot 100\cdot 5280=5280000$, somewhat higher since we don't account for the weakening of the gravitational force. □

EXAMPLE 9.5.3 How much work is done in lifting a 10 kilogram object from the surface of the earth to a distance D from the center of the earth? This is the same problem as before in different units, and we are not specifying a value for D. As before

$$W=\int_{r_0}^{D}\frac{k}{r^2}\,dr=-\left.\frac{k}{r}\right|_{r_0}^{D}=-\frac{k}{D}+\frac{k}{r_0}.$$

While "weight in pounds" is a measure of force, "weight in kilograms" is a measure of mass. To convert to force we need to use Newton's law $F=ma$. At the surface of the earth the acceleration due to gravity is approximately 9.8 meters per second squared, so the force is $F=10\cdot 9.8=98$. The units here are "kilogram-meters per second squared" or "kg m/s^2", also known as a Newton (N), so $F=98$ N. The radius of the earth is approximately 6378.1 kilometers or 6378100 meters. Now the problem proceeds as before. From $F=k/r^2$ we compute k: $98=k/6378100^2$, $k=3.986655642\cdot 10^{15}$. Then the work is:

$$W=-\frac{k}{D}+6.250538000\cdot 10^8 \quad \text{Newton-meters.}$$

As D increases W of course gets larger, since the quantity being subtracted, $-k/D$, gets smaller. But note that the work W will never exceed $6.250538000\cdot 10^8$, and in fact will approach this value as D gets larger. In short, with a finite amount of work, namely $6.250538000\cdot 10^8$ N-m, we can lift the 10 kilogram object as far as we wish from earth. □

Next is an example in which the force is constant, but there are many objects moving different distances.

EXAMPLE 9.5.4 Suppose that a water tank is shaped like a right circular cone with the tip at the bottom, and has height 10 meters and radius 2 meters at the top. If the tank is full, how much work is required to pump all the water out over the top? Here we have a large number of atoms of water that must be lifted different distances to get to the top of the tank. Fortunately, we don't really have to deal with individual atoms—we can consider all the atoms at a given depth together.

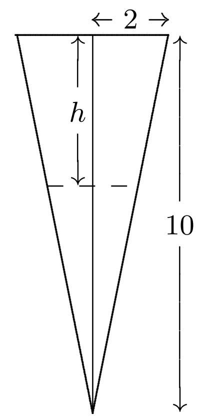


Figure 9.5.1 Cross-section of a conical water tank.

To approximate the work, we can divide the water in the tank into horizontal sections, approximate the volume of water in a section by a thin disk, and compute the amount of work required to lift each disk to the top of the tank. As usual, we take the limit as the sections get thinner and thinner to get the total work.

At depth h the circular cross-section through the tank has radius $r = (10-h)/5$, by similar triangles, and area $\pi(10-h)^2/25$. A section of the tank at depth h thus has volume approximately $\pi(10-h)^2/25\Delta h$ and so contains $\sigma\pi(10-h)^2/25\Delta h$ kilograms of water, where σ is the density of water in kilograms per cubic meter; $\sigma \approx 1000$. The force due to gravity on this much water is $9.8\sigma\pi(10-h)^2/25\Delta h$, and finally, this section of water must be lifted a distance h, which requires $h9.8\sigma\pi(10-h)^2/25\Delta h$ Newton-meters of work. The total work is therefore

$$W = \frac{9.8\sigma\pi}{25}\int_0^{10} h(10-h)^2\,dh = \frac{980000}{3}\pi \approx 1026254 \quad \text{Newton-meters.}$$

□

A spring has a "natural length," its length if nothing is stretching or compressing it. If the spring is either stretched or compressed the spring provides an opposing force; according to **Hooke's Law** the magnitude of this force is proportional to the distance the spring has been stretched or compressed: $F = kx$. The constant of proportionality, k, of course depends on the spring. Note that x here represents the *change* in length from the natural length.

EXAMPLE 9.5.5 Suppose $k = 5$ for a given spring that has a natural length of 0.1 meters. Suppose a force is applied that compresses the spring to length 0.08. What is the magnitude of the force? Assuming that the constant k has appropriate dimensions (namely, kg/s^2), the force is $5(0.1-0.08) = 5(0.02) = 0.1$ Newtons. □

EXAMPLE 9.5.6 How much work is done in compressing the spring in the previous example from its natural length to 0.08 meters? From 0.08 meters to 0.05 meters? How much work is done to stretch the spring from 0.1 meters to 0.15 meters? We can approximate the work by dividing the distance that the spring is compressed (or stretched) into small subintervals. Then the force exerted by the spring is approximately constant over the subinterval, so the work required to compress the spring from x_i to x_{i+1} is approximately $5(x_i - 0.1)\Delta x$. The total work is approximately

$$\sum_{i=0}^{n-1} 5(x_i - 0.1)\Delta x$$

and in the limit

$$W = \int_{0.1}^{0.08} 5(x - 0.1)\,dx = \left.\frac{5(x-0.1)^2}{2}\right|_{0.1}^{0.08} = \frac{5(0.08-0.1)^2}{2} - \frac{5(0.1-0.1)^2}{2} = \frac{1}{1000}\text{ N-m.}$$

The other values we seek simply use different limits. To compress the spring from 0.08 meters to 0.05 meters takes

$$W = \int_{0.08}^{0.05} 5(x - 0.1)\,dx = \left.\frac{5x^2}{2}\right|_{0.08}^{0.05} = \frac{5(0.05-0.1)^2}{2} - \frac{5(0.08-0.1)^2}{2} = \frac{21}{4000}\quad\text{N-m}$$

and to stretch the spring from 0.1 meters to 0.15 meters requires

$$W = \int_{0.1}^{0.15} 5(x - 0.1)\,dx = \left.\frac{5x^2}{2}\right|_{0.1}^{0.15} = \frac{5(0.15-0.1)^2}{2} - \frac{5(0.1-0.1)^2}{2} = \frac{1}{160}\quad\text{N-m.}$$

□

Exercises 9.5.

1. How much work is done in lifting a 100 kilogram weight from the surface of the earth to an orbit 35,786 kilometers above the surface of the earth? ⇒
2. How much work is done in lifting a 100 kilogram weight from an orbit 1000 kilometers above the surface of the earth to an orbit 35,786 kilometers above the surface of the earth? ⇒
3. A water tank has the shape of an upright cylinder with radius $r = 1$ meter and height 10 meters. If the depth of the water is 5 meters, how much work is required to pump all the water out the top of the tank? ⇒
4. Suppose the tank of the previous problem is lying on its side, so that the circular ends are vertical, and that it has the same amount of water as before. How much work is required to pump the water out the top of the tank (which is now 2 meters above the bottom of the tank)? ⇒

5. A water tank has the shape of the bottom half of a sphere with radius $r = 1$ meter. If the tank is full, how much work is required to pump all the water out the top of the tank? ⇒

6. A spring has constant $k = 10$ kg/s^2. How much work is done in compressing it 1/10 meter from its natural length? ⇒

7. A force of 2 Newtons will compress a spring from 1 meter (its natural length) to 0.8 meters. How much work is required to stretch the spring from 1.1 meters to 1.5 meters? ⇒

8. A 20 meter long steel cable has density 2 kilograms per meter, and is hanging straight down. How much work is required to lift the entire cable to the height of its top end? ⇒

9. The cable in the previous problem has a 100 kilogram bucket of concrete attached to its lower end. How much work is required to lift the entire cable and bucket to the height of its top end? ⇒

10. Consider again the cable and bucket of the previous problem. How much work is required to lift the bucket 10 meters by raising the cable 10 meters? (The top half of the cable ends up at the height of the top end of the cable, while the bottom half of the cable is lifted 10 meters.) ⇒

9.6 Center of Mass

Suppose a beam is 10 meters long, and that there are three weights on the beam: a 10 kilogram weight 3 meters from the left end, a 5 kilogram weight 6 meters from the left end, and a 4 kilogram weight 8 meters from the left end. Where should a fulcrum be placed so that the beam balances? Let's assign a scale to the beam, from 0 at the left end to 10 at the right, so that we can denote locations on the beam simply as x coordinates; the weights are at $x = 3$, $x = 6$, and $x = 8$, as in figure 9.6.1.

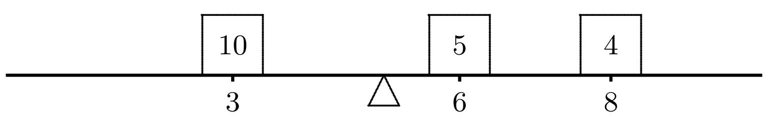


Figure 9.6.1 A beam with three masses.

Suppose to begin with that the fulcrum is placed at $x = 5$. What will happen? Each weight applies a force to the beam that tends to rotate it around the fulcrum; this effect is measured by a quantity called **torque**, proportional to the mass times the distance from the fulcrum. Of course, weights on different sides of the fulcrum rotate the beam in opposite directions. We can distinguish this by using a signed distance in the formula for torque. So with the fulcrum at 5, the torques induced by the three weights will be proportional to $(3 - 5)10 = -20$, $(6 - 5)5 = 5$, and $(8 - 5)4 = 12$. For the beam to balance, the sum of the torques must be zero; since the sum is $-20 + 5 + 12 = -3$, the beam rotates counter-clockwise, and to get the beam to balance we need to move the fulcrum to the left. To calculate exactly where the fulcrum should be, we let $\bar{x}$ denote the location of the fulcrum when the beam is in balance. The total torque on the beam is then

$(3-\bar{x})10+(6-\bar{x})5+(8-\bar{x})4 = 92-19\bar{x}$. Since the beam balances at $\bar{x}$ it must be that $92-19\bar{x}=0$ or $\bar{x}=92/19 \approx 4.84$, that is, the fulcrum should be placed at $x=92/19$ to balance the beam.

Now suppose that we have a beam with varying density—some portions of the beam contain more mass than other portions of the same size. We want to figure out where to put the fulcrum so that the beam balances.

m_0	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9

Figure 9.6.2 A solid beam.

EXAMPLE 9.6.1 Suppose the beam is 10 meters long and that the density is $1+x$ kilograms per meter at location x on the beam. To approximate the solution, we can think of the beam as a sequence of weights "on" a beam. For example, we can think of the portion of the beam between $x=0$ and $x=1$ as a weight sitting at $x=0$, the portion between $x=1$ and $x=2$ as a weight sitting at $x=1$, and so on, as indicated in figure 9.6.2. We then approximate the mass of the weights by assuming that each portion of the beam has constant density. So the mass of the first weight is approximately $m_0=(1+0)1=1$ kilograms, namely, $(1+0)$ kilograms per meter times 1 meter. The second weight is $m_1=(1+1)1=2$ kilograms, and so on to the tenth weight with $m_9=(1+9)1=10$ kilograms. So in this case the total torque is

$$(0-\bar{x})m_0+(1-\bar{x})m_1+\cdots+(9-\bar{x})m_9 = (0-\bar{x})1+(1-\bar{x})2+\cdots+(9-\bar{x})10.$$

If we set this to zero and solve for $\bar{x}$ we get $\bar{x}=6$. In general, if we divide the beam into n portions, the mass of weight number i will be $m_i=(1+x_i)(x_{i+1}-x_i)=(1+x_i)\Delta x$ and the torque induced by weight number i will be $(x_i-\bar{x})m_i=(x_i-\bar{x})(1+x_i)\Delta x$. The total torque is then

$$\begin{aligned}(x_0-\bar{x})(1+x_0)\Delta x + (x_1-\bar{x})(1+x_1)\Delta x + \cdots + (x_{n-1}-\bar{x})(1+x_{n-1})\Delta x \\ = \sum_{i=0}^{n-1} x_i(1+x_i)\Delta x - \sum_{i=0}^{n-1} \bar{x}(1+x_i)\Delta x \\ = \sum_{i=0}^{n-1} x_i(1+x_i)\Delta x - \bar{x}\sum_{i=0}^{n-1}(1+x_i)\Delta x.\end{aligned}$$

If we set this equal to zero and solve for $\bar{x}$ we get an approximation to the balance point of the beam:

$$\begin{aligned}
0 &= \sum_{i=0}^{n-1} x_i(1+x_i)\Delta x - \bar{x}\sum_{i=0}^{n-1}(1+x_i)\Delta x \\
\bar{x}\sum_{i=0}^{n-1}(1+x_i)\Delta x &= \sum_{i=0}^{n-1} x_i(1+x_i)\Delta x \\
\bar{x} &= \frac{\displaystyle\sum_{i=0}^{n-1} x_i(1+x_i)\Delta x}{\displaystyle\sum_{i=0}^{n-1}(1+x_i)\Delta x}.
\end{aligned}$$

The denominator of this fraction has a very familiar interpretation. Consider one term of the sum in the denominator: $(1+x_i)\Delta x$. This is the density near x_i times a short length, Δx, which in other words is approximately the mass of the beam between x_i and x_{i+1}. When we add these up we get approximately the mass of the beam.

Now each of the sums in the fraction has the right form to turn into an integral, which in turn gives us the exact value of $\bar{x}$:

$$\bar{x} = \frac{\displaystyle\int_0^{10} x(1+x)\,dx}{\displaystyle\int_0^{10}(1+x)\,dx}.$$

The numerator of this fraction is called the **moment** of the system around zero:

$$\int_0^{10} x(1+x)\,dx = \int_0^{10} x + x^2\,dx = \frac{1150}{3},$$

and the denominator is the mass of the beam:

$$\int_0^{10}(1+x)\,dx = 60,$$

and the balance point, officially called the **center of mass**, is

$$\bar{x} = \frac{1150}{3}\frac{1}{60} = \frac{115}{18} \approx 6.39.$$

□

It should be apparent that there was nothing special about the density function $\sigma(x) = 1 + x$ or the length of the beam, or even that the left end of the beam is at the origin. In general, if the density of the beam is $\sigma(x)$ and the beam covers the interval $[a, b]$, the moment of the beam around zero is

$$M_0 = \int_a^b x\sigma(x)\,dx$$

and the total mass of the beam is

$$M = \int_a^b \sigma(x)\,dx$$

and the center of mass is at

$$\bar{x} = \frac{M_0}{M}.$$

EXAMPLE 9.6.2 Suppose a beam lies on the x-axis between 20 and 30, and has density function $\sigma(x) = x - 19$. Find the center of mass. This is the same as the previous example except that the beam has been moved. Note that the density at the left end is $20 - 19 = 1$ and at the right end is $30 - 19 = 11$, as before. Hence the center of mass must be at approximately $20 + 6.39 = 26.39$. Let's see how the calculation works out.

$$\begin{aligned} M_0 &= \int_{20}^{30} x(x-19)\,dx = \int_{20}^{30} x^2 - 19x\,dx = \left.\frac{x^3}{3} - \frac{19x^2}{2}\right|_{20}^{30} = \frac{4750}{3} \\ M &= \int_{20}^{30} x - 19\,dx = \left.\frac{x^2}{2} - 19x\right|_{20}^{30} = 60 \\ \frac{M_0}{M} &= \frac{4750}{3}\frac{1}{60} = \frac{475}{18} \approx 26.39. \end{aligned}$$

□

EXAMPLE 9.6.3 Suppose a flat plate of uniform density has the shape contained by $y = x^2$, $y = 1$, and $x = 0$, in the first quadrant. Find the center of mass. (Since the density is constant, the center of mass depends only on the shape of the plate, not the density, or in other words, this is a purely geometric quantity. In such a case the center of mass is called the **centroid**.)

This is a two dimensional problem, but it can be solved as if it were two one dimensional problems: we need to find the x and y coordinates of the center of mass, $\bar{x}$ and $\bar{y}$, and fortunately we can do these independently. Imagine looking at the plate edge on, from below the x-axis. The plate will appear to be a beam, and the mass of a short section

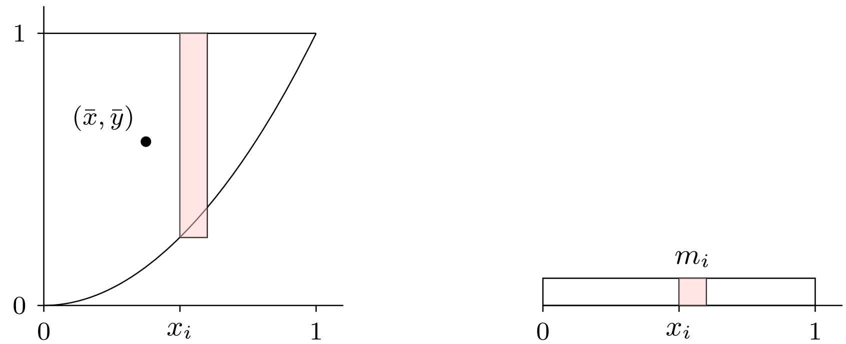


Figure 9.6.3 Center of mass for a two dimensional plate.

of the "beam", say between x_i and x_{i+1}, is the mass of a strip of the plate between x_i and x_{i+1}. See figure 9.6.3 showing the plate from above and as it appears edge on. Since the plate has uniform density we may as well assume that $\sigma = 1$. Then the mass of the plate between x_i and x_{i+1} is approximately $m_i = \sigma(1 - x_i^2)\Delta x = (1 - x_i^2)\Delta x$. Now we can compute the moment around the y-axis:

$$M_y = \int_0^1 x(1 - x^2)\,dx = \frac{1}{4}$$

and the total mass

$$M = \int_0^1 (1 - x^2)\,dx = \frac{2}{3}$$

and finally

$$\bar{x} = \frac{1}{4}\frac{3}{2} = \frac{3}{8}.$$

Next we do the same thing to find $\bar{y}$. The mass of the plate between y_i and y_{i+1} is approximately $n_i = \sqrt{y}\Delta y$, so

$$M_x = \int_0^1 y\sqrt{y}\,dy = \frac{2}{5}$$

and

$$\bar{y} = \frac{2}{5}\frac{3}{2} = \frac{3}{5},$$

since the total mass M is the same. The center of mass is shown in figure 9.6.3. □

EXAMPLE 9.6.4 Find the center of mass of a thin, uniform plate whose shape is the region between $y = \cos x$ and the x-axis between $x = -\pi/2$ and $x = \pi/2$. It is clear that $\bar{x} = 0$, but for practice let's compute it anyway. We will need the total mass, so we compute it first:

$$M = \int_{-\pi/2}^{\pi/2} \cos x\,dx = \sin x\Big|_{-\pi/2}^{\pi/2} = 2.$$

The moment around the y-axis is

$$M_y = \int_{-\pi/2}^{\pi/2} x\cos x\,dx = \cos x + x\sin x\Big|_{-\pi/2}^{\pi/2} = 0$$

and the moment around the x-axis is

$$M_x = \int_0^1 y\cdot 2\arccos y\,dy = y^2\arccos y - \frac{y\sqrt{1-y^2}}{2} + \frac{\arcsin y}{2}\Bigg|_0^1 = \frac{\pi}{4}.$$

Thus

$$\bar{x} = \frac{0}{2}, \quad \bar{y} = \frac{\pi}{8} \approx 0.393.$$

□

Exercises 9.6.

1. A beam 10 meters long has density $\sigma(x) = x^2$ at distance x from the left end of the beam. Find the center of mass $\bar{x}$. ⇒

2. A beam 10 meters long has density $\sigma(x) = \sin(\pi x/10)$ at distance x from the left end of the beam. Find the center of mass $\bar{x}$. ⇒

3. A beam 4 meters long has density $\sigma(x) = x^3$ at distance x from the left end of the beam. Find the center of mass $\bar{x}$. ⇒

4. Verify that $\displaystyle\int 2x\arccos x\,dx = x^2\arccos x - \frac{x\sqrt{1-x^2}}{2} + \frac{\arcsin x}{2} + C.$

5. A thin plate lies in the region between $y = x^2$ and the x-axis between $x = 1$ and $x = 2$. Find the centroid. ⇒

6. A thin plate fills the upper half of the unit circle $x^2 + y^2 = 1$. Find the centroid. ⇒

7. A thin plate lies in the region contained by $y = x$ and $y = x^2$. Find the centroid. ⇒

8. A thin plate lies in the region contained by $y = 4 - x^2$ and the x-axis. Find the centroid. ⇒

9. A thin plate lies in the region contained by $y = x^{1/3}$ and the x-axis between $x = 0$ and $x = 1$. Find the centroid. ⇒

10. A thin plate lies in the region contained by $\sqrt{x} + \sqrt{y} = 1$ and the axes in the first quadrant. Find the centroid. ⇒

11. A thin plate lies in the region between the circle $x^2 + y^2 = 4$ and the circle $x^2 + y^2 = 1$, above the x-axis. Find the centroid. ⇒

12. A thin plate lies in the region between the circle $x^2 + y^2 = 4$ and the circle $x^2 + y^2 = 1$ in the first quadrant. Find the centroid. ⇒

13. A thin plate lies in the region between the circle $x^2 + y^2 = 25$ and the circle $x^2 + y^2 = 16$ above the x-axis. Find the centroid. ⇒

9.7 Kinetic energy; improper integrals

Recall example 9.5.3 in which we computed the work required to lift an object from the surface of the earth to some large distance D away. Since $F = k/x^2$ we computed

$$\int_{r_0}^{D} \frac{k}{x^2}\,dx = -\frac{k}{D} + \frac{k}{r_0}.$$

We noticed that as D increases, k/D decreases to zero so that the amount of work increases to k/r_0. More precisely,

$$\lim_{D\to\infty} \int_{r_0}^{D} \frac{k}{x^2}\,dx = \lim_{D\to\infty} -\frac{k}{D} + \frac{k}{r_0} = \frac{k}{r_0}.$$

We might reasonably describe this calculation as computing the amount of work required to lift the object "to infinity," and abbreviate the limit as

$$\lim_{D\to\infty} \int_{r_0}^{D} \frac{k}{x^2}\,dx = \int_{r_0}^{\infty} \frac{k}{x^2}\,dx.$$

Such an integral, with a limit of infinity, is called an **improper integral**. This is a bit unfortunate, since it's not really "improper" to do this, nor is it really "an integral"—it is an abbreviation for the limit of a particular sort of integral. Nevertheless, we're stuck with the term, and the operation itself is perfectly legitimate. It may at first seem odd that a finite amount of work is sufficient to lift an object to "infinity", but sometimes surprising things are nevertheless true, and this is such a case. If the value of an improper integral is a finite number, as in this example, we say that the integral **converges**, and if not we say that the integral **diverges**.

Here's another way, perhaps even more surprising, to interpret this calculation. We know that one interpretation of

$$\int_{1}^{D} \frac{1}{x^2}\,dx$$

is the area under $y = 1/x^2$ from $x = 1$ to $x = D$. Of course, as D increases this area increases. But since

$$\int_{1}^{D} \frac{1}{x^2}\,dx = -\frac{1}{D} + \frac{1}{1},$$

while the area increases, it never exceeds 1, that is

$$\int_{1}^{\infty} \frac{1}{x^2}\,dx = 1.$$

The area of the infinite region under $y = 1/x^2$ from $x = 1$ to infinity is finite.

Consider a slightly different sort of improper integral: $\int_{-\infty}^{\infty} xe^{-x^2}\,dx$. There are two ways we might try to compute this. First, we could break it up into two more familiar integrals:

$$\int_{-\infty}^{\infty} xe^{-x^2}\,dx = \int_{-\infty}^{0} xe^{-x^2}\,dx + \int_{0}^{\infty} xe^{-x^2}\,dx.$$

Now we do these as before:

$$\int_{-\infty}^{0} xe^{-x^2}\,dx = \lim_{D\to\infty} -\frac{e^{-x^2}}{2}\bigg|_{D}^{0} = -\frac{1}{2},$$

and

$$\int_{0}^{\infty} xe^{-x^2}\,dx = \lim_{D\to\infty} -\frac{e^{-x^2}}{2}\bigg|_{0}^{D} = \frac{1}{2},$$

so

$$\int_{-\infty}^{\infty} xe^{-x^2}\,dx = -\frac{1}{2} + \frac{1}{2} = 0.$$

Alternately, we might try

$$\int_{-\infty}^{\infty} xe^{-x^2}\,dx = \lim_{D\to\infty} \int_{-D}^{D} xe^{-x^2}\,dx = \lim_{D\to\infty} -\frac{e^{-x^2}}{2}\bigg|_{-D}^{D} = \lim_{D\to\infty} -\frac{e^{-D^2}}{2} + \frac{e^{-D^2}}{2} = 0.$$

So we get the same answer either way. This does not always happen; sometimes the second approach gives a finite number, while the first approach does not; the exercises provide examples. In general, we interpret the integral $\int_{-\infty}^{\infty} f(x)\,dx$ according to the first method: both integrals $\int_{-\infty}^{a} f(x)\,dx$ and $\int_{a}^{\infty} f(x)\,dx$ must converge for the original integral to converge. The second approach does turn out to be useful; when $\lim_{D\to\infty} \int_{-D}^{D} f(x)\,dx = L$, and L is finite, then L is called the **Cauchy Principal Value** of $\int_{-\infty}^{\infty} f(x)\,dx$.

Here's a more concrete application of these ideas. We know that in general

$$W = \int_{x_0}^{x_1} F\,dx$$

is the work done against the force F in moving from x_0 to x_1. In the case that F is the force of gravity exerted by the earth, it is customary to make $F < 0$ since the force is "downward." This makes the work W negative when it should be positive, so typically the work in this case is defined as

$$W = -\int_{x_0}^{x_1} F\,dx.$$

Also, by Newton's Law, $F = ma(t)$. This means that

$$W = -\int_{x_0}^{x_1} ma(t)\,dx.$$

Unfortunately this integral is a bit problematic: $a(t)$ is in terms of t, while the limits and the "dx" are in terms of x. But x and t are certainly related here: $x = x(t)$ is the function that gives the position of the object at time t, so $v = v(t) = dx/dt = x'(t)$ is its velocity and $a(t) = v'(t) = x''(t)$. We can use $v = x'(t)$ as a substitution to convert the integral from "dx" to "dv" in the usual way, with a bit of cleverness along the way:

$$\begin{aligned}
dv &= x''(t)\,dt = a(t)\,dt = a(t)\frac{dt}{dx}\,dx \\
\frac{dx}{dt}\,dv &= a(t)\,dx \\
v\,dv &= a(t)\,dx.
\end{aligned}$$

Substituting in the integral:

$$W = -\int_{x_0}^{x_1} ma(t)\,dx = -\int_{v_0}^{v_1} mv\,dv = -\left.\frac{mv^2}{2}\right|_{v_0}^{v_1} = -\frac{mv_1^2}{2} + \frac{mv_0^2}{2}.$$

You may recall seeing the expression $mv^2/2$ in a physics course—it is called the **kinetic energy** of the object. We have shown here that the work done in moving the object from one place to another is the same as the change in kinetic energy.

We know that the work required to move an object from the surface of the earth to infinity is

$$W = \int_{r_0}^{\infty} \frac{k}{r^2}\,dr = \frac{k}{r_0}.$$

At the surface of the earth the acceleration due to gravity is approximately 9.8 meters per second squared, so the force on an object of mass m is $F = 9.8m$. The radius of the earth is approximately 6378.1 kilometers or 6378100 meters. Since the force due to gravity obeys an inverse square law, $F = k/r^2$ and $9.8m = k/6378100^2$, $k = 398665564178000m$ and $W = 62505380m$.

Now suppose that the initial velocity of the object, v_0, is just enough to get it to infinity, that is, just enough so that the object never slows to a stop, but so that its speed decreases to zero, i.e., so that $v_1 = 0$. Then

$$62505380m = W = -\frac{mv_1^2}{2} + \frac{mv_0^2}{2} = \frac{mv_0^2}{2}$$

so

$$v_0 = \sqrt{125010760} \approx 11181 \quad \text{meters per second,}$$

or about 40251 kilometers per hour. This speed is called the **escape velocity**. Notice that the mass of the object, m, canceled out at the last step; the escape velocity is the same for all objects. Of course, it takes considerably more energy to get a large object up to 40251 kph than a small one, so it is certainly more difficult to get a large object into deep space than a small one. Also, note that while we have computed the escape velocity for the earth, this speed would not in fact get an object "to infinity" because of the large mass in our neighborhood called the sun. Escape velocity for the sun *starting at the distance of the earth from the sun* is nearly 4 times the escape velocity we have calculated.

Exercises 9.7.

1. Is the area under $y = 1/x$ from 1 to infinity finite or infinite? If finite, compute the area. ⇒

2. Is the area under $y = 1/x^3$ from 1 to infinity finite or infinite? If finite, compute the area. ⇒

3. Does $\int_0^\infty x^2 + 2x - 1\,dx$ converge or diverge? If it converges, find the value. ⇒

4. Does $\int_1^\infty 1/\sqrt{x}\,dx$ converge or diverge? If it converges, find the value. ⇒

5. Does $\int_0^\infty e^{-x}\,dx$ converge or diverge? If it converges, find the value. ⇒

6. $\int_0^{1/2} (2x-1)^{-3}\,dx$ is an improper integral of a slightly different sort. Express it as a limit and determine whether it converges or diverges; if it converges, find the value. ⇒

7. Does $\int_0^1 1/\sqrt{x}\,dx$ converge or diverge? If it converges, find the value. ⇒

8. Does $\int_0^{\pi/2} \sec^2 x\,dx$ converge or diverge? If it converges, find the value. ⇒

9. Does $\int_{-\infty}^{\infty} \frac{x^2}{4+x^6}\,dx$ converge or diverge? If it converges, find the value. ⇒

10. Does $\int_{-\infty}^{\infty} x\,dx$ converge or diverge? If it converges, find the value. Also find the Cauchy Principal Value, if it exists. ⇒

11. Does $\int_{-\infty}^{\infty} \sin x\,dx$ converge or diverge? If it converges, find the value. Also find the Cauchy Principal Value, if it exists. ⇒

12. Does $\int_{-\infty}^{\infty} \cos x\, dx$ converge or diverge? If it converges, find the value. Also find the Cauchy Principal Value, if it exists. ⇒

13. Suppose the curve $y = 1/x$ is rotated around the x-axis generating a sort of funnel or horn shape, called **Gabriel's horn** or **Toricelli's trumpet**. Is the volume of this funnel from $x = 1$ to infinity finite or infinite? If finite, compute the volume. ⇒

14. An officially sanctioned baseball must be between 142 and 149 grams. How much work, in Newton-meters, does it take to throw a ball at 80 miles per hour? At 90 mph? At 100.9 mph? (According to the Guinness Book of World Records, at `http://www.baseball-almanac.com/recbooks/rb_guin.shtml`, "The greatest reliably recorded speed at which a baseball has been pitched is 100.9 mph by Lynn Nolan Ryan (California Angels) at Anaheim Stadium in California on August 20, 1974.") ⇒

9.8 Probability

You perhaps have at least a rudimentary understanding of **discrete probability**, which measures the likelihood of an "event" when there are a finite number of possibilities. For example, when an ordinary six-sided die is rolled, the probability of getting any particular number is $1/6$. In general, the probability of an event is the number of ways the event can happen divided by the number of ways that "anything" can happen.

For a slightly more complicated example, consider the case of two six-sided dice. The dice are physically distinct, which means that rolling a 2–5 is different than rolling a 5–2; each is an equally likely event out of a total of 36 ways the dice can land, so each has a probability of $1/36$.

Most interesting events are not so simple. More interesting is the probability of rolling a certain sum out of the possibilities 2 through 12. It is clearly not true that all sums are equally likely: the only way to roll a 2 is to roll 1–1, while there are many ways to roll a 7. Because the number of possibilities is quite small, and because a pattern quickly becomes evident, it is easy to see that the probabilities of the various sums are:

$$
\begin{aligned}
P(2) = P(12) &= 1/36\\
P(3) = P(11) &= 2/36\\
P(4) = P(10) &= 3/36\\
P(5) = P(9) &= 4/36\\
P(6) = P(8) &= 5/36\\
P(7) &= 6/36
\end{aligned}
$$

Here we use $P(n)$ to mean "the probability of rolling an n." Since we have correctly accounted for all possibilities, the sum of all these probabilities is $36/36 = 1$; the probability that the sum is one of 2 through 12 is 1, because there are no other possibilities.

The study of probability is concerned with more difficult questions as well; for example, suppose the two dice are rolled many times. On the average, what sum will come up? In the language of probability, this average is called the **expected value** of the sum. This is at first a little misleading, as it does not tell us what to "expect" when the two dice are rolled, but what we expect the long term average will be.

Suppose that two dice are rolled 36 million times. Based on the probabilities, we would expect about 1 million rolls to be 2, about 2 million to be 3, and so on, with a roll of 7 topping the list at about 6 million. The sum of all rolls would be 1 million times 2 plus 2 million times 3, and so on, and dividing by 36 million we would get the average:

$$\begin{aligned}\bar{x} &= (2\cdot 10^6 + 3(2\cdot 10^6) + \cdots + 7(6\cdot 10^6) + \cdots + 12\cdot 10^6)\frac{1}{36\cdot 10^6}\\ &= 2\frac{10^6}{36\cdot 10^6} + 3\frac{2\cdot 10^6}{36\cdot 10^6} + \cdots + 7\frac{6\cdot 10^6}{36\cdot 10^6} + \cdots + 12\frac{10^6}{36\cdot 10^6}\\ &= 2P(2) + 3P(3) + \cdots + 7P(7) + \cdots + 12P(12)\\ &= \sum_{i=2}^{12} iP(i) = 7.\end{aligned}$$

There is nothing special about the 36 million in this calculation. No matter what the number of rolls, once we simplify the average, we get the same $\sum_{i=2}^{12} iP(i)$. While the actual average value of a large number of rolls will not be exactly 7, the average should be close to 7 when the number of rolls is large. Turning this around, if the average is not close to 7, we should suspect that the dice are not fair.

A variable, say X, that can take certain values, each with a corresponding probability, is called a **random variable**; in the example above, the random variable was the sum of the two dice. If the possible values for X are x_1, x_2,…,x_n, then the expected value of the random variable is $E(X) = \sum_{i=1}^{n} x_i P(x_i)$. The expected value is also called the **mean**.

When the number of possible values for X is finite, we say that X is a discrete random variable. In many applications of probability, the number of possible values of a random variable is very large, perhaps even infinite. To deal with the infinite case we need a different approach, and since there is a sum involved, it should not be wholly surprising that integration turns out to be a useful tool. It then turns out that even when the number of possibilities is large but finite, it is frequently easier to pretend that the number is infinite. Suppose, for example, that a dart is thrown at a dart board. Since the dart board consists of a finite number of atoms, there are in some sense only a finite number of places for the dart to land, but it is easier to explore the probabilities involved by pretending that the dart can land on any point in the usual x-y plane.

DEFINITION 9.8.1 Let $f : \mathbb{R} \to \mathbb{R}$ be a function. If $f(x) \geq 0$ for every x and $\int_{-\infty}^{\infty} f(x)\,dx = 1$ then f is a **probability density function**. □

We associate a probability density function with a random variable X by stipulating that the probability that X is between a and b is $\int_a^b f(x)\,dx$. Because of the requirement that the integral from $-\infty$ to ∞ be 1, all probabilities are less than or equal to 1, and the probability that X takes on some value between $-\infty$ and ∞ is 1, as it should be.

EXAMPLE 9.8.2 Consider again the two dice example; we can view it in a way that more resembles the probability density function approach. Consider a random variable X that takes on any real value with probabilities given by the probability density function in figure 9.8.1. The function f consists of just the top edges of the rectangles, with vertical sides drawn for clarity; the function is zero below 1.5 and above 12.5. The area of each rectangle is the probability of rolling the sum in the middle of the bottom of the rectangle, or

$$P(n) = \int_{n-1/2}^{n+1/2} f(x)\,dx.$$

The probability of rolling a 4, 5, or 6 is

$$P(n) = \int_{7/2}^{13/2} f(x)\,dx.$$

Of course, we could also compute probabilities that don't make sense in the context of the dice, such as the probability that X is between 4 and 5.8. □

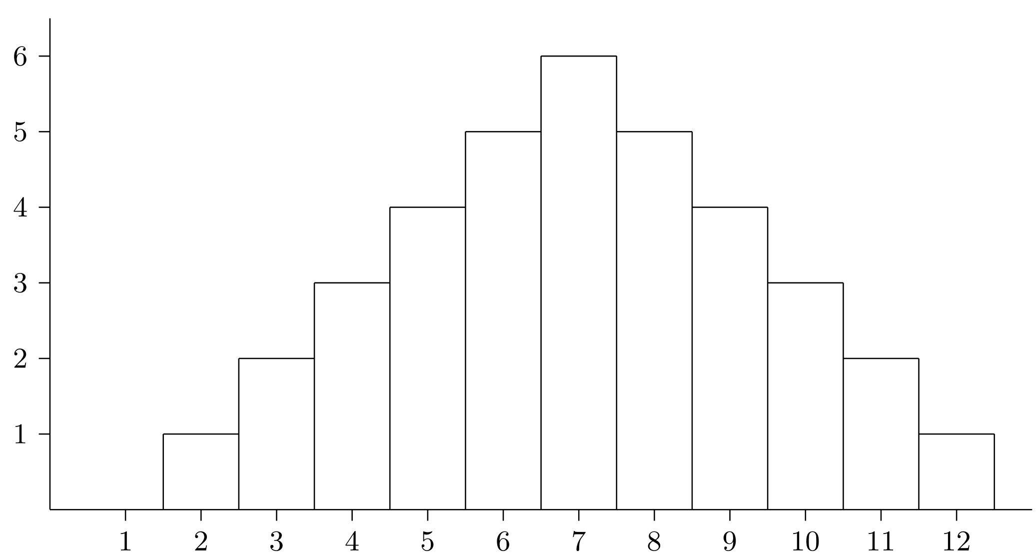


Figure 9.8.1 A probability density function for two dice.

The function

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

is called the **cumulative distribution function** or simply (probability) distribution.

EXAMPLE 9.8.3 Suppose that $a < b$ and

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise.} \end{cases}$$

Then $f(x)$ is the **uniform probability density function** on $[a, b]$. and the corresponding distribution is the **uniform distribution** on $[a, b]$. □

EXAMPLE 9.8.4 Consider the function $f(x) = e^{-x^2/2}$. What can we say about

$$\int_{-\infty}^{\infty} e^{-x^2/2}\, dx?$$

We cannot find an antiderivative of f, but we can see that this integral is some finite number. Notice that $0 < f(x) = e^{-x^2/2} \le e^{-x/2}$ for $|x| > 1$. This implies that the area under $e^{-x^2/2}$ is less than the area under $e^{-x/2}$, over the interval $[1, \infty)$. It is easy to compute the latter area, namely

$$\int_{1}^{\infty} e^{-x/2}\, dx = \frac{2}{\sqrt{e}},$$

so

$$\int_{1}^{\infty} e^{-x^2/2}\, dx$$

is some finite number smaller than $2/\sqrt{e}$. Because f is symmetric around the y-axis,

$$\int_{-\infty}^{-1} e^{-x^2/2}\, dx = \int_{1}^{\infty} e^{-x^2/2}\, dx.$$

This means that

$$\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \int_{-\infty}^{-1} e^{-x^2/2}\, dx + \int_{-1}^{1} e^{-x^2/2}\, dx + \int_{1}^{\infty} e^{-x^2/2}\, dx = A$$

for some finite positive number A. Now if we let $g(x) = f(x)/A$,

$$\int_{-\infty}^{\infty} g(x)\, dx = \frac{1}{A}\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \frac{1}{A}A = 1,$$

so g is a probability density function. It turns out to be very useful, and is called the **standard normal probability density function** or more informally the **bell curve**,

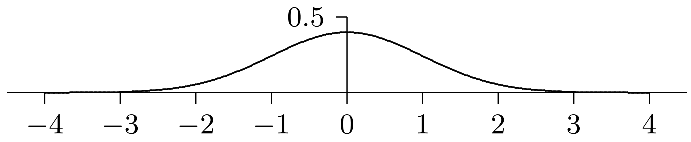


Figure 9.8.2 The bell curve.

giving rise to the **standard normal distribution**. See figure 9.8.2 for the graph of the bell curve. □

We have shown that A is some finite number without computing it; we cannot compute it with the techniques we have available. By using some techniques from multivariable calculus, it can be shown that $A = \sqrt{2\pi}$.

EXAMPLE 9.8.5 The **exponential distribution** has probability density function

$$f(x) = \begin{cases} 0 & x < 0 \\ ce^{-cx} & x \ge 0 \end{cases}$$

where c is a positive constant. □

The mean or expected value of a random variable is quite useful, as hinted at in our discussion of dice. Recall that the mean for a discrete random variable is $E(X) = \sum_{i=1}^{n} x_i P(x_i)$. In the more general context we use an integral in place of the sum.

DEFINITION 9.8.6 The **mean** of a random variable X with probability density function f is $\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$, provided the integral converges. □

When the mean exists it is unique, since it is the result of an explicit calculation. The mean does not always exist.

The mean might look familiar; it is essentially identical to the center of mass of a one-dimensional beam, as discussed in section 9.6. The probability density function f plays the role of the physical density function, but now the "beam" has infinite length. If we consider only a finite portion of the beam, say between a and b, then the center of mass is

$$\bar{x} = \frac{\int_a^b x f(x)\,dx}{\int_a^b f(x)\,dx}.$$

If we extend the beam to infinity, we get

$$\bar{x} = \frac{\displaystyle\int_{-\infty}^{\infty} x f(x)\,dx}{\displaystyle\int_{-\infty}^{\infty} f(x)\,dx} = \int_{-\infty}^{\infty} x f(x)\,dx = E(X),$$

because $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$. In the center of mass interpretation, this integral is the total mass of the beam, which is always 1 when f is a probability density function.

EXAMPLE 9.8.7 The mean of the standard normal distribution is

$$\int_{-\infty}^{\infty} x\frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx.$$

We compute the two halves:

$$\int_{-\infty}^{0} x\frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx = \lim_{D\to-\infty} -\left.\frac{e^{-x^2/2}}{\sqrt{2\pi}}\right|_{D}^{0} = -\frac{1}{\sqrt{2\pi}}$$

and

$$\int_{0}^{\infty} x\frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx = \lim_{D\to\infty} -\left.\frac{e^{-x^2/2}}{\sqrt{2\pi}}\right|_{0}^{D} = \frac{1}{\sqrt{2\pi}}.$$

The sum of these is 0, which is the mean. □

While the mean is very useful, it typically is not enough information to properly evaluate a situation. For example, suppose we could manufacture an 11-sided die, with the faces numbered 2 through 12 so that each face is equally likely to be down when the die is rolled. The value of a roll is the value on this lower face. Rolling the die gives the same range of values as rolling two ordinary dice, but now each value occurs with probability 1/11. The expected value of a roll is

$$\frac{2}{11} + \frac{3}{11} + \cdots + \frac{12}{11} = 7.$$

The mean does not distinguish the two cases, though of course they are quite different.

If f is a probability density function for a random variable X, with mean μ, we would like to measure how far a "typical" value of X is from μ. One way to measure this distance is $(X-\mu)^2$; we square the difference so as to measure all distances as positive. To get the typical such squared distance, we compute the mean. For two dice, for example, we get

$$(2-7)^2\frac{1}{36}+(3-7)^2\frac{2}{36}+\cdots+(7-7)^2\frac{6}{36}+\cdots(11-7)^2\frac{2}{36}+(12-7)^2\frac{1}{36}=\frac{35}{36}.$$

Because we squared the differences this does not directly measure the typical distance we seek; if we take the square root of this we do get such a measure, $\sqrt{35/36}\approx 2.42$. Doing the computation for the strange 11-sided die we get

$$(2-7)^2\frac{1}{11}+(3-7)^2\frac{1}{11}+\cdots+(7-7)^2\frac{1}{11}+\cdots(11-7)^2\frac{1}{11}+(12-7)^2\frac{1}{11}=10,$$

with square root approximately 3.16. Comparing 2.42 to 3.16 tells us that the two-dice rolls clump somewhat more closely near 7 than the rolls of the weird die, which of course we already knew because these examples are quite simple.

To perform the same computation for a probability density function the sum is replaced by an integral, just as in the computation of the mean. The expected value of the squared distances is

$$V(X)=\int_{-\infty}^{\infty}(x-\mu)^2 f(x)\,dx,$$

called the **variance**. The square root of the variance is the **standard deviation**, denoted σ.

EXAMPLE 9.8.8 We compute the standard deviation of the standard normal distrubution. The variance is

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^2e^{-x^2/2}\,dx.$$

To compute the antiderivative, use integration by parts, with $u=x$ and $dv=xe^{-x^2/2}\,dx$. This gives

$$\int x^2e^{-x^2/2}\,dx=-xe^{-x^2/2}+\int e^{-x^2/2}\,dx.$$

We cannot do the new integral, but we know its value when the limits are $-\infty$ to ∞, from our discussion of the standard normal distribution. Thus

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^2e^{-x^2/2}\,dx=-\frac{1}{\sqrt{2\pi}}xe^{-x^2/2}\bigg|_{-\infty}^{\infty}+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-x^2/2}\,dx=0+\frac{1}{\sqrt{2\pi}}\sqrt{2\pi}=1.$$

The standard deviation is then $\sqrt{1}=1$. □

EXAMPLE 9.8.9 Here is a simple example showing how these ideas can be useful. Suppose it is known that, in the long run, 1 out of every 100 computer memory chips produced by a certain manufacturing plant is defective when the manufacturing process is running correctly. Suppose 1000 chips are selected at random and 15 of them are defective. This is more than the 'expected' number (10), but is it so many that we should suspect that something has gone wrong in the manufacturing process? We are interested in the probability that various numbers of defective chips arise; the probability distribution is discrete: there can only be a whole number of defective chips. But (under reasonable assumptions) the distribution is very close to a normal distribution, namely this one:

$$f(x) = \frac{1}{\sqrt{2\pi}\sqrt{1000(.01)(.99)}} \exp\left(\frac{-(x-10)^2}{2(1000)(.01)(.99)}\right),$$

which is pictured in figure 9.8.3 (recall that $\exp(x) = e^x$).

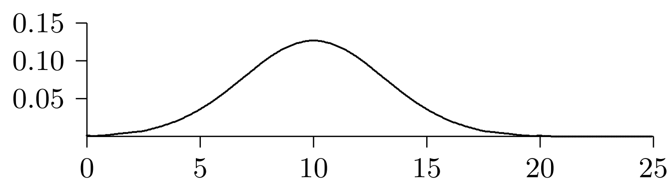


Figure 9.8.3 Normal density function for the defective chips example.

Now how do we measure how unlikely it is that under normal circumstances we would see 15 defective chips? We can't compute the probability of exactly 15 defective chips, as this would be $\int_{15}^{15} f(x)\,dx = 0$. We could compute $\int_{14.5}^{15.5} f(x)\,dx \approx 0.036$; this means there is only a 3.6% chance that the number of defective chips is 15. (We cannot compute these integrals exactly; computer software has been used to approximate the integral values in this discussion.) But this is misleading: $\int_{9.5}^{10.5} f(x)\,dx \approx 0.126$, which is larger, certainly, but still small, even for the "most likely" outcome. The most useful question, in most circumstances, is this: how likely is it that the number of defective chips is "far from" the mean? For example, how likely, or unlikely, is it that the number of defective chips is different by 5 or more from the expected value of 10? This is the probability that the number of defective chips is less than 5 or larger than 15, namely

$$\int_{-\infty}^{5} f(x)\,dx + \int_{15}^{\infty} f(x)\,dx \approx 0.11.$$

So there is an 11% chance that this happens—not large, but not tiny. Hence the 15 defective chips does not appear to be cause for alarm: about one time in nine we would expect to see the number of defective chips 5 or more away from the expected 10. How about 20? Here we compute

$$\int_{-\infty}^{0} f(x)\,dx + \int_{20}^{\infty} f(x)\,dx \approx 0.0015.$$

So there is only a 0.15% chance that the number of defective chips is more than 10 away from the mean; this would typically be interpreted as too suspicious to ignore—it shouldn't happen if the process is running normally.

The big question, of course, is what level of improbability should trigger concern? It depends to some degree on the application, and in particular on the consequences of getting it wrong in one direction or the other. If we're wrong, do we lose a little money? A lot of money? Do people die? In general, the standard choices are 5% and 1%. So what we should do is find the number of defective chips that has only, let us say, a 1% chance of occurring under normal circumstances, and use that as the relevant number. In other words, we want to know when

$$\int_{-\infty}^{10-r} f(x)\,dx + \int_{10+r}^{\infty} f(x)\,dx < 0.01.$$

A bit of trial and error shows that with $r = 8$ the value is about 0.011, and with $r = 9$ it is about 0.004, so if the number of defective chips is 19 or more, or 1 or fewer, we should look for problems. If the number is high, we worry that the manufacturing process has a problem, or conceivably that the process that tests for defective chips is not working correctly and is flagging good chips as defective. If the number is too low, we suspect that the testing procedure is broken, and is not detecting defective chips. □

Exercises 9.8.

1. Verify that $\displaystyle\int_{1}^{\infty} e^{-x/2}\,dx = 2/\sqrt{e}$.
2. Show that the function in example 9.8.5 is a probability density function. Compute the mean and standard deviation. ⇒
3. Compute the mean and standard deviation of the uniform distribution on $[a, b]$. (See example 9.8.3.) ⇒
4. What is the expected value of one roll of a fair six-sided die? ⇒
5. What is the expected sum of one roll of three fair six-sided dice? ⇒
6. Let μ and σ be real numbers with $\sigma > 0$. Show that

$$N(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

is a probability density function. You will not be able to compute this integral directly; use a substitution to convert the integral into the one from example 9.8.4. The function N is the probability density function of the **normal distribution** with mean μ and standard deviation σ. Show that the mean of the normal distribution is μ and the standard deviation is σ.

7. Let

$$f(x) = \begin{cases} \frac{1}{x^2} & x \ge 1 \\ 0 & x < 1 \end{cases}$$

Show that f is a probability density function, and that the distribution has no mean.

8. Let

$$f(x) = \begin{cases} x & -1 \le x \le 1 \\ 1 & 1 < x \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Is f a probability density function? Justify your answer.

9. If you have access to appropriate software, find r so that

$$\int_{-\infty}^{10-r} f(x)\,dx + \int_{10+r}^{\infty} f(x)\,dx \approx 0.05,$$

using the function of example 9.8.9. Discuss the impact of using this new value of r to decide whether to investigate the chip manufacturing process. ⇒

9.9 Arc Length

Here is another geometric application of the integral: find the length of a portion of a curve. As usual, we need to think about how we might approximate the length, and turn the approximation into an integral.

We already know how to compute one simple arc length, that of a line segment. If the endpoints are $P_0(x_0, y_0)$ and $P_1(x_1, y_1)$ then the length of the segment is the distance between the points, $\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}$, from the Pythagorean theorem, as illustrated in figure 9.9.1.

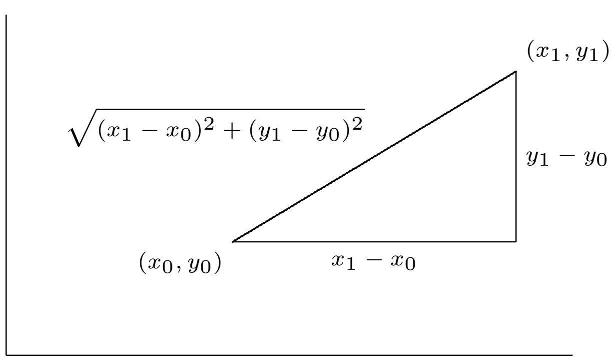


Figure 9.9.1 The length of a line segment.

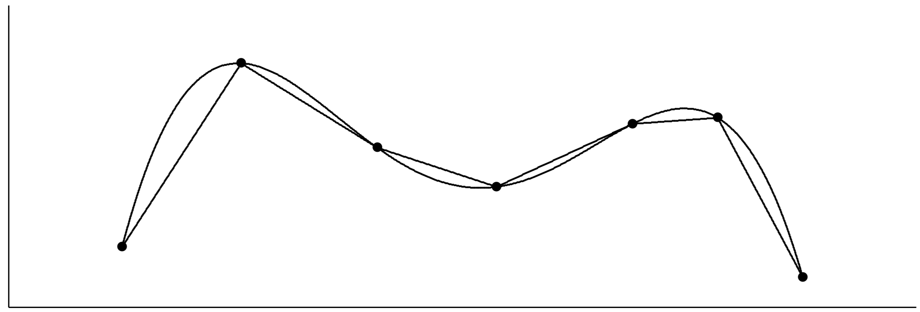


Figure 9.9.2 Approximating arc length with line segments.

Now if the graph of f is "nice" (say, differentiable) it appears that we can approximate the length of a portion of the curve with line segments, and that as the number of segments increases, and their lengths decrease, the sum of the lengths of the line segments will approach the true arc length; see figure 9.9.2.

Now we need to write a formula for the sum of the lengths of the line segments, in a form that we know becomes an integral in the limit. So we suppose we have divided the interval $[a, b]$ into n subintervals as usual, each with length $\Delta x = (b-a)/n$, and endpoints $a = x_0$, x_1, x_2, $\ldots$, $x_n = b$. The length of a typical line segment, joining $(x_i, f(x_i))$ to $(x_{i+1}, f(x_{i+1}))$, is $\sqrt{(\Delta x)^2 + (f(x_{i+1}) - f(x_i))^2}$. By the Mean Value Theorem (6.5.2), there is a number t_i in (x_i, x_{i+1}) such that $f'(t_i)\Delta x = f(x_{i+1}) - f(x_i)$, so the length of the line segment can be written as

$$\sqrt{(\Delta x)^2 + (f'(t_i))^2 \Delta x^2} = \sqrt{1 + (f'(t_i))^2}\,\Delta x.$$

The arc length is then

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} \sqrt{1 + (f'(t_i))^2}\,\Delta x = \int_a^b \sqrt{1 + (f'(x))^2}\,dx.$$

Note that the sum looks a bit different than others we have encountered, because the approximation contains a t_i instead of an x_i. In the past we have always used left endpoints (namely, x_i) to get a representative value of f on $[x_i, x_{i+1}]$; now we are using a different point, but the principle is the same.

To summarize, to compute the length of a curve on the interval $[a, b]$, we compute the integral

$$\int_a^b \sqrt{1 + (f'(x))^2}\,dx.$$

Unfortunately, integrals of this form are typically difficult or impossible to compute exactly, because usually none of our methods for finding antiderivatives will work. In practice this means that the integral will usually have to be approximated.

EXAMPLE 9.9.1 Let $f(x) = \sqrt{r^2 - x^2}$, the upper half circle of radius r. The length of this curve is half the circumference, namely πr. Let's compute this with the arc length formula. The derivative f' is $-x/\sqrt{r^2 - x^2}$ so the integral is

$$\int_{-r}^{r} \sqrt{1 + \frac{x^2}{r^2 - x^2}}\,dx = \int_{-r}^{r} \sqrt{\frac{r^2}{r^2 - x^2}}\,dx = r\int_{-r}^{r} \sqrt{\frac{1}{r^2 - x^2}}\,dx.$$

Using a trigonometric substitution, we find the antiderivative, namely $\arcsin(x/r)$. Notice that the integral is improper at both endpoints, as the function $\sqrt{1/(r^2 - x^2)}$ is undefined when $x = \pm r$. So we need to compute

$$\lim_{D\to -r^+} \int_{D}^{0} \sqrt{\frac{1}{r^2 - x^2}}\,dx + \lim_{D\to r^-} \int_{0}^{D} \sqrt{\frac{1}{r^2 - x^2}}\,dx.$$

This is not difficult, and has value π, so the original integral, with the extra r in front, has value πr as expected. □

Exercises 9.9.

1. Find the arc length of $f(x) = x^{3/2}$ on $[0, 2]$. ⇒
2. Find the arc length of $f(x) = x^2/8 - \ln x$ on $[1, 2]$. ⇒
3. Find the arc length of $f(x) = (1/3)(x^2 + 2)^{3/2}$ on the interval $[0, a]$. ⇒
4. Find the arc length of $f(x) = \ln(\sin x)$ on the interval $[\pi/4, \pi/3]$. ⇒
5. Let $a > 0$. Show that the length of $y = \cosh x$ on $[0, a]$ is equal to $\displaystyle\int_0^a \cosh x\,dx$.
6. Find the arc length of $f(x) = \cosh x$ on $[0, \ln 2]$. ⇒
7. Set up the integral to find the arc length of $\sin x$ on the interval $[0, \pi]$; do not evaluate the integral. If you have access to appropriate software, approximate the value of the integral. ⇒
8. Set up the integral to find the arc length of $y = xe^{-x}$ on the interval $[2, 3]$; do not evaluate the integral. If you have access to appropriate software, approximate the value of the integral. ⇒
9. Find the arc length of $y = e^x$ on the interval $[0, 1]$. (This can be done exactly; it is a bit tricky and a bit long.) ⇒

9.10 Surface Area

Another geometric question that arises naturally is: "What is the surface area of a volume?" For example, what is the surface area of a sphere? More advanced techniques are required to approach this question in general, but we can compute the areas of some volumes generated by revolution.

As usual, the question is: how might we approximate the surface area? For a surface obtained by rotating a curve around an axis, we can take a polygonal approximation to the curve, as in the last section, and rotate it around the same axis. This gives a surface composed of many "truncated cones;" a truncated cone is called a **frustum** of a cone. Figure 9.10.1 illustrates this approximation.

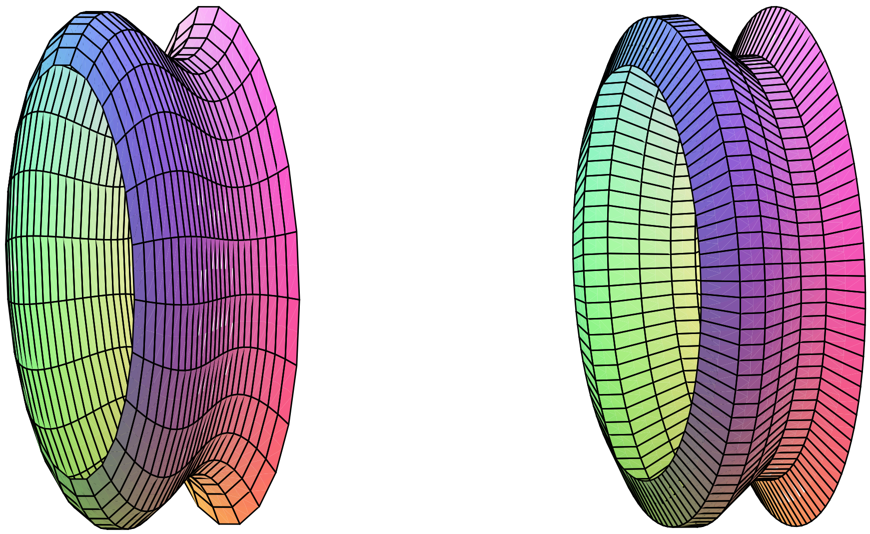


Figure 9.10.1 Approximating a surface (left) by portions of cones (right).

So we need to be able to compute the area of a frustum of a cone. Since the frustum can be formed by removing a small cone from the top of a larger one, we can compute the desired area if we know the surface area of a cone. Suppose a right circular cone has base radius r and slant height h. If we cut the cone from the vertex to the base circle and flatten it out, we obtain a sector of a circle with radius h and arc length $2\pi r$, as in figure 9.10.2. The angle at the center, in radians, is then $2\pi r/h$, and the area of the cone is equal to the area of the sector of the circle. Let A be the area of the sector; since the area of the entire circle is πh^2, we have

$$\frac{A}{\pi h^2} = \frac{2\pi r/h}{2\pi}$$
$$A = \pi r h.$$

Now suppose we have a frustum of a cone with slant height h and radii r_0 and r_1, as in figure 9.10.3. The area of the entire cone is $\pi r_1(h_0 + h)$, and the area of the small cone is $\pi r_0 h_0$; thus, the area of the frustum is $\pi r_1(h_0 + h) - \pi r_0 h_0 = \pi((r_1 - r_0)h_0 + r_1 h)$. By

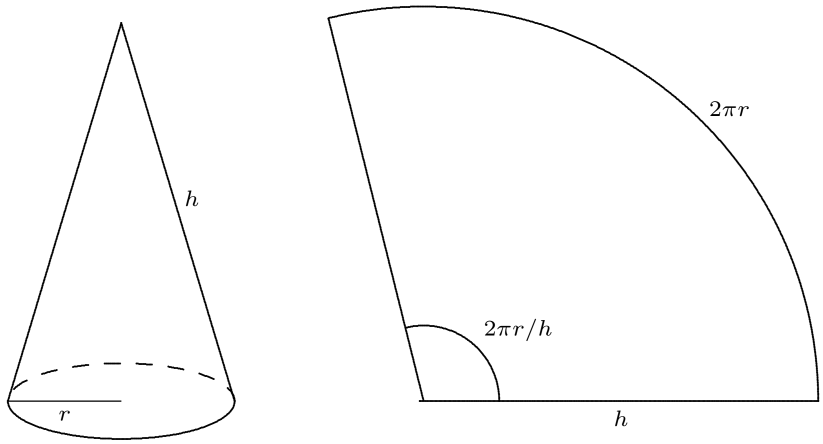


Figure 9.10.2 The area of a cone.

similar triangles,

$$\frac{h_0}{r_0} = \frac{h_0 + h}{r_1}.$$

With a bit of algebra this becomes $(r_1 - r_0)h_0 = r_0 h$; substitution into the area gives

$$\pi((r_1 - r_0)h_0 + r_1 h) = \pi(r_0 h + r_1 h) = \pi h(r_0 + r_1) = 2\pi \frac{r_0 + r_1}{2} h = 2\pi r h.$$

The final form is particularly easy to remember, with r equal to the average of r_0 and r_1, as it is also the formula for the area of a cylinder. (Think of a cylinder of radius r and height h as the frustum of a cone of infinite height.)

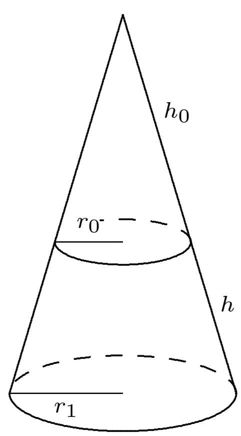


Figure 9.10.3 The area of a frustum.

Now we are ready to approximate the area of a surface of revolution. On one subinterval, the situation is as shown in figure 9.10.4. When the line joining two points on the curve is rotated around the x-axis, it forms a frustum of a cone. The area is

$$2\pi rh = 2\pi \frac{f(x_i) + f(x_{i+1})}{2} \sqrt{1 + (f'(t_i))^2}\,\Delta x.$$

Here $\sqrt{1 + (f'(t_i))^2}\,\Delta x$ is the length of the line segment, as we found in the previous section. Assuming f is a continuous function, there must be some x_i^* in $[x_i, x_{i+1}]$ such that $(f(x_i) + f(x_{i+1}))/2 = f(x_i^*)$, so the approximation for the surface area is

$$\sum_{i=0}^{n-1} 2\pi f(x_i^*) \sqrt{1 + (f'(t_i))^2}\,\Delta x.$$

This is not quite the sort of sum we have seen before, as it contains two different values in the interval $[x_i, x_{i+1}]$, namely x_i^* and t_i. Nevertheless, using more advanced techniques than we have available here, it turns out that

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} 2\pi f(x_i^*) \sqrt{1 + (f'(t_i))^2}\,\Delta x = \int_a^b 2\pi f(x) \sqrt{1 + (f'(x))^2}\,dx$$

is the surface area we seek. (Roughly speaking, this is because while x_i^* and t_i are distinct values in $[x_i, x_{i+1}]$, they get closer and closer to each other as the length of the interval shrinks.)

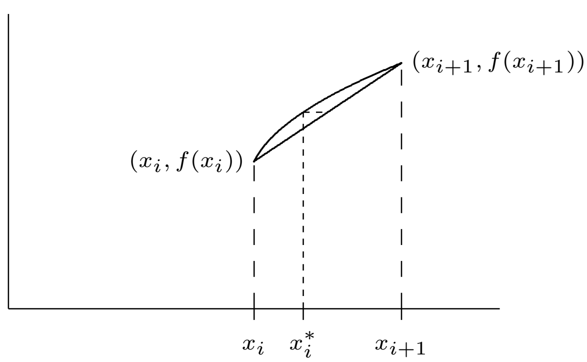


Figure 9.10.4 One subinterval.

EXAMPLE 9.10.1 We compute the surface area of a sphere of radius r. The sphere can be obtained by rotating the graph of $f(x) = \sqrt{r^2 - x^2}$ about the x-axis. The derivative

f' is $-x/\sqrt{r^2-x^2}$, so the surface area is given by

$$\begin{aligned} A &= 2\pi \int_{-r}^{r} \sqrt{r^2-x^2}\sqrt{1+\frac{x^2}{r^2-x^2}}\,dx \\ &= 2\pi \int_{-r}^{r} \sqrt{r^2-x^2}\sqrt{\frac{r^2}{r^2-x^2}}\,dx \\ &= 2\pi \int_{-r}^{r} r\,dx = 2\pi r \int_{-r}^{r} 1\,dx = 4\pi r^2 \end{aligned}$$

□

If the curve is rotated around the y axis, the formula is nearly identical, because the length of the line segment we use to approximate a portion of the curve doesn't change. Instead of the radius $f(x_i^*)$, we use the new radius $\bar{x}_i = (x_i + x_{i+1})/2$, and the surface area integral becomes

$$\int_a^b 2\pi x\sqrt{1+(f'(x))^2}\,dx.$$

EXAMPLE 9.10.2 Compute the area of the surface formed when $f(x) = x^2$ between 0 and 2 is rotated around the y-axis.

We compute $f'(x) = 2x$, and then

$$2\pi \int_0^2 x\sqrt{1+4x^2}\,dx = \frac{\pi}{6}(17^{3/2}-1),$$

by a simple substitution. □

Exercises 9.10.

1. Compute the area of the surface formed when $f(x) = 2\sqrt{1-x}$ between -1 and 0 is rotated around the x-axis. ⇒
2. Compute the surface area of example 9.10.2 by rotating $f(x) = \sqrt{x}$ around the x-axis.
3. Compute the area of the surface formed when $f(x) = x^3$ between 1 and 3 is rotated around the x-axis. ⇒
4. Compute the area of the surface formed when $f(x) = 2 + \cosh(x)$ between 0 and 1 is rotated around the x-axis. ⇒
5. Consider the surface obtained by rotating the graph of $f(x) = 1/x$, $x \geq 1$, around the x-axis. This surface is called **Gabriel's horn** or **Toricelli's trumpet**. In exercise 13 in section 9.7 we saw that Gabriel's horn has finite volume. Show that Gabriel's horn has infinite surface area.
6. Consider the circle $(x-2)^2 + y^2 = 1$. Sketch the surface obtained by rotating this circle about the y-axis. (The surface is called a **torus**.) What is the surface area? ⇒

7. Consider the ellipse with equation $x^2/4 + y^2 = 1$. If the ellipse is rotated around the x-axis it forms an **ellipsoid**. Compute the surface area. ⇒

8. Generalize the preceding result: rotate the ellipse given by $x^2/a^2 + y^2/b^2 = 1$ about the x-axis and find the surface area of the resulting ellipsoid. You should consider two cases, when $a > b$ and when $a < b$. Compare to the area of a sphere. ⇒

10

Polar Coordinates, Parametric Equations

10.1 Polar Coordinates

Coordinate systems are tools that let us use algebraic methods to understand geometry. While the **rectangular** (also called **Cartesian**) coordinates that we have been using are the most common, some problems are easier to analyze in alternate coordinate systems.

A coordinate system is a scheme that allows us to identify any point in the plane or in three-dimensional space by a set of numbers. In rectangular coordinates these numbers are interpreted, roughly speaking, as the lengths of the sides of a rectangle. In **polar coordinates** a point in the plane is identified by a pair of numbers (r, θ). The number θ measures the angle between the positive x-axis and a ray that goes through the point, as shown in figure 10.1.1; the number r measures the distance from the origin to the point. Figure 10.1.1 shows the point with rectangular coordinates $(1, \sqrt{3})$ and polar coordinates $(2, \pi/3)$, 2 units from the origin and $\pi/3$ radians from the positive x-axis.

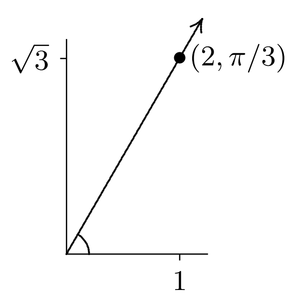


Figure 10.1.1 Polar coordinates of the point $(1, \sqrt{3})$.

Just as we describe curves in the plane using equations involving x and y, so can we describe curves using equations involving r and θ. Most common are equations of the form $r = f(\theta)$.

EXAMPLE 10.1.1 Graph the curve given by $r = 2$. All points with $r = 2$ are at distance 2 from the origin, so $r = 2$ describes the circle of radius 2 with center at the origin. □

EXAMPLE 10.1.2 Graph the curve given by $r = 1 + \cos\theta$. We first consider $y = 1 + \cos x$, as in figure 10.1.2. As θ goes through the values in $[0, 2\pi]$, the value of r tracks the value of y, forming the "cardioid" shape of figure 10.1.2. For example, when $\theta = \pi/2$, $r = 1 + \cos(\pi/2) = 1$, so we graph the point at distance 1 from the origin along the positive y-axis, which is at an angle of $\pi/2$ from the positive x-axis. When $\theta = 7\pi/4$, $r = 1 + \cos(7\pi/4) = 1 + \sqrt{2}/2 \approx 1.71$, and the corresponding point appears in the fourth quadrant. This illustrates one of the potential benefits of using polar coordinates: the equation for this curve in rectangular coordinates would be quite complicated. □

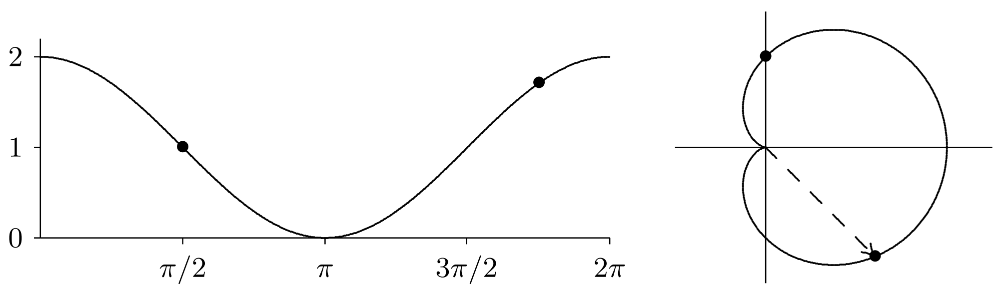


Figure 10.1.2 A cardioid: $y = 1 + \cos x$ on the left, $r = 1 + \cos\theta$ on the right.

Each point in the plane is associated with exactly one pair of numbers in the rectangular coordinate system; each point is associated with an infinite number of pairs in polar coordinates. In the cardioid example, we considered only the range $0 \le \theta \le 2\pi$, and already there was a duplicate: $(2, 0)$ and $(2, 2\pi)$ are the same point. Indeed, every value of θ outside the interval $[0, 2\pi)$ duplicates a point on the curve $r = 1 + \cos\theta$ when $0 \le \theta < 2\pi$. We can even make sense of polar coordinates like $(-2, \pi/4)$: go to the direction $\pi/4$ and then move a distance 2 in the opposite direction; see figure 10.1.3. As usual, a negative angle θ means an angle measured clockwise from the positive x-axis. The point in figure 10.1.3 also has coordinates $(2, 5\pi/4)$ and $(2, -3\pi/4)$.

The relationship between rectangular and polar coordinates is quite easy to understand. The point with polar coordinates (r, θ) has rectangular coordinates $x = r\cos\theta$ and $y = r\sin\theta$; this follows immediately from the definition of the sine and cosine functions. Using figure 10.1.3 as an example, the point shown has rectangular coordinates

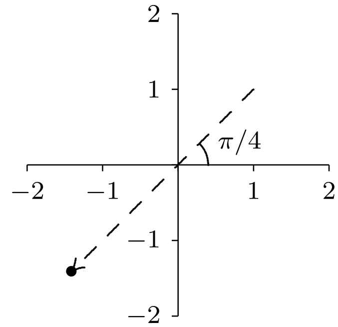


Figure 10.1.3 The point $(-2, \pi/4) = (2, 5\pi/4) = (2, -3\pi/4)$ in polar coordinates.

$x = (-2)\cos(\pi/4) = -\sqrt{2} \approx 1.4142$ and $y = (-2)\sin(\pi/4) = -\sqrt{2}$. This makes it very easy to convert equations from rectangular to polar coordinates.

EXAMPLE 10.1.3 Find the equation of the line $y = 3x + 2$ in polar coordinates. We merely substitute: $r \sin\theta = 3r\cos\theta + 2$, or $r = \dfrac{2}{\sin\theta - 3\cos\theta}$. □

EXAMPLE 10.1.4 Find the equation of the circle $(x - 1/2)^2 + y^2 = 1/4$ in polar coordinates. Again substituting: $(r\cos\theta - 1/2)^2 + r^2\sin^2\theta = 1/4$. A bit of algebra turns this into $r = \cos(t)$. You should try plotting a few (r, θ) values to convince yourself that this makes sense. □

EXAMPLE 10.1.5 Graph the polar equation $r = \theta$. Here the distance from the origin exactly matches the angle, so a bit of thought makes it clear that when $\theta \ge 0$ we get the spiral of Archimedes in figure 10.1.4. When $\theta < 0$, r is also negative, and so the full graph is the right hand picture in the figure. □

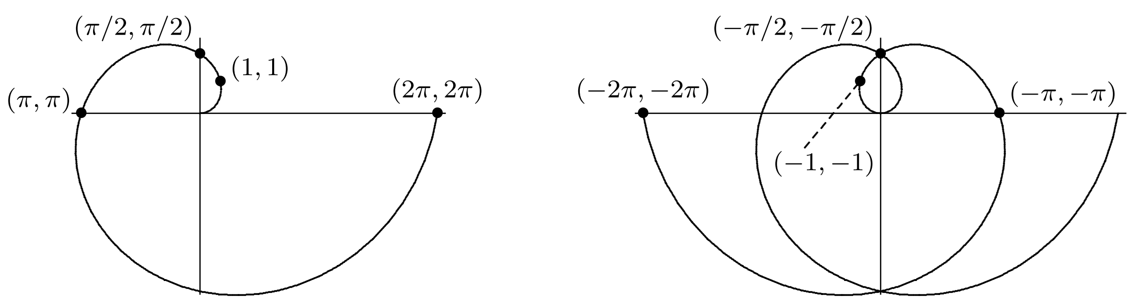


Figure 10.1.4 The spiral of Archimedes and the full graph of $r = \theta$.

Converting polar equations to rectangular equations can be somewhat trickier, and graphing polar equations directly is also not always easy.

EXAMPLE 10.1.6 Graph $r = 2\sin\theta$. Because the sine is periodic, we know that we will get the entire curve for values of θ in $[0, 2\pi)$. As θ runs from 0 to $\pi/2$, r increases from 0 to 2. Then as θ continues to π, r decreases again to 0. When θ runs from π to 2π, r is negative, and it is not hard to see that the first part of the curve is simply traced out again, so in fact we get the whole curve for values of θ in $[0, \pi)$. Thus, the curve looks something like figure 10.1.5. Now, this suggests that the curve could possibly be a circle, and if it is, it would have to be the circle $x^2 + (y-1)^2 = 1$. Having made this guess, we can easily check it. First we substitute for x and y to get $(r\cos\theta)^2 + (r\sin\theta - 1)^2 = 1$; expanding and simplifying does indeed turn this into $r = 2\sin\theta$. □

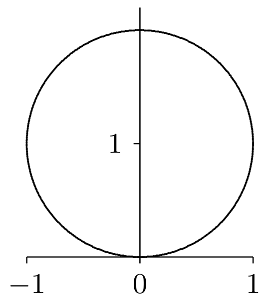


Figure 10.1.5 Graph of $r = 2\sin\theta$.

Exercises 10.1.

1. Plot these polar coordinate points on one graph: $(2, \pi/3)$, $(-3, \pi/2)$, $(-2, -\pi/4)$, $(1/2, \pi)$, $(1, 4\pi/3)$, $(0, 3\pi/2)$.

Find an equation in polar coordinates that has the same graph as the given equation in rectangular coordinates.

2. $y = 3x \Rightarrow$

3. $y = -4 \Rightarrow$

4. $xy^2 = 1 \Rightarrow$

5. $x^2 + y^2 = 5 \Rightarrow$

6. $y = x^3 \Rightarrow$

7. $y = \sin x \Rightarrow$

8. $y = 5x + 2 \Rightarrow$

9. $x = 2 \Rightarrow$

10. $y = x^2 + 1 \Rightarrow$

11. $y = 3x^2 - 2x \Rightarrow$

12. $y = x^2 + y^2 \Rightarrow$

Sketch the curve.

13. $r = \cos\theta$

14. $r = \sin(\theta + \pi/4)$

15. $r = -\sec\theta$

16. $r = \theta/2$, $\theta \geq 0$

17. $r = 1 + \theta^1/\pi^2$

18. $r = \cot\theta\csc\theta$

19. $r = \dfrac{1}{\sin\theta + \cos\theta}$

20. $r^2 = -2\sec\theta\csc\theta$

In the exercises below, find an equation in rectangular coordinates that has the same graph as the given equation in polar coordinates.

21. $r = \sin(3\theta)$ ⇒

22. $r = \sin^2\theta$ ⇒

23. $r = \sec\theta\csc\theta$ ⇒

24. $r = \tan\theta$ ⇒

10.2 Slopes in polar coordinates

When we describe a curve using polar coordinates, it is still a curve in the x-y plane. We would like to be able to compute slopes and areas for these curves using polar coordinates.

We have seen that $x = r\cos\theta$ and $y = r\sin\theta$ describe the relationship between polar and rectangular coordinates. If in turn we are interested in a curve given by $r = f(\theta)$, then we can write $x = f(\theta)\cos\theta$ and $y = f(\theta)\sin\theta$, describing x and y in terms of θ alone. The first of these equations describes θ implicitly in terms of x, so using the chain rule we may compute

$$\frac{dy}{dx} = \frac{dy}{d\theta}\frac{d\theta}{dx}.$$

Since $d\theta/dx = 1/(dx/d\theta)$, we can instead compute

$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{f(\theta)\cos\theta + f'(\theta)\sin\theta}{-f(\theta)\sin\theta + f'(\theta)\cos\theta}.$$

EXAMPLE 10.2.1 Find the points at which the curve given by $r = 1 + \cos\theta$ has a vertical or horizontal tangent line. Since this function has period 2π, we may restrict our attention to the interval $[0, 2\pi)$ or $(-\pi, \pi]$, as convenience dictates. First, we compute the slope:

$$\frac{dy}{dx} = \frac{(1+\cos\theta)\cos\theta - \sin\theta\sin\theta}{-(1+\cos\theta)\sin\theta - \sin\theta\cos\theta} = \frac{\cos\theta + \cos^2\theta - \sin^2\theta}{-\sin\theta - 2\sin\theta\cos\theta}.$$

This fraction is zero when the numerator is zero (and the denominator is not zero). The numerator is $2\cos^2\theta + \cos\theta - 1$ so by the quadratic formula

$$\cos\theta = \frac{-1 \pm \sqrt{1 + 4\cdot 2}}{4} = -1 \quad \text{or} \quad \frac{1}{2}.$$

This means θ is π or $\pm\pi/3$. However, when $\theta = \pi$, the denominator is also 0, so we cannot conclude that the tangent line is horizontal.

Setting the denominator to zero we get

$$-\theta - 2\sin\theta\cos\theta = 0$$
$$\sin\theta(1 + 2\cos\theta) = 0,$$

so either $\sin\theta = 0$ or $\cos\theta = -1/2$. The first is true when θ is 0 or π, the second when θ is $2\pi/3$ or $4\pi/3$. However, as above, when $\theta = \pi$, the numerator is also 0, so we cannot conclude that the tangent line is vertical. Figure 10.2.1 shows points corresponding to θ equal to 0, $\pm\pi/3$, $2\pi/3$ and $4\pi/3$ on the graph of the function. Note that when $\theta = \pi$ the curve hits the origin and does not have a tangent line. □

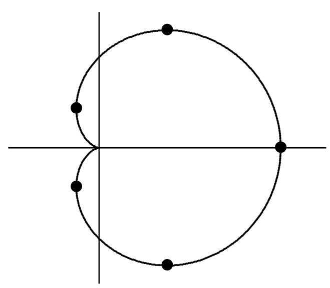


Figure 10.2.1 Points of vertical and horizontal tangency for $r = 1 + \cos\theta$.

We know that the second derivative $f''(x)$ is useful in describing functions, namely, in describing concavity. We can compute $f''(x)$ in terms of polar coordinates as well. We already know how to write $dy/dx = y'$ in terms of θ, then

$$\frac{d}{dx}\frac{dy}{dx} = \frac{dy'}{dx} = \frac{dy'}{d\theta}\frac{d\theta}{dx} = \frac{dy'/d\theta}{dx/d\theta}.$$

EXAMPLE 10.2.2 We find the second derivative for the cardioid $r = 1 + \cos\theta$:

$$\frac{d}{d\theta}\frac{\cos\theta + \cos^2\theta - \sin^2\theta}{-\sin\theta - 2\sin\theta\cos\theta} \cdot \frac{1}{dx/d\theta} = \cdots = \frac{3(1+\cos\theta)}{(\sin\theta + 2\sin\theta\cos\theta)^2} \cdot \frac{1}{-(\sin\theta + 2\sin\theta\cos\theta)}$$
$$= \frac{-3(1+\cos\theta)}{(\sin\theta + 2\sin\theta\cos\theta)^3}.$$

The ellipsis here represents rather a substantial amount of algebra. We know from above that the cardioid has horizontal tangents at $\pm\pi/3$; substituting these values into the second derivative we get $y''(\pi/3) = -\sqrt{3}/2$ and $y''(-\pi/3) = \sqrt{3}/2$, indicating concave down and concave up respectively. This agrees with the graph of the function. □

Exercises 10.2.

Compute $y' = dy/dx$ and $y'' = d^2y/dx^2$.

1. $r = \theta \Rightarrow$

2. $r = 1 + \sin\theta \Rightarrow$

3. $r = \cos\theta \Rightarrow$

4. $r = \sin\theta \Rightarrow$

5. $r = \sec\theta \Rightarrow$

6. $r = \sin(2\theta) \Rightarrow$

Sketch the curves over the interval $[0, 2\pi]$ unless otherwise stated.

7. $r = \sin\theta + \cos\theta$

8. $r = 2 + 2\sin\theta$

9. $r = \dfrac{3}{2} + \sin\theta$

10. $r = 2 + \cos\theta$

11. $r = \dfrac{1}{2} + \cos\theta$

12. $r = \cos(\theta/2), 0 \le \theta \le 4\pi$

13. $r = \sin(\theta/3), 0 \le \theta \le 6\pi$

14. $r = \sin^2\theta$

15. $r = 1 + \cos^2(2\theta)$

16. $r = \sin^2(3\theta)$

17. $r = \tan\theta$

18. $r = \sec(\theta/2), 0 \le \theta \le 4\pi$

19. $r = 1 + \sec\theta$

20. $r = \dfrac{1}{1 - \cos\theta}$

21. $r = \dfrac{1}{1 + \sin\theta}$

22. $r = \cot(2\theta)$

23. $r = \pi/\theta, 0 \le \theta \le \infty$

24. $r = 1 + \pi/\theta, 0 \le \theta \le \infty$

25. $r = \sqrt{\pi/\theta}, 0 \le \theta \le \infty$

10.3 Areas in polar coordinates

We can use the equation of a curve in polar coordinates to compute some areas bounded by such curves. The basic approach is the same as with any application of integration: find an approximation that approaches the true value. For areas in rectangular coordinates, we approximated the region using rectangles; in polar coordinates, we use sectors of circles, as depicted in figure 10.3.1. Recall that the area of a sector of a circle is $\alpha r^2/2$, where α is the angle subtended by the sector. If the curve is given by $r = f(\theta)$, and the angle subtended by a small sector is $\Delta\theta$, the area is $(\Delta\theta)(f(\theta))^2/2$. Thus we approximate the total area as

$$\sum_{i=0}^{n-1} \frac{1}{2} f(\theta_i)^2 \, \Delta\theta.$$

In the limit this becomes

$$\int_a^b \frac{1}{2} f(\theta)^2 \, d\theta.$$

EXAMPLE 10.3.1 We find the area inside the cardioid $r = 1 + \cos\theta$.

$$\int_0^{2\pi} \frac{1}{2}(1+\cos\theta)^2\, d\theta = \frac{1}{2}\int_0^{2\pi} 1+2\cos\theta+\cos^2\theta\, d\theta = \frac{1}{2}\,(\theta + 2\sin\theta + \frac{\theta}{2} + \frac{\sin 2\theta}{4})\Big|_0^{2\pi} = \frac{3\pi}{2}.$$

□

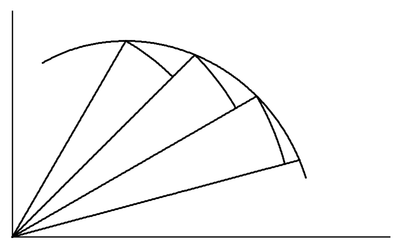


Figure 10.3.1 Approximating area by sectors of circles.

EXAMPLE 10.3.2 We find the area between the circles $r = 2$ and $r = 4\sin\theta$, as shown in figure 10.3.2. The two curves intersect where $2 = 4\sin\theta$, or $\sin\theta = 1/2$, so $\theta = \pi/6$ or $5\pi/6$. The area we want is then

$$\frac{1}{2}\int_{\pi/6}^{5\pi/6} 16\sin^2\theta - 4\, d\theta = \frac{4}{3}\pi + 2\sqrt{3}.$$

□

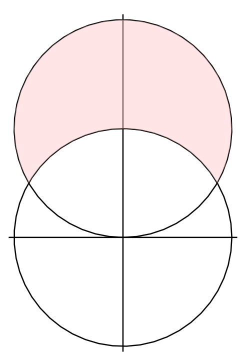


Figure 10.3.2 An area between curves.

This example makes the process appear more straightforward than it is. Because points have many different representations in polar coordinates, it is not always so easy to identify points of intersection.

EXAMPLE 10.3.3 We find the shaded area in the first graph of figure 10.3.3 as the difference of the other two shaded areas. The cardioid is $r = 1 + \sin\theta$ and the circle is $r = 3\sin\theta$. We attempt to find points of intersection:

$$\begin{aligned} 1 + \sin\theta &= 3\sin\theta \\ 1 &= 2\sin\theta \\ 1/2 &= \sin\theta. \end{aligned}$$

This has solutions $\theta = \pi/6$ and $5\pi/6$; $\pi/6$ corresponds to the intersection in the first quadrant that we need. Note that no solution of this equation corresponds to the intersection point at the origin, but fortunately that one is obvious. The cardioid goes through the origin when $\theta = -\pi/2$; the circle goes through the origin at multiples of π, starting with 0.

Now the larger region has area

$$\frac{1}{2}\int_{-\pi/2}^{\pi/6} (1 + \sin\theta)^2\, d\theta = \frac{\pi}{2} - \frac{9}{16}\sqrt{3}$$

and the smaller has area

$$\frac{1}{2}\int_{0}^{\pi/6} (3\sin\theta)^2\, d\theta = \frac{3\pi}{8} - \frac{9}{16}\sqrt{3}$$

so the area we seek is $\pi/8$. □

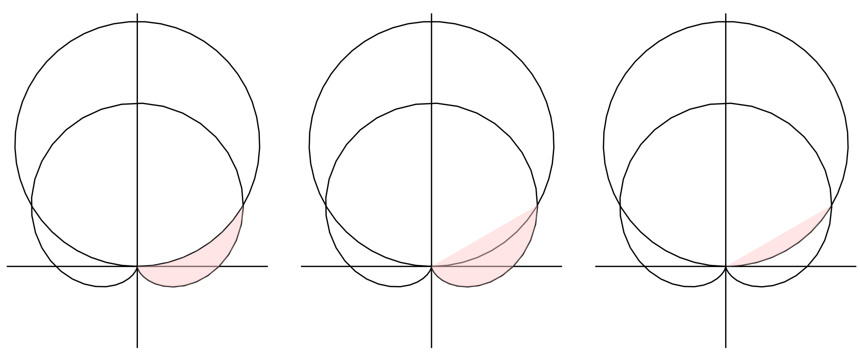


Figure 10.3.3 An area between curves.

Exercises 10.3.

Find the area enclosed by the curve.

1. $r = \sqrt{\sin\theta}$ ⇒

2. $r = 2 + \cos\theta$ ⇒

3. $r = \sec\theta, \pi/6 \le \theta \le \pi/3$ ⇒

4. $r = \cos\theta, 0 \le \theta \le \pi/3$ ⇒

5. $r = 2a\cos\theta, a > 0$ ⇒

6. $r = 4 + 3\sin\theta$ ⇒

7. Find the area inside the loop formed by $r = \tan(\theta/2)$. ⇒

8. Find the area inside one loop of $r = \cos(3\theta)$. ⇒

9. Find the area inside one loop of $r = \sin^2\theta$. ⇒

10. Find the area inside the small loop of $r = (1/2) + \cos\theta$. ⇒

11. Find the area inside $r = (1/2) + \cos\theta$, including the area inside the small loop. ⇒

12. Find the area inside one loop of $r^2 = \cos(2\theta)$. ⇒

13. Find the area enclosed by $r = \tan\theta$ and $r = \dfrac{\csc\theta}{\sqrt{2}}$. ⇒

14. Find the area inside $r = 2\cos\theta$ and outside $r = 1$. ⇒

15. Find the area inside $r = 2\sin\theta$ and above the line $r = (3/2)\csc\theta$. ⇒

16. Find the area inside $r = \theta$, $0 \le \theta \le 2\pi$. ⇒

17. Find the area inside $r = \sqrt{\theta}$, $0 \le \theta \le 2\pi$. ⇒

18. Find the area inside both $r = \sqrt{3}\cos\theta$ and $r = \sin\theta$. ⇒

19. Find the area inside both $r = 1 - \cos\theta$ and $r = \cos\theta$. ⇒

20. The center of a circle of radius 1 is on the circumference of a circle of radius 2. Find the area of the region inside both circles. ⇒

21. Find the shaded area in figure 10.3.4. The curve is $r = \theta$, $0 \le \theta \le 3\pi$. ⇒

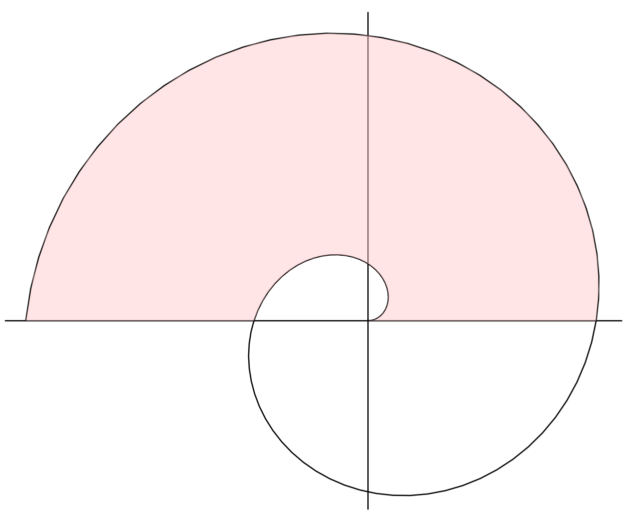


Figure 10.3.4 An area bounded by the spiral of Archimedes.

10.4 Parametric Equations

When we computed the derivative dy/dx using polar coordinates, we used the expressions $x = f(\theta)\cos\theta$ and $y = f(\theta)\sin\theta$. These two equations completely specify the curve, though the form $r = f(\theta)$ is simpler. The expanded form has the virtue that it can easily be generalized to describe a wider range of curves than can be specified in rectangular or polar coordinates.

Suppose $f(t)$ and $g(t)$ are functions. Then the equations $x = f(t)$ and $y = g(t)$ describe a curve in the plane. In the case of the polar coordinates equations, the variable t is replaced by θ which has a natural geometric interpretation. But t in general is simply an arbitrary variable, often called in this case a **parameter**, and this method of specifying a curve is known as **parametric equations**. One important interpretation of t is *time*. In this interpretation, the equations $x = f(t)$ and $y = g(t)$ give the position of an object at time t.

EXAMPLE 10.4.1 Describe the path of an object that moves so that its position at time t is given by $x = \cos t$, $y = \cos^2 t$. We see immediately that $y = x^2$, so the path lies on this parabola. The path is not the entire parabola, however, since $x = \cos t$ is always between -1 and 1. It is now easy to see that the object oscillates back and forth on the parabola between the endpoints $(1, 1)$ and $(-1, 1)$, and is at point $(1, 1)$ at time $t = 0$. □

It is sometimes quite easy to describe a complicated path in parametric equations when rectangular and polar coordinate expressions are difficult or impossible to devise.

EXAMPLE 10.4.2 A wheel of radius 1 rolls along a straight line, say the x-axis. A point on the rim of the wheel will trace out a curve, called a cycloid. Assume the point starts at the origin; find parametric equations for the curve.

Figure 10.4.1 illustrates the generation of the curve (click on the AP link to see an animation). The wheel is shown at its starting point, and again after it has rolled through about 490 degrees. We take as our parameter t the angle through which the wheel has turned, measured as shown clockwise from the line connecting the center of the wheel to the ground. Because the radius is 1, the center of the wheel has coordinates $(t, 1)$. We seek to write the coordinates of the point on the rim as $(t + \Delta x, 1 + \Delta y)$, where Δx and Δy are as shown in figure 10.4.2. These values are nearly the sine and cosine of the angle t, from the unit circle definition of sine and cosine. However, some care is required because we are measuring t from a nonstandard starting line and in a clockwise direction, as opposed to the usual counterclockwise direction. A bit of thought reveals that $\Delta x = -\sin t$ and $\Delta y = -\cos t$. Thus the parametric equations for the cycloid are $x = t - \sin t$, $y = 1 - \cos t$. □

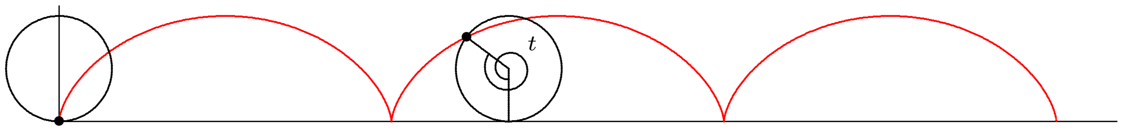


Figure 10.4.1 A cycloid. (AP)

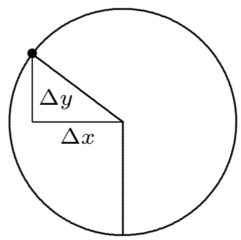


Figure 10.4.2 The wheel.

Exercises 10.4.

1. What curve is described by $x = t^2$, $y = t^4$? If t is interpreted as time, describe how the object moves on the curve.
2. What curve is described by $x = 3\cos t$, $y = 3\sin t$? If t is interpreted as time, describe how the object moves on the curve.
3. What curve is described by $x = 3\cos t$, $y = 2\sin t$? If t is interpreted as time, describe how the object moves on the curve.
4. What curve is described by $x = 3\sin t$, $y = 3\cos t$? If t is interpreted as time, describe how the object moves on the curve.
5. Sketch the curve described by $x = t^3 - t$, $y = t^2$. If t is interpreted as time, describe how the object moves on the curve.
6. A wheel of radius 1 rolls along a straight line, say the x-axis. A point P is located halfway between the center of the wheel and the rim; assume P starts at the point $(0, 1/2)$. As the wheel rolls, P traces a curve; find parametric equations for the curve. ⇒
7. A wheel of radius 1 rolls around the outside of a circle of radius 3. A point P on the rim of the wheel traces out a curve called a **hypercycloid**, as indicated in figure 10.4.3. Assuming P starts at the point $(3, 0)$, find parametric equations for the curve. ⇒
8. A wheel of radius 1 rolls around the inside of a circle of radius 3. A point P on the rim of the wheel traces out a curve called a **hypocycloid**, as indicated in figure 10.4.3. Assuming P starts at the point $(3, 0)$, find parametric equations for the curve. ⇒
9. An **involute** of a circle is formed as follows: Imagine that a long (that is, infinite) string is wound tightly around a circle, and that you grasp the end of the string and begin to unwind it, keeping the string taut. The end of the string traces out the involute. Find parametric equations for this curve, using a circle of radius 1, and assuming that the string unwinds counter-clockwise and the end of the string is initially at $(1, 0)$. Figure 10.4.4 shows part of the curve; the dotted lines represent the string at a few different times. ⇒

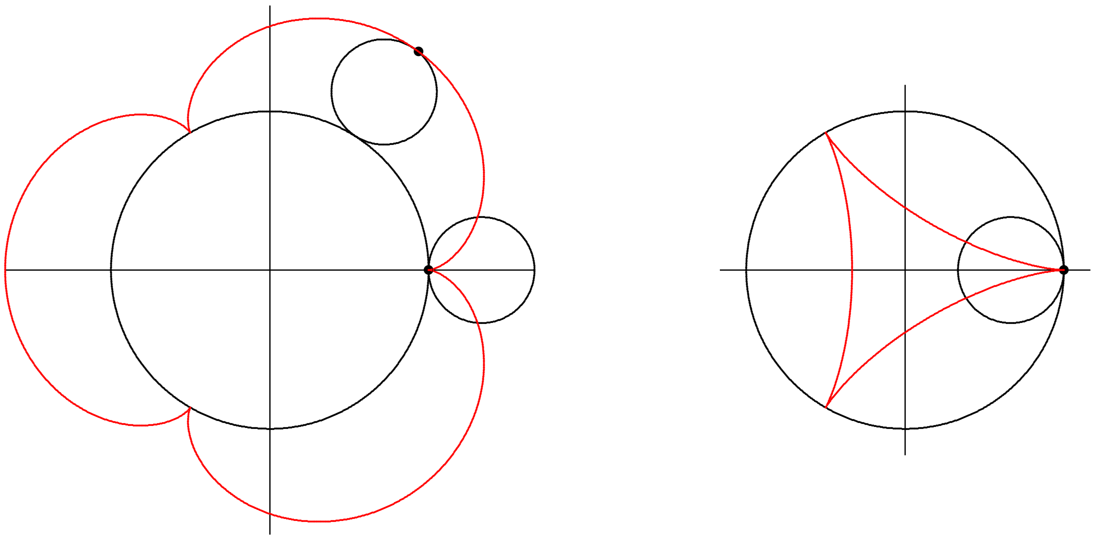


Figure 10.4.3 A hypercycloid and a hypocycloid.

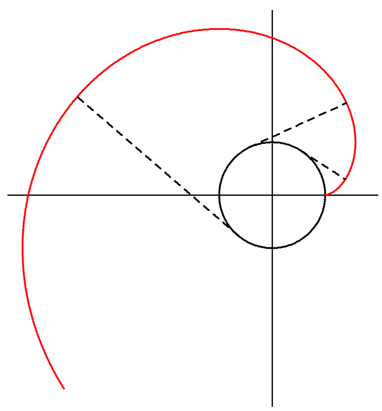


Figure 10.4.4 An involute of a circle.

10.5 Calculus with Parametric Equations

We have already seen how to compute slopes of curves given by parametric equations—it is how we computed slopes in polar coordinates.

EXAMPLE 10.5.1 Find the slope of the cycloid $x = t - \sin t$, $y = 1 - \cos t$. We compute $x' = 1 - \cos t$, $y' = \sin t$, so

$$\frac{dy}{dx} = \frac{\sin t}{1 - \cos t}.$$

Note that when t is an odd multiple of π, like π or 3π, this is $(0/2) = 0$, so there is a horizontal tangent line, in agreement with figure 10.4.1. At even multiples of π, the fraction is $0/0$, which is undefined. The figure shows that there is no tangent line at such points. □

Areas can be a bit trickier with parametric equations, depending on the curve and the area desired. We can potentially compute areas between the curve and the x-axis quite easily.

EXAMPLE 10.5.2 Find the area under one arch of the cycloid $x = t - \sin t$, $y = 1 - \cos t$. We would like to compute

$$\int_0^{2\pi} y\, dx,$$

but we do not know y in terms of x. However, the parametric equations allow us to make a substitution: use $y = 1 - \cos t$ to replace y, and compute $dx = (1 - \cos t)\, dt$. Then the integral becomes

$$\int_0^{2\pi} (1 - \cos t)(1 - \cos t)\, dt = 3\pi.$$

Note that we need to convert the original x limits to t limits using $x = t - \sin t$. When $x = 0$, $t = \sin t$, which happens only when $t = 0$. Likewise, when $x = 2\pi$, $t - 2\pi = \sin t$ and $t = 2\pi$. Alternately, because we understand how the cycloid is produced, we can see directly that one arch is generated by $0 \le t \le 2\pi$. In general, of course, the t limits will be different than the x limits. □

This technique will allow us to compute some quite interesting areas, as illustrated by the exercises.

As a final example, we see how to compute the length of a curve given by parametric equations. Section 9.9 investigates arc length for functions given as y in terms of x, and develops the formula for length:

$$\int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

Using some properties of derivatives, including the chain rule, we can convert this to use parametric equations $x = f(t)$, $y = g(t)$:

$$\begin{aligned}
\int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx &= \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dx}{dt}\right)^2 \left(\frac{dy}{dx}\right)^2}\, \frac{dt}{dx}\, dx \\
&= \int_u^v \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt \\
&= \int_u^v \sqrt{(f'(t))^2 + (g'(t))^2}\, dt.
\end{aligned}$$

Here u and v are the t limits corresponding to the x limits a and b.

EXAMPLE 10.5.3 Find the length of one arch of the cycloid. From $x = t - \sin t$, $y = 1 - \cos t$, we get the derivatives $f' = 1 - \cos t$ and $g' = \sin t$, so the length is

$$\int_0^{2\pi} \sqrt{(1-\cos t)^2 + \sin^2 t}\, dt = \int_0^{2\pi} \sqrt{2 - 2\cos t}\, dt.$$

Now we use the formula $\sin^2(t/2) = (1 - \cos(t))/2$ or $4\sin^2(t/2) = 2 - 2\cos t$ to get

$$\int_0^{2\pi} \sqrt{4\sin^2(t/2)}\, dt.$$

Since $0 \le t \le 2\pi$, $\sin(t/2) \ge 0$, so we can rewrite this as

$$\int_0^{2\pi} 2\sin(t/2)\, dt = 8.$$

□

Exercises 10.5.

1. Consider the curve of exercise 6 in section 10.4. Find all values of t for which the curve has a horizontal tangent line. ⇒
2. Consider the curve of exercise 6 in section 10.4. Find the area under one arch of the curve. ⇒
3. Consider the curve of exercise 6 in section 10.4. Set up an integral for the length of one arch of the curve. ⇒
4. Consider the hypercycloid of exercise 7 in section 10.4. Find all points at which the curve has a horizontal tangent line. ⇒
5. Consider the hypercycloid of exercise 7 in section 10.4. Find the area between the large circle and one arch of the curve. ⇒
6. Consider the hypercycloid of exercise 7 in section 10.4. Find the length of one arch of the curve. ⇒
7. Consider the hypocycloid of exercise 8 in section 10.4. Find the area inside the curve. ⇒
8. Consider the hypocycloid of exercise 8 in section 10.4. Find the length of one arch of the curve. ⇒
9. Recall the involute of a circle from exercise 9 in section 10.4. Find the point in the first quadrant in figure 10.4.4 at which the tangent line is vertical. ⇒
10. Recall the involute of a circle from exercise 9 in section 10.4. Instead of an infinite string, suppose we have a string of length π attached to the unit circle at $(-1, 0)$, and initially laid around the top of the circle with its end at $(1, 0)$. If we grasp the end of the string and begin to unwind it, we get a piece of the involute, until the string is vertical. If we then keep the string taut and continue to rotate it counter-clockwise, the end traces out a semi-circle with center at $(-1, 0)$, until the string is vertical again. Continuing, the end of the string traces out the mirror image of the initial portion of the curve; see figure 10.5.1. Find the area of the region inside this curve and outside the unit circle. ⇒

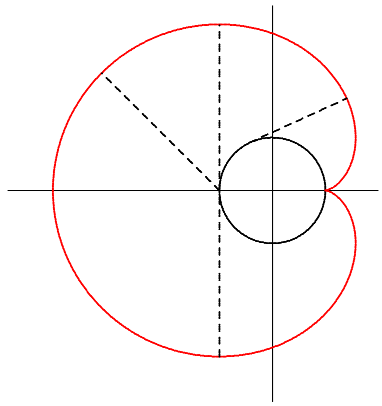


Figure 10.5.1 A region formed by the end of a string.

11. Find the length of the curve from the previous exercise, shown in figure 10.5.1. ⇒

12. Find the length of the spiral of Archimedes (figure 10.3.4) for $0 \le \theta \le 2\pi$. ⇒

11

Sequences and Series

Consider the following sum:

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\cdots+\frac{1}{2^i}+\cdots$$

The dots at the end indicate that the sum goes on forever. Does this make sense? Can we assign a numerical value to an infinite sum? While at first it may seem difficult or impossible, we have certainly done something similar when we talked about one quantity getting "closer and closer" to a fixed quantity. Here we could ask whether, as we add more and more terms, the sum gets closer and closer to some fixed value. That is, look at

$$\begin{aligned}\frac{1}{2}&=\frac{1}{2}\\ \frac{3}{4}&=\frac{1}{2}+\frac{1}{4}\\ \frac{7}{8}&=\frac{1}{2}+\frac{1}{4}+\frac{1}{8}\\ \frac{15}{16}&=\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}\end{aligned}$$

and so on, and ask whether these values have a limit. It seems pretty clear that they do, namely 1. In fact, as we will see, it's not hard to show that

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\cdots+\frac{1}{2^i}=\frac{2^i-1}{2^i}=1-\frac{1}{2^i}$$

and then

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

There is one place that you have long accepted this notion of infinite sum without really thinking of it as a sum:

$$0.3333\bar{3} = \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \frac{3}{10000} + \cdots = \frac{1}{3},$$

for example, or

$$3.14159\ldots = 3 + \frac{1}{10} + \frac{4}{100} + \frac{1}{1000} + \frac{5}{10000} + \frac{9}{100000} + \cdots = \pi.$$

Our first task, then, to investigate infinite sums, called **series**, is to investigate limits of **sequences** of numbers. That is, we officially call

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

a series, while

$$\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}, \ldots, \frac{2^i - 1}{2^i}, \ldots$$

is a sequence, and

$$\sum_{i=1}^{\infty} \frac{1}{2^i} = \lim_{i\to\infty} \frac{2^i - 1}{2^i},$$

that is, the value of a series is the limit of a particular sequence.

11.1 Sequences

While the idea of a sequence of numbers, $a_1, a_2, a_3, \ldots$ is straightforward, it is useful to think of a sequence as a function. We have up until now dealt with functions whose domains are the real numbers, or a subset of the real numbers, like $f(x) = \sin x$. A sequence is a function with domain the natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$ or the non-negative integers, $\mathbb{Z}^{\geq 0} = \{0, 1, 2, 3, \ldots\}$. The range of the function is still allowed to be the real numbers; in symbols, we say that a sequence is a function $f\colon \mathbb{N} \to \mathbb{R}$. Sequences are written in a few different ways, all equivalent; these all mean the same thing:

$$a_1, a_2, a_3, \ldots$$
$$\{a_n\}_{n=1}^{\infty}$$
$$\{f(n)\}_{n=1}^{\infty}$$

As with functions on the real numbers, we will most often encounter sequences that can be expressed by a formula. We have already seen the sequence $a_i = f(i) = 1 - 1/2^i$, and others are easy to come by:

$$\begin{aligned}
f(i) &= \frac{i}{i+1} \\
f(n) &= \frac{1}{2^n} \\
f(n) &= \sin(n\pi/6) \\
f(i) &= \frac{(i-1)(i+2)}{2^i}
\end{aligned}$$

Frequently these formulas will make sense if thought of either as functions with domain $\mathbb{R}$ or $\mathbb{N}$, though occasionally one will make sense only for integer values.

Faced with a sequence we are interested in the limit

$$\lim_{i\to\infty} f(i) = \lim_{i\to\infty} a_i.$$

We already understand

$$\lim_{x\to\infty} f(x)$$

when x is a real valued variable; now we simply want to restrict the "input" values to be integers. No real difference is required in the definition of limit, except that we specify, perhaps implicitly, that the variable is an integer. Compare this definition to definition 4.10.4.

DEFINITION 11.1.1 Suppose that $\{a_n\}_{n=1}^{\infty}$ is a sequence. We say that $\lim_{n\to\infty} a_n = L$ if for every $\epsilon > 0$ there is an $N > 0$ so that whenever $n > N$, $|a_n - L| < \epsilon$. If $\lim_{n\to\infty} a_n = L$ we say that the sequence **converges**, otherwise it **diverges**. □

If $f(i)$ defines a sequence, and $f(x)$ makes sense, and $\lim_{x\to\infty} f(x) = L$, then it is clear that $\lim_{i\to\infty} f(i) = L$ as well, but it is important to note that the converse of this statement is not true. For example, since $\lim_{x\to\infty}(1/x) = 0$, it is clear that also $\lim_{i\to\infty}(1/i) = 0$, that is, the numbers

$$\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \ldots$$

get closer and closer to 0. Consider this, however: Let $f(n) = \sin(n\pi)$. This is the sequence

$$\sin(0\pi), \sin(1\pi), \sin(2\pi), \sin(3\pi), \ldots = 0, 0, 0, 0, \ldots$$

since $\sin(n\pi) = 0$ when n is an integer. Thus $\lim_{n\to\infty} f(n) = 0$. But $\lim_{x\to\infty} f(x)$, when x is real, does not exist: as x gets bigger and bigger, the values $\sin(x\pi)$ do not get closer and closer to a single value, but take on all values between −1 and 1 over and over. In general, whenever you want to know $\lim_{n\to\infty} f(n)$ you should first attempt to compute $\lim_{x\to\infty} f(x)$, since if the latter exists it is also equal to the first limit. But if for some reason $\lim_{x\to\infty} f(x)$ does not exist, it may still be true that $\lim_{n\to\infty} f(n)$ exists, but you'll have to figure out another way to compute it.

It is occasionally useful to think of the graph of a sequence. Since the function is defined only for integer values, the graph is just a sequence of dots. In figure 11.1.1 we see the graphs of two sequences and the graphs of the corresponding real functions.

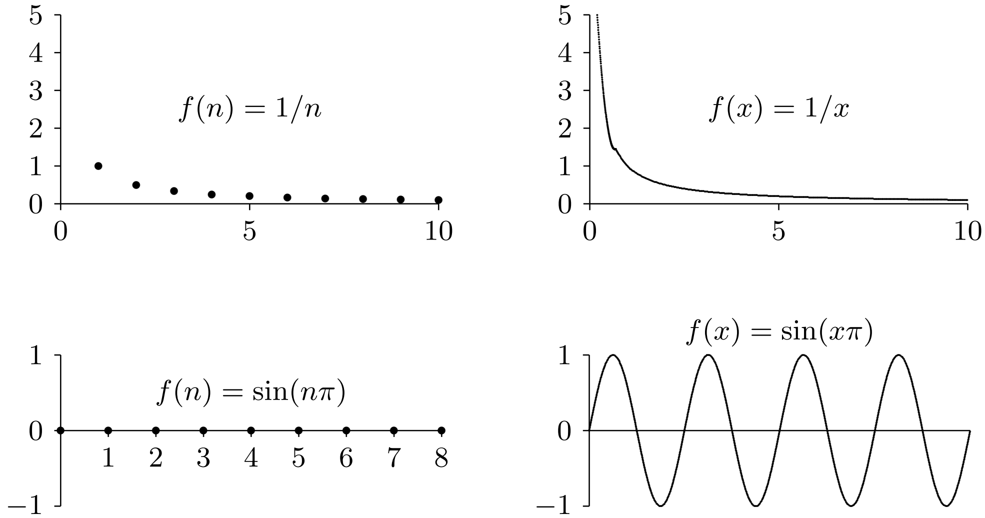


Figure 11.1.1 Graphs of sequences and their corresponding real functions.

Not surprisingly, the properties of limits of real functions translate into properties of sequences quite easily. Theorem 2.3.6 about limits becomes

THEOREM 11.1.2 Suppose that $\lim_{n\to\infty} a_n = L$ and $\lim_{n\to\infty} b_n = M$ and k is some constant. Then

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

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

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

$$\lim_{n\to\infty} (a_n b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n = LM$$

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n} = \frac{L}{M}, \text{ if } M \text{ is not } 0$$

■

Likewise the Squeeze Theorem (4.3.1) becomes

THEOREM 11.1.3 Suppose that $a_n \le b_n \le c_n$ for all $n > N$, for some N. If $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = L$, then $\lim_{n\to\infty} b_n = L$. ■

And a final useful fact:

THEOREM 11.1.4 $\lim_{n\to\infty} |a_n| = 0$ if and only if $\lim_{n\to\infty} a_n = 0$. ■

This says simply that the size of a_n gets close to zero if and only if a_n gets close to zero.

EXAMPLE 11.1.5 Determine whether $\left\{\dfrac{n}{n+1}\right\}_{n=0}^{\infty}$ converges or diverges. If it converges, compute the limit. Since this makes sense for real numbers we consider

$$\lim_{x\to\infty} \frac{x}{x+1} = \lim_{x\to\infty} 1 - \frac{1}{x+1} = 1 - 0 = 1.$$

Thus the sequence converges to 1. □

EXAMPLE 11.1.6 Determine whether $\left\{\dfrac{\ln n}{n}\right\}_{n=1}^{\infty}$ converges or diverges. If it converges, compute the limit. We compute

$$\lim_{x\to\infty} \frac{\ln x}{x} = \lim_{x\to\infty} \frac{1/x}{1} = 0,$$

using L'Hôpital's Rule. Thus the sequence converges to 0. □

EXAMPLE 11.1.7 Determine whether $\{(-1)^n\}_{n=0}^{\infty}$ converges or diverges. If it converges, compute the limit. This does not make sense for all real exponents, but the sequence is easy to understand: it is

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

and clearly diverges. □

EXAMPLE 11.1.8 Determine whether $\{(-1/2)^n\}_{n=0}^{\infty}$ converges or diverges. If it converges, compute the limit. We consider the sequence $\{|(-1/2)^n|\}_{n=0}^{\infty} = \{(1/2)^n\}_{n=0}^{\infty}$. Then

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

so by theorem 11.1.4 the sequence converges to 0. □

EXAMPLE 11.1.9 Determine whether $\{(\sin n)/\sqrt{n}\}_{n=1}^{\infty}$ converges or diverges. If it converges, compute the limit. Since $|\sin n| \le 1$, $0 \le |\sin n/\sqrt{n}| \le 1/\sqrt{n}$ and we can use theorem 11.1.3 with $a_n = 0$ and $c_n = 1/\sqrt{n}$. Since $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = 0$, $\lim_{n\to\infty} \sin n/\sqrt{n} = 0$ and the sequence converges to 0. □

EXAMPLE 11.1.10 A particularly common and useful sequence is $\{r^n\}_{n=0}^{\infty}$, for various values of r. Some are quite easy to understand: If $r = 1$ the sequence converges to 1 since every term is 1, and likewise if $r = 0$ the sequence converges to 0. If $r = -1$ this is the sequence of example 11.1.7 and diverges. If $r > 1$ or $r < -1$ the terms r^n get large without limit, so the sequence diverges. If $0 < r < 1$ then the sequence converges to 0. If $-1 < r < 0$ then $|r^n| = |r|^n$ and $0 < |r| < 1$, so the sequence $\{|r|^n\}_{n=0}^{\infty}$ converges to 0, so also $\{r^n\}_{n=0}^{\infty}$ converges to 0. converges. In summary, $\{r^n\}$ converges precisely when $-1 < r \le 1$ in which case

$$\lim_{n\to\infty} r^n = \begin{cases} 0 & \text{if } -1 < r < 1 \\ 1 & \text{if } r = 1 \end{cases}$$

□

Sometimes we will not be able to determine the limit of a sequence, but we still would like to know whether it converges. In some cases we can determine this even without being able to compute the limit.

A sequence is called **increasing** or sometimes **strictly increasing** if $a_i < a_{i+1}$ for all i. It is called **non-decreasing** or sometimes (unfortunately) **increasing** if $a_i \le a_{i+1}$ for all i. Similarly a sequence is **decreasing** if $a_i > a_{i+1}$ for all i and **non-increasing** if $a_i \ge a_{i+1}$ for all i. If a sequence has any of these properties it is called **monotonic**.

EXAMPLE 11.1.11 The sequence

$$\left\{\frac{2^i - 1}{2^i}\right\}_{i=1}^{\infty} = \frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}, \ldots,$$

is increasing, and

$$\left\{\frac{n+1}{n}\right\}_{i=1}^{\infty} = \frac{2}{1}, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \ldots$$

is decreasing. □

A sequence is **bounded above** if there is some number N such that $a_n \le N$ for every n, and **bounded below** if there is some number N such that $a_n \ge N$ for every n. If a sequence is bounded above and bounded below it is **bounded**. If a sequence $\{a_n\}_{n=0}^{\infty}$ is increasing or non-decreasing it is bounded below (by a_0), and if it is decreasing or non-increasing it is bounded above (by a_0). Finally, with all this new terminology we can state an important theorem.

THEOREM 11.1.12 If a sequence is bounded and monotonic then it converges. ■

We will not prove this; the proof appears in many calculus books. It is not hard to believe: suppose that a sequence is increasing and bounded, so each term is larger than the one before, yet never larger than some fixed value N. The terms must then get closer and closer to some value between a_0 and N. It need not be N, since N may be a "too-generous" upper bound; the limit will be the smallest number that is above all of the terms a_i.

EXAMPLE 11.1.13 All of the terms $(2^i - 1)/2^i$ are less than 2, and the sequence is increasing. As we have seen, the limit of the sequence is 1—1 is the smallest number that is bigger than all the terms in the sequence. Similarly, all of the terms $(n+1)/n$ are bigger than 1/2, and the limit is 1—1 is the largest number that is smaller than the terms of the sequence. □

We don't actually need to know that a sequence is monotonic to apply this theorem—it is enough to know that the sequence is "eventually" monotonic, that is, that at some point it becomes increasing or decreasing. For example, the sequence 10, 9, 8, 15, 3, 21, 4, 3/4, 7/8, 15/16, 31/32, ... is not increasing, because among the first few terms it is not. But starting with the term 3/4 it is increasing, so the theorem tells us that the sequence $3/4, 7/8, 15/16, 31/32, \ldots$ converges. Since convergence depends only on what happens as n gets large, adding a few terms at the beginning can't turn a convergent sequence into a divergent one.

EXAMPLE 11.1.14 Show that $\{n^{1/n}\}$ converges.

We first show that this sequence is decreasing, that is, that $n^{1/n} > (n+1)^{1/(n+1)}$. Consider the real function $f(x) = x^{1/x}$ when $x \geq 1$. We can compute the derivative, $f'(x) = x^{1/x}(1-\ln x)/x^2$, and note that when $x \geq 3$ this is negative. Since the function has negative slope, $n^{1/n} > (n+1)^{1/(n+1)}$ when $n \geq 3$. Since all terms of the sequence are positive, the sequence is decreasing and bounded when $n \geq 3$, and so the sequence converges. (As it happens, we can compute the limit in this case, but we know it converges even without knowing the limit; see exercise 1.) □

EXAMPLE 11.1.15 Show that $\{n!/n^n\}$ converges.

Again we show that the sequence is decreasing, and since each term is positive the sequence converges. We can't take the derivative this time, as $x!$ doesn't make sense for x real. But we note that if $a_{n+1}/a_n < 1$ then $a_{n+1} < a_n$, which is what we want to know. So we look at a_{n+1}/a_n:

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)!}{(n+1)^{n+1}}\frac{n^n}{n!} = \frac{(n+1)!}{n!}\frac{n^n}{(n+1)^{n+1}} = \frac{n+1}{n+1}\left(\frac{n}{n+1}\right)^n = \left(\frac{n}{n+1}\right)^n < 1.$$

(Again it is possible to compute the limit; see exercise 2.) □

Exercises 11.1.

1. Compute $\lim_{x\to\infty} x^{1/x}$. ⇒
2. Use the squeeze theorem to show that $\lim_{n\to\infty} \frac{n!}{n^n} = 0$.
3. Determine whether $\{\sqrt{n+47} - \sqrt{n}\}_{n=0}^{\infty}$ converges or diverges. If it converges, compute the limit. ⇒
4. Determine whether $\left\{\frac{n^2+1}{(n+1)^2}\right\}_{n=0}^{\infty}$ converges or diverges. If it converges, compute the limit. ⇒
5. Determine whether $\left\{\frac{n+47}{\sqrt{n^2+3n}}\right\}_{n=1}^{\infty}$ converges or diverges. If it converges, compute the limit. ⇒
6. Determine whether $\left\{\frac{2^n}{n!}\right\}_{n=0}^{\infty}$ converges or diverges. ⇒

11.2 Series

While much more can be said about sequences, we now turn to our principal interest, series. Recall that a series, roughly speaking, is the sum of a sequence: if $\{a_n\}_{n=0}^{\infty}$ is a sequence then the associated series is

$$\sum_{i=0}^{\infty} a_n = a_0 + a_1 + a_2 + \cdots$$

Associated with a series is a second sequence, called the **sequence of partial sums** $\{s_n\}_{n=0}^{\infty}$:

$$s_n = \sum_{i=0}^{n} a_i.$$

So

$$s_0 = a_0, \quad s_1 = a_0 + a_1, \quad s_2 = a_0 + a_1 + a_2, \quad \ldots$$

A series converges if the sequence of partial sums converges, and otherwise the series diverges.

EXAMPLE 11.2.1 If $a_n = kx^n$, $\sum_{n=0}^{\infty} a_n$ is called a **geometric series**. A typical partial sum is

$$s_n = k + kx + kx^2 + kx^3 + \cdots + kx^n = k(1 + x + x^2 + x^3 + \cdots + x^n).$$

We note that

$$\begin{aligned} s_n(1-x) &= k(1+x+x^2+x^3+\cdots+x^n)(1-x) \\ &= k(1+x+x^2+x^3+\cdots+x^n)1 - k(1+x+x^2+x^3+\cdots+x^{n-1}+x^n)x \\ &= k(1+x+x^2+x^3+\cdots+x^n - x - x^2 - x^3 - \cdots - x^n - x^{n+1}) \\ &= k(1-x^{n+1}) \end{aligned}$$

so

$$\begin{aligned} s_n(1-x) &= k(1-x^{n+1}) \\ s_n &= k\frac{1-x^{n+1}}{1-x}. \end{aligned}$$

If $|x| < 1$, $\lim_{n\to\infty} x^n = 0$ so

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} k\frac{1-x^{n+1}}{1-x} = k\frac{1}{1-x}.$$

Thus, when $|x| < 1$ the geometric series converges to $k/(1-x)$. When, for example, $k = 1$ and $x = 1/2$:

$$s_n = \frac{1-(1/2)^{n+1}}{1-1/2} = \frac{2^{n+1}-1}{2^n} = 2 - \frac{1}{2^n} \quad \text{and} \quad \sum_{n=0}^{\infty} \frac{1}{2^n} = \frac{1}{1-1/2} = 2.$$

We began the chapter with the series

$$\sum_{n=1}^{\infty} \frac{1}{2^n},$$

namely, the geometric series without the first term 1. Each partial sum of this series is 1 less than the corresponding partial sum for the geometric series, so of course the limit is also one less than the value of the geometric series, that is,

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = 1.$$

□

It is not hard to see that the following theorem follows from theorem 11.1.2.

THEOREM 11.2.2 Suppose that $\sum a_n$ and $\sum b_n$ are convergent series, and c is a constant. Then

1. $\sum ca_n$ is convergent and $\sum ca_n = c\sum a_n$

2. $\sum(a_n+b_n)$ is convergent and $\sum(a_n+b_n)=\sum a_n+\sum b_n$.

∎

The two parts of this theorem are subtly different. Suppose that $\sum a_n$ diverges; does $\sum ca_n$ also diverge if c is non-zero? Yes: suppose instead that $\sum ca_n$ converges; then by the theorem, $\sum(1/c)ca_n$ converges, but this is the same as $\sum a_n$, which by assumption diverges. Hence $\sum ca_n$ also diverges. Note that we are applying the theorem with a_n replaced by ca_n and c replaced by $(1/c)$.

Now suppose that $\sum a_n$ and $\sum b_n$ diverge; does $\sum(a_n+b_n)$ also diverge? Now the answer is no: Let $a_n=1$ and $b_n=-1$, so certainly $\sum a_n$ and $\sum b_n$ diverge. But $\sum(a_n+b_n)=\sum(1+-1)=\sum 0=0$. Of course, sometimes $\sum(a_n+b_n)$ will also diverge, for example, if $a_n=b_n=1$, then $\sum(a_n+b_n)=\sum(1+1)=\sum 2$ diverges.

In general, the sequence of partial sums s_n is harder to understand and analyze than the sequence of terms a_n, and it is difficult to determine whether series converge and if so to what. Sometimes things are relatively simple, starting with the following.

THEOREM 11.2.3 If $\sum a_n$ converges then $\lim\limits_{n\to\infty} a_n=0$.

Proof. Since $\sum a_n$ converges, $\lim\limits_{n\to\infty} s_n=L$ and $\lim\limits_{n\to\infty} s_{n-1}=L$, because this really says the same thing but "renumbers" the terms. By theorem 11.1.2,

$$\lim_{n\to\infty}(s_n-s_{n-1})=\lim_{n\to\infty} s_n-\lim_{n\to\infty} s_{n-1}=L-L=0.$$

But

$$s_n-s_{n-1}=(a_0+a_1+a_2+\cdots+a_n)-(a_0+a_1+a_2+\cdots+a_{n-1})=a_n,$$

so as desired $\lim\limits_{n\to\infty} a_n=0$. ∎

This theorem presents an easy divergence test: if given a series $\sum a_n$ the limit $\lim\limits_{n\to\infty} a_n$ does not exist or has a value other than zero, the series diverges. Note well that the converse is *not* true: If $\lim\limits_{n\to\infty} a_n=0$ then the series does not necessarily converge.

EXAMPLE 11.2.4 Show that $\displaystyle\sum_{n=1}^{\infty}\frac{n}{n+1}$ diverges.

We compute the limit:

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

Looking at the first few terms perhaps makes it clear that the series has no chance of converging:

$$\frac{1}{2}+\frac{2}{3}+\frac{3}{4}+\frac{4}{5}+\cdots$$

will just get larger and larger; indeed, after a bit longer the series starts to look very much like $\cdots + 1 + 1 + 1 + 1 + \cdots$, and of course if we add up enough 1's we can make the sum as large as we desire. □

EXAMPLE 11.2.5 Show that $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.

Here the theorem does not apply: $\lim_{n\to\infty} 1/n = 0$, so it looks like perhaps the series converges. Indeed, if you have the fortitude (or the software) to add up the first 1000 terms you will find that

$$\sum_{n=1}^{1000} \frac{1}{n} \approx 7.49,$$

so it might be reasonable to speculate that the series converges to something in the neighborhood of 10. But in fact the partial sums do go to infinity; they just get big very, very slowly. Consider the following:

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} > 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} = 1 + \frac{1}{2} + \frac{1}{2}$$

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}$$

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{16} > 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{8} + \frac{1}{16} + \cdots + \frac{1}{16} = 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}$$

and so on. By swallowing up more and more terms we can always manage to add at least another 1/2 to the sum, and by adding enough of these we can make the partial sums as big as we like. In fact, it's not hard to see from this pattern that

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{2^n} > 1 + \frac{n}{2},$$

so to make sure the sum is over 100, for example, we'd add up terms until we get to around $1/2^{198}$, that is, about $4 \cdot 10^{59}$ terms. This series, $\sum(1/n)$, is called the **harmonic series**. □

Exercises 11.2.

1. Explain why $\displaystyle\sum_{n=1}^{\infty} \frac{n^2}{2n^2+1}$ diverges. ⇒

2. Explain why $\displaystyle\sum_{n=1}^{\infty} \frac{5}{2^{1/n}+14}$ diverges. ⇒

3. Explain why $\displaystyle\sum_{n=1}^{\infty} \frac{3}{n}$ diverges. ⇒

4. Compute $\displaystyle\sum_{n=0}^{\infty} \frac{4}{(-3)^n} - \frac{3}{3^n}$. ⇒

5. Compute $\displaystyle\sum_{n=0}^{\infty} \frac{3}{2^n} + \frac{4}{5^n}$. ⇒

6. Compute $\displaystyle\sum_{n=0}^{\infty} \frac{4^{n+1}}{5^n}$. ⇒

7. Compute $\displaystyle\sum_{n=0}^{\infty} \frac{3^{n+1}}{7^{n+1}}$. ⇒

8. Compute $\displaystyle\sum_{n=1}^{\infty} \left(\frac{3}{5}\right)^n$. ⇒

9. Compute $\displaystyle\sum_{n=1}^{\infty} \frac{3^n}{5^{n+1}}$. ⇒

11.3 The Integral Test

It is generally quite difficult, often impossible, to determine the value of a series exactly. In many cases it is possible at least to determine whether or not the series converges, and so we will spend most of our time on this problem.

If all of the terms a_n in a series are non-negative, then clearly the sequence of partial sums s_n is non-decreasing. This means that if we can show that the sequence of partial sums is bounded, the series must converge. We know that if the series converges, the terms a_n approach zero, but this does not mean that $a_n \ge a_{n+1}$ for every n. Many useful and interesting series do have this property, however, and they are among the easiest to understand. Let's look at an example.

EXAMPLE 11.3.1 Show that $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges.

The terms $1/n^2$ are positive and decreasing, and since $\displaystyle\lim_{x\to\infty} 1/x^2 = 0$, the terms $1/n^2$ approach zero. We seek an upper bound for all the partial sums, that is, we want to find a number N so that $s_n \le N$ for every n. The upper bound is provided courtesy of integration, and is inherent in figure 11.3.1.

The figure shows the graph of $y = 1/x^2$ together with some rectangles that lie completely below the curve and that all have base length one. Because the heights of the rectangles are determined by the height of the curve, the areas of the rectangles are $1/1^2$, $1/2^2$, $1/3^2$, and so on—in other words, exactly the terms of the series. The partial sum s_n is simply the sum of the areas of the first n rectangles. Because the rectangles all lie between the curve and the x-axis, any sum of rectangle areas is less than the corresponding

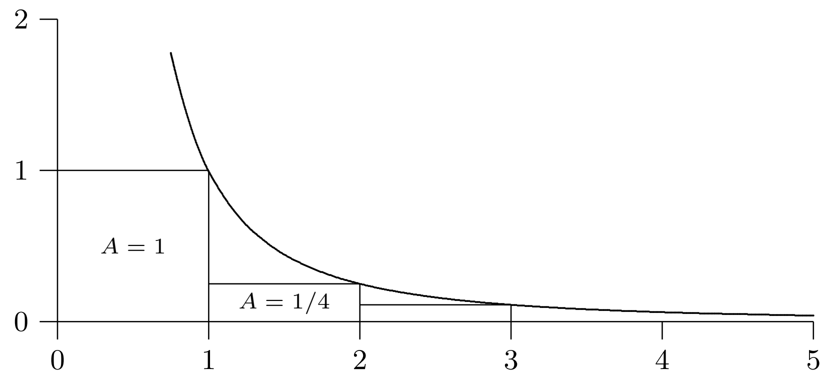


Figure 11.3.1 Graph of $y = 1/x^2$ with rectangles.

area under the curve, and so of course any sum of rectangle areas is less than the area under the entire curve, that is, all the way to infinity. There is a bit of trouble at the left end, where there is an asymptote, but we can work around that easily. Here it is:

$$s_n = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} < 1 + \int_1^n \frac{1}{x^2}\,dx < 1 + \int_1^\infty \frac{1}{x^2}\,dx = 1 + 1 = 2,$$

recalling that we computed this improper integral in section 9.7. Since the sequence of partial sums s_n is increasing and bounded above by 2, we know that $\lim_{n\to\infty} s_n = L < 2$, and so the series converges to some number less than 2. In fact, it is possible, though difficult, to show that $L = \pi^2/6 \approx 1.6$. □

We already know that $\sum 1/n$ diverges. What goes wrong if we try to apply this technique to it? Here's the calculation:

$$s_n = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} < 1 + \int_1^n \frac{1}{x}\,dx < 1 + \int_1^\infty \frac{1}{x}\,dx = 1 + \infty.$$

The problem is that the improper integral doesn't converge. Note well that this does *not* prove that $\sum 1/n$ diverges, just that this particular calculation fails to prove that it converges. A slight modification, however, allows us to prove in a second way that $\sum 1/n$ diverges.

EXAMPLE 11.3.2 Consider a slightly altered version of figure 11.3.1, shown in figure 11.3.2.

The rectangles this time are above the curve, that is, each rectangle completely contains the corresponding area under the curve. This means that

$$s_n = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} > \int_1^{n+1} \frac{1}{x}\,dx = \ln x\Big|_1^{n+1} = \ln(n+1).$$

As n gets bigger, $\ln(n+1)$ goes to infinity, so the sequence of partial sums s_n must also go to infinity, so the harmonic series diverges. □

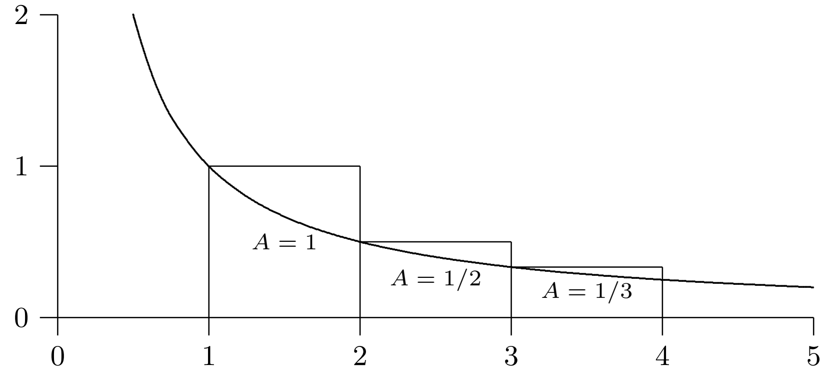


Figure 11.3.2 Graph of $y = 1/x$ with rectangles.

The important fact that clinches this example is that

$$\lim_{n\to\infty} \int_1^{n+1} \frac{1}{x}\,dx = \infty,$$

which we can rewrite as

$$\int_1^{\infty} \frac{1}{x}\,dx = \infty.$$

So these two examples taken together indicate that we can prove that a series converges or prove that it diverges with a single calculation of an improper integral. This is known as the **integral test**, which we state as a theorem.

THEOREM 11.3.3 Suppose that $f(x) > 0$ and is decreasing on the infinite interval $[k, \infty)$ (for some $k \ge 1$) and that $a_n = f(n)$. Then the series $\displaystyle\sum_{n=1}^{\infty} a_n$ converges if and only if the improper integral $\displaystyle\int_1^{\infty} f(x)\,dx$ converges. ■

The two examples we have seen are called p-series; a p-series is any series of the form $\sum 1/n^p$. If $p \le 0$, $\lim_{n\to\infty} 1/n^p \ne 0$, so the series diverges. For positive values of p we can determine precisely which series converge.

THEOREM 11.3.4 A p-series with $p > 0$ converges if and only if $p > 1$.

Proof. We use the integral test; we have already done $p = 1$, so assume that $p \ne 1$.

$$\int_1^{\infty} \frac{1}{x^p}\,dx = \lim_{D\to\infty} \left.\frac{x^{1-p}}{1-p}\right|_1^D = \lim_{D\to\infty} \frac{D^{1-p}}{1-p} - \frac{1}{1-p}.$$

If $p > 1$ then $1 - p < 0$ and $\lim_{D\to\infty} D^{1-p} = 0$, so the integral converges. If $0 < p < 1$ then $1 - p > 0$ and $\lim_{D\to\infty} D^{1-p} = \infty$, so the integral diverges. ■

EXAMPLE 11.3.5 Show that $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^3}$ converges.

We could of course use the integral test, but now that we have the theorem we may simply note that this is a p-series with $p > 1$. □

EXAMPLE 11.3.6 Show that $\displaystyle\sum_{n=1}^{\infty} \frac{5}{n^4}$ converges.

We know that if $\displaystyle\sum_{n=1}^{\infty} 1/n^4$ converges then $\displaystyle\sum_{n=1}^{\infty} 5/n^4$ also converges, by theorem 11.2.2. Since $\displaystyle\sum_{n=1}^{\infty} 1/n^4$ is a convergent p-series, $\displaystyle\sum_{n=1}^{\infty} 5/n^4$ converges also. □

EXAMPLE 11.3.7 Show that $\displaystyle\sum_{n=1}^{\infty} \frac{5}{\sqrt{n}}$ diverges.

This also follows from theorem 11.2.2: Since $\displaystyle\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$ is a p-series with $p = 1/2 < 1$, it diverges, and so does $\displaystyle\sum_{n=1}^{\infty} \frac{5}{\sqrt{n}}$. □

Since it is typically difficult to compute the value of a series exactly, a good approximation is frequently required. In a real sense, a good approximation is only as good as we know it is, that is, while an approximation may in fact be good, it is only valuable in practice if we can guarantee its accuracy to some degree. This guarantee is usually easy to come by for series with decreasing positive terms.

EXAMPLE 11.3.8 Approximate $\sum 1/n^2$ to two decimal places.

Referring to figure 11.3.1, if we approximate the sum by $\displaystyle\sum_{n=1}^{N} 1/n^2$, the error we make is the total area of the remaining rectangles, all of which lie under the curve $1/x^2$ from $x = N$ out to infinity. So we know the true value of the series is larger than the approximation, and no bigger than the approximation plus the area under the curve from N to infinity. Roughly, then, we need to find N so that

$$\int_N^{\infty} \frac{1}{x^2}\,dx < 1/100.$$

We can compute the integral:

$$\int_N^\infty \frac{1}{x^2}\,dx = \frac{1}{N},$$

so $N = 100$ is a good starting point. Adding up the first 100 terms gives approximately 1.634983900, and that plus $1/100$ is 1.644983900, so approximating the series by the value halfway between these will be at most $1/200 = 0.005$ in error. The midpoint is 1.639983900, but while this is correct to ± 0.005, we can't tell if the correct two-decimal approximation is 1.63 or 1.64. We need to make N big enough to reduce the guaranteed error, perhaps to around 0.004 to be safe, so we would need $1/N \approx 0.008$, or $N = 125$. Now the sum of the first 125 terms is approximately 1.636965982, and that plus 0.008 is 1.644965982 and the point halfway between them is 1.640965982. The true value is then 1.640965982 ± 0.004, and all numbers in this range round to 1.64, so 1.64 is correct to two decimal places. We have mentioned that the true value of this series can be shown to be $\pi^2/6 \approx 1.644934068$ which rounds down to 1.64 (just barely) and is indeed below the upper bound of 1.644965982, again just barely. Frequently approximations will be even better than the "guaranteed" accuracy, but not always, as this example demonstrates. □

Exercises 11.3.

Determine whether each series converges or diverges.

1. $\displaystyle\sum_{n=1}^\infty \frac{1}{n^{\pi/4}}$ ⇒

2. $\displaystyle\sum_{n=1}^\infty \frac{n}{n^2+1}$ ⇒

3. $\displaystyle\sum_{n=1}^\infty \frac{\ln n}{n^2}$ ⇒

4. $\displaystyle\sum_{n=1}^\infty \frac{1}{n^2+1}$ ⇒

5. $\displaystyle\sum_{n=1}^\infty \frac{1}{e^n}$ ⇒

6. $\displaystyle\sum_{n=1}^\infty \frac{n}{e^n}$ ⇒

7. $\displaystyle\sum_{n=2}^\infty \frac{1}{n\ln n}$ ⇒

8. $\displaystyle\sum_{n=2}^\infty \frac{1}{n(\ln n)^2}$ ⇒

9. Find an N so that $\displaystyle\sum_{n=1}^\infty \frac{1}{n^4}$ is between $\displaystyle\sum_{n=1}^N \frac{1}{n^4}$ and $\displaystyle\sum_{n=1}^N \frac{1}{n^4} + 0.005$. ⇒

10. Find an N so that $\displaystyle\sum_{n=0}^\infty \frac{1}{e^n}$ is between $\displaystyle\sum_{n=0}^N \frac{1}{e^n}$ and $\displaystyle\sum_{n=0}^N \frac{1}{e^n} + 10^{-4}$. ⇒

11. Find an N so that $\displaystyle\sum_{n=1}^\infty \frac{\ln n}{n^2}$ is between $\displaystyle\sum_{n=1}^N \frac{\ln n}{n^2}$ and $\displaystyle\sum_{n=1}^N \frac{\ln n}{n^2} + 0.005$. ⇒

12. Find an N so that $\displaystyle\sum_{n=2}^\infty \frac{1}{n(\ln n)^2}$ is between $\displaystyle\sum_{n=2}^N \frac{1}{n(\ln n)^2}$ and $\displaystyle\sum_{n=2}^N \frac{1}{n(\ln n)^2} + 0.005$. ⇒

11.4 Alternating Series

Next we consider series with both positive and negative terms, but in a regular pattern: they alternate, as in the **alternating harmonic series** for example:

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = \frac{1}{1} + \frac{-1}{2} + \frac{1}{3} + \frac{-1}{4} + \cdots = \frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots.$$

In this series the sizes of the terms decrease, that is, $|a_n|$ forms a decreasing sequence, but this is not required in an alternating series. As with positive term series, however, when the terms do have decreasing sizes it is easier to analyze the series, much easier, in fact, than positive term series. Consider pictorially what is going on in the alternating harmonic series, shown in figure 11.4.1. Because the sizes of the terms a_n are decreasing, the partial sums s_1, s_3, s_5, and so on, form a decreasing sequence that is bounded below by s_2, so this sequence must converge. Likewise, the partial sums s_2, s_4, s_6, and so on, form an increasing sequence that is bounded above by s_1, so this sequence also converges. Since all the even numbered partial sums are less than all the odd numbered ones, and since the "jumps" (that is, the a_i terms) are getting smaller and smaller, the two sequences must converge to the same value, meaning the entire sequence of partial sums $s_1, s_2, s_3, \ldots$ converges as well.

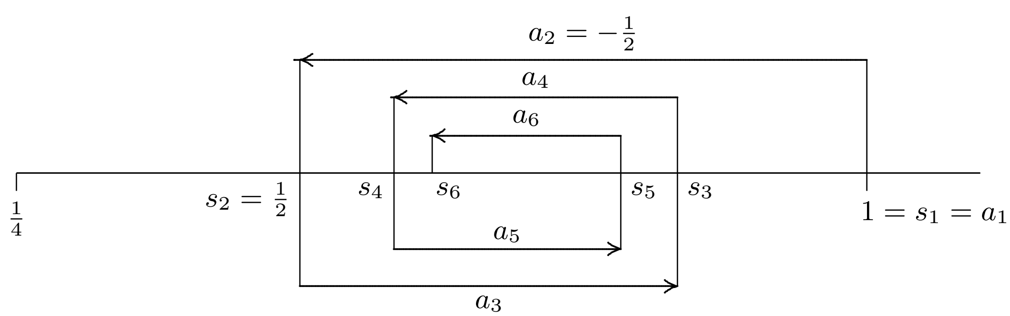


Figure 11.4.1 The alternating harmonic series.

There's nothing special about the alternating harmonic series—the same argument works for any alternating sequence with decreasing size terms. The alternating series test is worth calling a theorem.

THEOREM 11.4.1 Suppose that $\{a_n\}_{n=1}^{\infty}$ is a non-increasing sequence of positive numbers and $\lim_{n\to\infty} a_n = 0$. Then the alternating series $\sum_{n=1}^{\infty}(-1)^{n-1}a_n$ converges.

Proof. The odd numbered partial sums, s_1, s_3, s_5, and so on, form a non-increasing sequence, because $s_{2k+3} = s_{2k+1} - a_{2k+2} + a_{2k+3} \le s_{2k+1}$, since $a_{2k+2} \ge a_{2k+3}$. This sequence is bounded below by s_2, so it must converge, say $\lim_{k\to\infty} s_{2k+1} = L$. Likewise, the partial sums s_2, s_4, s_6, and so on, form a non-decreasing sequence that is bounded above by s_1, so this sequence also converges, say $\lim_{k\to\infty} s_{2k} = M$. Since $\lim_{n\to\infty} a_n = 0$ and $s_{2k+1} = s_{2k} + a_{2k+1}$,

$$L = \lim_{k\to\infty} s_{2k+1} = \lim_{k\to\infty} (s_{2k} + a_{2k+1}) = \lim_{k\to\infty} s_{2k} + \lim_{k\to\infty} a_{2k+1} = M + 0 = M,$$

so $L = M$, the two sequences of partial sums converge to the same limit, and this means the entire sequence of partial sums also converges to L. ■

Another useful fact is implicit in this discussion. Suppose that

$$L = \sum_{n=1}^{\infty} (-1)^{n-1} a_n$$

and that we approximate L by a finite part of this sum, say

$$L \approx \sum_{n=1}^{N} (-1)^{n-1} a_n.$$

Because the terms are decreasing in size, we know that the true value of L must be between this approximation and the next one, that is, between

$$\sum_{n=1}^{N} (-1)^{n-1} a_n \quad \text{and} \quad \sum_{n=1}^{N+1} (-1)^{n-1} a_n.$$

Depending on whether N is odd or even, the second will be larger or smaller than the first.

EXAMPLE 11.4.2 Approximate the alternating harmonic series to one decimal place.

We need to go roughly to the point at which the next term to be added or subtracted is 1/10. Adding up the first nine and the first ten terms we get approximately 0.746 and 0.646. These are 1/10 apart, but it is not clear how the correct value would be rounded. It turns out that we are able to settle the question by computing the sums of the first eleven and twelve terms, which give 0.737 and 0.653, so correct to one place the value is 0.7. □

We have considered alternating series with first index 1, and in which the first term is positive, but a little thought shows this is not crucial. The same test applies to any similar series, such as $\sum_{n=0}^{\infty} (-1)^n a_n$, $\sum_{n=1}^{\infty} (-1)^n a_n$, $\sum_{n=17}^{\infty} (-1)^n a_n$, etc.

Exercises 11.4.

Determine whether the following series converge or diverge.

1. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n+5}$ ⇒

2. $\displaystyle\sum_{n=4}^{\infty} \frac{(-1)^{n-1}}{\sqrt{n-3}}$ ⇒

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

4. $\displaystyle\sum_{n=1}^{\infty} (-1)^{n-1}\frac{\ln n}{n}$ ⇒

5. Approximate $\displaystyle\sum_{n=1}^{\infty} (-1)^{n-1}\frac{1}{n^3}$ to two decimal places. ⇒

6. Approximate $\displaystyle\sum_{n=1}^{\infty} (-1)^{n-1}\frac{1}{n^4}$ to two decimal places. ⇒

11.5 Comparison Tests

As we begin to compile a list of convergent and divergent series, new ones can sometimes be analyzed by comparing them to ones that we already understand.

EXAMPLE 11.5.1 Does $\displaystyle\sum_{n=2}^{\infty} \frac{1}{n^2 \ln n}$ converge?

The obvious first approach, based on what we know, is the integral test. Unfortunately, we can't compute the required antiderivative. But looking at the series, it would appear that it must converge, because the terms we are adding are smaller than the terms of a p-series, that is,

$$\frac{1}{n^2 \ln n} < \frac{1}{n^2},$$

when $n \ge 3$. Since adding up the terms $1/n^2$ doesn't get "too big", the new series "should" also converge. Let's make this more precise.

The series $\displaystyle\sum_{n=2}^{\infty} \frac{1}{n^2 \ln n}$ converges if and only if $\displaystyle\sum_{n=3}^{\infty} \frac{1}{n^2 \ln n}$ converges—all we've done is dropped the initial term. We know that $\displaystyle\sum_{n=3}^{\infty} \frac{1}{n^2}$ converges. Looking at two typical partial sums:

$$s_n = \frac{1}{3^2 \ln 3} + \frac{1}{4^2 \ln 4} + \frac{1}{5^2 \ln 5} + \cdots + \frac{1}{n^2 \ln n} < \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \cdots + \frac{1}{n^2} = t_n.$$

Since the p-series converges, say to L, and since the terms are positive, $t_n < L$. Since the terms of the new series are positive, the s_n form an increasing sequence and $s_n < t_n < L$ for all n. Hence the sequence $\{s_n\}$ is bounded and so converges. □

Sometimes, even when the integral test applies, comparison to a known series is easier, so it's generally a good idea to think about doing a comparison before doing the integral test.

EXAMPLE 11.5.2 Does $\displaystyle\sum_{n=2}^{\infty} \frac{|\sin n|}{n^2}$ converge?

We can't apply the integral test here, because the terms of this series are not decreasing. Just as in the previous example, however,

$$\frac{|\sin n|}{n^2} \le \frac{1}{n^2},$$

because $|\sin n| \le 1$. Once again the partial sums are non-decreasing and bounded above by $\sum 1/n^2 = L$, so the new series converges. □

Like the integral test, the comparison test can be used to show both convergence and divergence. In the case of the integral test, a single calculation will confirm whichever is the case. To use the comparison test we must first have a good idea as to convergence or divergence and pick the sequence for comparison accordingly.

EXAMPLE 11.5.3 Does $\displaystyle\sum_{n=2}^{\infty} \frac{1}{\sqrt{n^2-3}}$ converge?

We observe that the -3 should have little effect compared to the n^2 inside the square root, and therefore guess that the terms are enough like $1/\sqrt{n^2} = 1/n$ that the series should diverge. We attempt to show this by comparison to the harmonic series. We note that

$$\frac{1}{\sqrt{n^2-3}} > \frac{1}{\sqrt{n^2}} = \frac{1}{n},$$

so that

$$s_n = \frac{1}{\sqrt{2^2-3}} + \frac{1}{\sqrt{3^2-3}} + \cdots + \frac{1}{\sqrt{n^2-3}} > \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = t_n,$$

where t_n is 1 less than the corresponding partial sum of the harmonic series (because we start at $n = 2$ instead of $n = 1$). Since $\lim\limits_{n\to\infty} t_n = \infty$, $\lim\limits_{n\to\infty} s_n = \infty$ as well. □

So the general approach is this: If you believe that a new series is convergent, attempt to find a convergent series whose terms are larger than the terms of the new series; if you believe that a new series is divergent, attempt to find a divergent series whose terms are smaller than the terms of the new series.

EXAMPLE 11.5.4 Does $\displaystyle\sum_{n=1}^{\infty} \frac{1}{\sqrt{n^2+3}}$ converge?

Just as in the last example, we guess that this is very much like the harmonic series and so diverges. Unfortunately,

$$\frac{1}{\sqrt{n^2+3}} < \frac{1}{n},$$

so we can't compare the series directly to the harmonic series. A little thought leads us to

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

so if $\sum 1/(2n)$ diverges then the given series diverges. But since $\sum 1/(2n) = (1/2)\sum 1/n$, theorem 11.2.2 implies that it does indeed diverge. □

For reference we summarize the comparison test in a theorem.

THEOREM 11.5.5 Suppose that a_n and b_n are non-negative for all n and that $a_n \le b_n$ when $n \ge N$, for some N.

If $\displaystyle\sum_{n=0}^{\infty} b_n$ converges, so does $\displaystyle\sum_{n=0}^{\infty} a_n$.

If $\displaystyle\sum_{n=0}^{\infty} a_n$ diverges, so does $\displaystyle\sum_{n=0}^{\infty} b_n$.

■

Exercises 11.5.

Determine whether the series converge or diverge.

1. $\displaystyle\sum_{n=1}^{\infty} \frac{1}{2n^2+3n+5}$ ⇒

2. $\displaystyle\sum_{n=2}^{\infty} \frac{1}{2n^2+3n-5}$ ⇒

3. $\displaystyle\sum_{n=1}^{\infty} \frac{1}{2n^2-3n-5}$ ⇒

4. $\displaystyle\sum_{n=1}^{\infty} \frac{3n+4}{2n^2+3n+5}$ ⇒

5. $\displaystyle\sum_{n=1}^{\infty} \frac{3n^2+4}{2n^2+3n+5}$ ⇒

6. $\displaystyle\sum_{n=1}^{\infty} \frac{\ln n}{n}$ ⇒

7. $\displaystyle\sum_{n=1}^{\infty} \frac{\ln n}{n^3}$ ⇒

8. $\displaystyle\sum_{n=2}^{\infty} \frac{1}{\ln n}$ ⇒

9. $\displaystyle\sum_{n=1}^{\infty} \frac{3^n}{2^n+5^n}$ ⇒

10. $\displaystyle\sum_{n=1}^{\infty} \frac{3^n}{2^n+3^n}$ ⇒

11.6 Absolute Convergence

Roughly speaking there are two ways for a series to converge: As in the case of $\sum 1/n^2$, the individual terms get small very quickly, so that the sum of all of them stays finite, or, as in the case of $\sum(-1)^{n-1}/n$, the terms don't get small fast enough ($\sum 1/n$ diverges), but a mixture of positive and negative terms provides enough cancellation to keep the sum finite. You might guess from what we've seen that if the terms get small fast enough to do the job, then whether or not some terms are negative and some positive the series converges.

THEOREM 11.6.1 If $\displaystyle\sum_{n=0}^{\infty} |a_n|$ converges, then $\displaystyle\sum_{n=0}^{\infty} a_n$ converges.

Proof. Note that $0 \le a_n+|a_n| \le 2|a_n|$ so by the comparison test $\displaystyle\sum_{n=0}^{\infty}(a_n+|a_n|)$ converges. Now

$$\sum_{n=0}^{\infty}(a_n+|a_n|) - \sum_{n=0}^{\infty}|a_n| = \sum_{n=0}^{\infty} a_n + |a_n| - |a_n| = \sum_{n=0}^{\infty} a_n$$

converges by theorem 11.2.2. ∎

So given a series $\sum a_n$ with both positive and negative terms, you should first ask whether $\sum |a_n|$ converges. This may be an easier question to answer, because we have tests that apply specifically to terms with non-negative terms. If $\sum |a_n|$ converges then you know that $\sum a_n$ converges as well. If $\sum |a_n|$ diverges then it still may be true that $\sum a_n$ converges—you will have to do more work to decide the question. Another way to think of this result is: it is (potentially) easier for $\sum a_n$ to converge than for $\sum |a_n|$ to converge, because the latter series cannot take advantage of cancellation.

If $\sum |a_n|$ converges we say that $\sum a_n$ converges **absolutely**; to say that $\sum a_n$ converges absolutely is to say that any cancellation that happens to come along is not really needed, as the terms already get small so fast that convergence is guaranteed by that alone. If $\sum a_n$ converges but $\sum |a_n|$ does not, we say that $\sum a_n$ converges **conditionally**. For example $\displaystyle\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n^2}$ converges absolutely, while $\displaystyle\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}$ converges conditionally.

EXAMPLE 11.6.2 Does $\displaystyle\sum_{n=2}^{\infty}\frac{\sin n}{n^2}$ converge?

In example 11.5.2 we saw that $\displaystyle\sum_{n=2}^{\infty}\frac{|\sin n|}{n^2}$ converges, so the given series converges absolutely. □

EXAMPLE 11.6.3 Does $\displaystyle\sum_{n=0}^{\infty}(-1)^n\frac{3n+4}{2n^2+3n+5}$ converge?

Taking the absolute value, $\displaystyle\sum_{n=0}^{\infty}\frac{3n+4}{2n^2+3n+5}$ diverges by comparison to $\displaystyle\sum_{n=1}^{\infty}\frac{3}{10n}$, so if the series converges it does so conditionally. It is true that $\lim_{n\to\infty}(3n+4)/(2n^2+3n+5)=0$, so to apply the alternating series test we need to know whether the terms are decreasing. If we let $f(x)=(3x+4)/(2x^2+3x+5)$ then $f'(x)=-(6x^2+16x-3)/(2x^2+3x+5)^2$, and it is not hard to see that this is negative for $x\ge 1$, so the series is decreasing and by the alternating series test it converges. □

Exercises 11.6.

Determine whether each series converges absolutely, converges conditionally, or diverges.

1. $\displaystyle\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{2n^2+3n+5}$ ⇒

2. $\displaystyle\sum_{n=1}^{\infty}(-1)^{n-1}\frac{3n^2+4}{2n^2+3n+5}$ ⇒

3. $\displaystyle\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\ln n}{n}$ ⇒

4. $\displaystyle\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\ln n}{n^3}$ ⇒

5. $\displaystyle\sum_{n=2}^{\infty}(-1)^{n}\frac{1}{\ln n}$ ⇒

6. $\displaystyle\sum_{n=0}^{\infty}(-1)^{n}\frac{3^n}{2^n+5^n}$ ⇒

7. $\displaystyle\sum_{n=0}^{\infty}(-1)^{n}\frac{3^n}{2^n+3^n}$ ⇒

8. $\displaystyle\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\arctan n}{n}$ ⇒

11.7 The Ratio and Root Tests

Does the series $\displaystyle\sum_{n=0}^{\infty}\frac{n^5}{5^n}$ converge? It is possible, but a bit unpleasant, to approach this with the integral test or the comparison test, but there is an easier way. Consider what happens as we move from one term to the next in this series:

$$\cdots+\frac{n^5}{5^n}+\frac{(n+1)^5}{5^{n+1}}+\cdots$$

The denominator goes up by a factor of 5, $5^{n+1}=5\cdot 5^n$, but the numerator goes up by much less: $(n+1)^5=n^5+5n^4+10n^3+10n^2+5n+1$, which is much less than $5n^5$ when n is large, because $5n^4$ is much less than n^5. So we might guess that in the long run it begins to look as if each term is 1/5 of the previous term. We have seen series that behave like this:

$$\sum_{n=0}^{\infty} \frac{1}{5^n} = \frac{5}{4},$$

a geometric series. So we might try comparing the given series to some variation of this geometric series. This is possible, but a bit messy. We can in effect do the same thing, but bypass most of the unpleasant work.

The key is to notice that

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \lim_{n\to\infty} \frac{(n+1)^5}{5^{n+1}} \frac{5^n}{n^5} = \lim_{n\to\infty} \frac{(n+1)^5}{n^5} \frac{1}{5} = 1 \cdot \frac{1}{5} = \frac{1}{5}.$$

This is really just what we noticed above, done a bit more officially: in the long run, each term is one fifth of the previous term. Now pick some number between 1/5 and 1, say 1/2. Because

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \frac{1}{5},$$

then when n is big enough, say $n \ge N$ for some N,

$$\frac{a_{n+1}}{a_n} < \frac{1}{2} \quad \text{and} \quad a_{n+1} < \frac{a_n}{2}.$$

So $a_{N+1} < a_N/2$, $a_{N+2} < a_{N+1}/2 < a_N/4$, $a_{N+3} < a_{N+2}/2 < a_{N+1}/4 < a_N/8$, and so on. The general form is $a_{N+k} < a_N/2^k$. So if we look at the series

$$\sum_{k=0}^{\infty} a_{N+k} = a_N + a_{N+1} + a_{N+2} + a_{N+3} + \cdots + a_{N+k} + \cdots,$$

its terms are less than or equal to the terms of the sequence

$$a_N + \frac{a_N}{2} + \frac{a_N}{4} + \frac{a_N}{8} + \cdots + \frac{a_N}{2^k} + \cdots = \sum_{k=0}^{\infty} \frac{a_N}{2^k} = 2a_N.$$

So by the comparison test, $\sum_{k=0}^{\infty} a_{N+k}$ converges, and this means that $\sum_{n=0}^{\infty} a_n$ converges, since we've just added the fixed number $a_0 + a_1 + \cdots + a_{N-1}$.

Under what circumstances could we do this? What was crucial was that the limit of a_{n+1}/a_n, say L, was less than 1 so that we could pick a value r so that $L < r < 1$. The fact that $L < r$ ($1/5 < 1/2$ in our example) means that we can compare the series $\sum a_n$ to $\sum r^n$, and the fact that $r < 1$ guarantees that $\sum r^n$ converges. That's really all that is required to make the argument work. We also made use of the fact that the terms of the series were positive; in general we simply consider the absolute values of the terms and we end up testing for absolute convergence.

THEOREM 11.7.1 The Ratio Test Suppose that $\lim_{n\to\infty} |a_{n+1}/a_n| = L$. If $L < 1$ the series $\sum a_n$ converges absolutely, if $L > 1$ the series diverges, and if $L = 1$ this test gives no information.

Proof. The example above essentially proves the first part of this, if we simply replace $1/5$ by L and $1/2$ by r. Suppose that $L > 1$, and pick r so that $1 < r < L$. Then for $n \ge N$, for some N,

$$\frac{|a_{n+1}|}{|a_n|} > r \quad \text{and} \quad |a_{n+1}| > r|a_n|.$$

This implies that $|a_{N+k}| > r^k|a_N|$, but since $r > 1$ this means that $\lim_{k\to\infty} |a_{N+k}| \ne 0$, which means also that $\lim_{n\to\infty} a_n \ne 0$. By the divergence test, the series diverges.

To see that we get no information when $L = 1$, we need to exhibit two series with $L = 1$, one that converges and one that diverges. It is easy to see that $\sum 1/n^2$ and $\sum 1/n$ do the job. ∎

EXAMPLE 11.7.2 The ratio test is particularly useful for series involving the factorial function. Consider $\sum_{n=0}^{\infty} 5^n/n!$.

$$\lim_{n\to\infty} \frac{5^{n+1}}{(n+1)!}\frac{n!}{5^n} = \lim_{n\to\infty} \frac{5^{n+1}}{5^n}\frac{n!}{(n+1)!} = \lim_{n\to\infty} 5\frac{1}{(n+1)} = 0.$$

Since $0 < 1$, the series converges. □

A similar argument, which we will not do, justifies a similar test that is occasionally easier to apply.

THEOREM 11.7.3 The Root Test Suppose that $\lim_{n\to\infty} |a_n|^{1/n} = L$. If $L < 1$ the series $\sum a_n$ converges absolutely, if $L > 1$ the series diverges, and if $L = 1$ this test gives no information. ∎

The proof of the root test is actually easier than that of the ratio test, and is a good exercise.

EXAMPLE 11.7.4 Analyze $\displaystyle\sum_{n=0}^{\infty} \frac{5^n}{n^n}$.

The ratio test turns out to be a bit difficult on this series (try it). Using the root test:

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

Since $0 < 1$, the series converges. □

The root test is frequently useful when n appears as an exponent in the general term of the series.

Exercises 11.7.

1. Compute $\lim_{n\to\infty} |a_{n+1}/a_n|$ for the series $\sum 1/n^2$.

2. Compute $\lim_{n\to\infty} |a_{n+1}/a_n|$ for the series $\sum 1/n$.

3. Compute $\lim_{n\to\infty} |a_n|^{1/n}$ for the series $\sum 1/n^2$.

4. Compute $\lim_{n\to\infty} |a_n|^{1/n}$ for the series $\sum 1/n$.

Determine whether the series converge.

5. $\displaystyle\sum_{n=0}^{\infty}(-1)^n\frac{3^n}{5^n}$ ⇒

6. $\displaystyle\sum_{n=1}^{\infty}\frac{n!}{n^n}$ ⇒

7. $\displaystyle\sum_{n=1}^{\infty}\frac{n^5}{n^n}$ ⇒

8. $\displaystyle\sum_{n=1}^{\infty}\frac{(n!)^2}{n^n}$ ⇒

9. Prove theorem 11.7.3, the root test.

11.8 Power Series

Recall that we were able to analyze all geometric series "simultaneously" to discover that

$$\sum_{n=0}^{\infty} kx^n = \frac{k}{1-x},$$

if $|x| < 1$, and that the series diverges when $|x| \ge 1$. At the time, we thought of x as an unspecified constant, but we could just as well think of it as a variable, in which case the series

$$\sum_{n=0}^{\infty} kx^n$$

is a function, namely, the function $k/(1-x)$, as long as $|x|<1$. While $k/(1-x)$ is a reasonably easy function to deal with, the more complicated $\sum kx^n$ does have its attractions: it appears to be an infinite version of one of the simplest function types—a polynomial. This leads naturally to the questions: Do other functions have representations as series? Is there an advantage to viewing them in this way?

The geometric series has a special feature that makes it unlike a typical polynomial—the coefficients of the powers of x are the same, namely k. We will need to allow more general coefficients if we are to get anything other than the geometric series.

DEFINITION 11.8.1 A power series has the form

$$\sum_{n=0}^{\infty} a_n x^n,$$

with the understanding that a_n may depend on n but not on x. □

EXAMPLE 11.8.2 $\displaystyle\sum_{n=1}^{\infty} \frac{x^n}{n}$ is a power series. We can investigate convergence using the ratio test:

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

Thus when $|x|<1$ the series converges and when $|x|>1$ it diverges, leaving only two values in doubt. When $x=1$ the series is the harmonic series and diverges; when $x=-1$ it is the alternating harmonic series (actually the negative of the usual alternating harmonic series) and converges. Thus, we may think of $\displaystyle\sum_{n=1}^{\infty} \frac{x^n}{n}$ as a function from the interval $[-1,1)$ to the real numbers. □

A bit of thought reveals that the ratio test applied to a power series will always have the same nice form. In general, we will compute

$$\lim_{n\to\infty} \frac{|a_{n+1}||x|^{n+1}}{|a_n||x|^n} = \lim_{n\to\infty} |x|\frac{|a_{n+1}|}{|a_n|} = |x|\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = L|x|,$$

assuming that $\lim |a_{n+1}|/|a_n|$ exists. Then the series converges if $L|x|<1$, that is, if $|x|<1/L$, and diverges if $|x|>1/L$. Only the two values $x=\pm 1/L$ require further investigation. Thus the series will definitely define a function on the interval $(-1/L, 1/L)$, and perhaps will extend to one or both endpoints as well. Two special cases deserve mention: if $L = 0$ the limit is 0 no matter what value x takes, so the series converges for all x and the function is defined for all real numbers. If $L = \infty$, then no matter what value x takes the limit is infinite and the series converges only when $x = 0$. The value $1/L$ is called the **radius of convergence** of the series, and the interval on which the series converges is the **interval of convergence**.

Consider again the geometric series,

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}.$$

Whatever benefits there might be in using the series form of this function are only available to us when x is between -1 and 1. Frequently we can address this shortcoming by modifying the power series slightly. Consider this series:

$$\sum_{n=0}^{\infty} \frac{(x+2)^n}{3^n} = \sum_{n=0}^{\infty} \left(\frac{x+2}{3}\right)^n = \frac{1}{1-\frac{x+2}{3}} = \frac{3}{1-x},$$

because this is just a geometric series with x replaced by $(x+2)/3$. Multiplying both sides by $1/3$ gives

$$\sum_{n=0}^{\infty} \frac{(x+2)^n}{3^{n+1}} = \frac{1}{1-x},$$

the same function as before. For what values of x does this series converge? Since it is a geometric series, we know that it converges when

$$\begin{aligned} |x+2|/3 &< 1 \\ |x+2| &< 3 \\ -3 < x+2 &< 3 \\ -5 < x &< 1. \end{aligned}$$

So we have a series representation for $1/(1-x)$ that works on a larger interval than before, at the expense of a somewhat more complicated series. The endpoints of the interval of convergence now are -5 and 1, but note that they can be more compactly described as -2 ± 3. We say that 3 is the radius of convergence, and we now say that the series is centered at -2.

DEFINITION 11.8.3 A power series centered at a has the form

$$\sum_{n=0}^{\infty} a_n (x-a)^n,$$

with the understanding that a_n may depend on n but not on x. □

Exercises 11.8.

Find the radius and interval of convergence for each series. In exercises 3 and 4, do not attempt to determine whether the endpoints are in the interval of convergence.

1. $\sum_{n=0}^{\infty} nx^n$ ⇒

2. $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ ⇒

3. $\sum_{n=1}^{\infty} \frac{n!}{n^n} x^n$ ⇒

4. $\sum_{n=1}^{\infty} \frac{n!}{n^n} (x-2)^n$ ⇒

5. $\sum_{n=1}^{\infty} \frac{(n!)^2}{n^n} (x-2)^n$ ⇒

6. $\sum_{n=1}^{\infty} \frac{(x+5)^n}{n(n+1)}$ ⇒

11.9 Calculus with Power Series

Now we know that some functions can be expressed as power series, which look like infinite polynomials. Since calculus, that is, computation of derivatives and antiderivatives, is easy for polynomials, the obvious question is whether the same is true for infinite series. The answer is yes:

THEOREM 11.9.1 Suppose the power series $f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n$ has radius of convergence R. Then

$$f'(x) = \sum_{n=0}^{\infty} n a_n (x-a)^{n-1},$$

$$\int f(x)\,dx = C + \sum_{n=0}^{\infty} \frac{a_n}{n+1} (x-a)^{n+1},$$

and these two series have radius of convergence R as well. ■

EXAMPLE 11.9.2 Starting with the geometric series:

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

$$\int \frac{1}{1-x}\,dx = -\ln|1-x| = \sum_{n=0}^{\infty} \frac{1}{n+1} x^{n+1}$$

$$\ln|1-x| = \sum_{n=0}^{\infty} -\frac{1}{n+1} x^{n+1}$$

when $|x| < 1$. The series does not converge when $x = 1$ but does converge when $x = -1$ or $1 - x = 2$. The interval of convergence is $[-1, 1)$, or $0 < 1 - x \leq 2$, so we can use the series to represent $\ln(x)$ when $0 < x \leq 2$. For example

$$\ln(3/2) = \ln(1 - -1/2) = \sum_{n=0}^{\infty} (-1)^n \frac{1}{n+1} \frac{1}{2^{n+1}}$$

and so

$$\ln(3/2) \approx \frac{1}{2} - \frac{1}{8} + \frac{1}{24} - \frac{1}{64} + \frac{1}{160} - \frac{1}{384} + \frac{1}{896} = \frac{909}{2240} \approx 0.406.$$

Because this is an alternating series with decreasing terms, we know that the true value is between $909/2240$ and $909/2240 - 1/2048 = 29053/71680 \approx .4053$, so correct to two decimal places the value is 0.41.

What about $\ln(9/4)$? Since $9/4$ is larger than 2 we cannot use the series directly, but

$$\ln(9/4) = \ln((3/2)^2) = 2\ln(3/2) \approx 0.82,$$

so in fact we get a lot more from this one calculation than first meets the eye. To estimate the true value accurately we actually need to be a bit more careful. When we multiply by two we know that the true value is between 0.8106 and 0.812, so rounded to two decimal places the true value is 0.81. □

Exercises 11.9.

1. Find a series representation for $\ln 2$. ⇒
2. Find a power series representation for $1/(1-x)^2$. ⇒
3. Find a power series representation for $2/(1-x)^3$. ⇒
4. Find a power series representation for $1/(1-x)^3$. What is the radius of convergence? ⇒
5. Find a power series representation for $\int \ln(1-x)\,dx$. ⇒

11.10 Taylor Series

We have seen that some functions can be represented as series, which may give valuable information about the function. So far, we have seen only those examples that result from manipulation of our one fundamental example, the geometric series. We would like to start with a given function and produce a series to represent it, if possible.

Suppose that $f(x) = \sum_{n=0}^{\infty} a_n x^n$ on some interval of convergence. Then we know that we can compute derivatives of f by taking derivatives of the terms of the series. Let's look at the first few in general:

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1} = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \cdots$$

$$f''(x) = \sum_{n=2}^{\infty} n(n-1) a_n x^{n-2} = 2a_2 + 3 \cdot 2a_3 x + 4 \cdot 3a_4 x^2 + \cdots$$

$$f'''(x) = \sum_{n=3}^{\infty} n(n-1)(n-2) a_n x^{n-3} = 3 \cdot 2a_3 + 4 \cdot 3 \cdot 2a_4 x + \cdots$$

By examining these it's not hard to discern the general pattern. The kth derivative must be

$$\begin{aligned} f^{(k)}(x) &= \sum_{n=k}^{\infty} n(n-1)(n-2) \cdots (n-k+1) a_n x^{n-k} \\ &= k(k-1)(k-2) \cdots (2)(1) a_k + (k+1)(k) \cdots (2) a_{k+1} x + \\ &\quad + (k+2)(k+1) \cdots (3) a_{k+2} x^2 + \cdots \end{aligned}$$

We can shrink this quite a bit by using factorial notation:

$$f^{(k)}(x) = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!} a_n x^{n-k} = k! a_k + (k+1)! a_{k+1} x + \frac{(k+2)!}{2!} a_{k+2} x^2 + \cdots$$

Now substitute $x = 0$:

$$f^{(k)}(0) = k! a_k + \sum_{n=k+1}^{\infty} \frac{n!}{(n-k)!} a_n 0^{n-k} = k! a_k,$$

and solve for a_k:

$$a_k = \frac{f^{(k)}(0)}{k!}.$$

Note the special case, obtained from the series for f itself, that gives $f(0) = a_0$.

So if a function f can be represented by a series, we know just what series it is. Given a function f, the series

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

is called the **Maclaurin series** for f.

EXAMPLE 11.10.1 Find the Maclaurin series for $f(x) = 1/(1-x)$. We need to compute the derivatives of f (and hope to spot a pattern).

$$\begin{aligned}
f(x) &= (1-x)^{-1} \\
f'(x) &= (1-x)^{-2} \\
f''(x) &= 2(1-x)^{-3} \\
f'''(x) &= 6(1-x)^{-4} \\
f^{(4)}(x) &= 4!(1-x)^{-5} \\
&\vdots \\
f^{(n)}(x) &= n!(1-x)^{-n-1}
\end{aligned}$$

So

$$\frac{f^{(n)}(0)}{n!} = \frac{n!(1-0)^{-n-1}}{n!} = 1$$

and the Maclaurin series is

$$\sum_{n=0}^{\infty} 1 \cdot x^n = \sum_{n=0}^{\infty} x^n,$$

the geometric series. □

A warning is in order here. Given a function f we may be able to compute the Maclaurin series, but that does not mean we have found a series representation for f. We still need to know where the series converges, and if, where it converges, it converges to $f(x)$. While for most commonly encountered functions the Maclaurin series does indeed converge to f on some interval, this is not true of all functions, so care is required.

As a practical matter, if we are interested in using a series to approximate a function, we will need some finite number of terms of the series. Even for functions with messy derivatives we can compute these using computer software like Sage. If we want to know the whole series, that is, a typical term in the series, we need a function whose derivatives fall into a pattern that we can discern. A few of the most important functions are fortunately very easy.

EXAMPLE 11.10.2 Find the Maclaurin series for $\sin x$.

The derivatives are quite easy: $f'(x) = \cos x$, $f''(x) = -\sin x$, $f'''(x) = -\cos x$, $f^{(4)}(x) = \sin x$, and then the pattern repeats. We want to know the derivatives at zero: 1, 0, −1, 0, 1, 0, −1, 0,..., and so the Maclaurin series is

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

We should always determine the radius of convergence:

$$\lim_{n\to\infty} \frac{|x|^{2n+3}}{(2n+3)!}\frac{(2n+1)!}{|x|^{2n+1}} = \lim_{n\to\infty} \frac{|x|^2}{(2n+3)(2n+2)} = 0,$$

so the series converges for every x. Since it turns out that this series does indeed converge to $\sin x$ everywhere, we have a series representation for $\sin x$ for every x. Here is an interactive plot of the sine and some of its series approximations. □

Sometimes the formula for the nth derivative of a function f is difficult to discover, but a combination of a known Maclaurin series and some algebraic manipulation leads easily to the Maclaurin series for f.

EXAMPLE 11.10.3 Find the Maclaurin series for $x\sin(-x)$.

To get from $\sin x$ to $x\sin(-x)$ we substitute $-x$ for x and then multiply by x. We can do the same thing to the series for $\sin x$:

$$x\sum_{n=0}^{\infty}(-1)^n \frac{(-x)^{2n+1}}{(2n+1)!} = x\sum_{n=0}^{\infty}(-1)^n(-1)^{2n+1} \frac{x^{2n+1}}{(2n+1)!} = \sum_{n=0}^{\infty}(-1)^{n+1} \frac{x^{2n+2}}{(2n+1)!}.$$

□

As we have seen, a general power series can be centered at a point other than zero, and the method that produces the Maclaurin series can also produce such series.

EXAMPLE 11.10.4 Find a series centered at −2 for $1/(1-x)$.

If the series is $\sum_{n=0}^{\infty} a_n(x+2)^n$ then looking at the kth derivative:

$$k!(1-x)^{-k-1} = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!}a_n(x+2)^{n-k}$$

and substituting $x = -2$ we get $k!3^{-k-1} = k!a_k$ and $a_k = 3^{-k-1} = 1/3^{k+1}$, so the series is

$$\sum_{n=0}^{\infty} \frac{(x+2)^n}{3^{n+1}}.$$

We've already seen this, on page 284. □

Such a series is called the **Taylor series** for the function, and the general term has the form

$$\frac{f^{(n)}(a)}{n!}(x-a)^n.$$

A Maclaurin series is simply a Taylor series with $a = 0$.

Exercises 11.10.

For each function, find the Maclaurin series or Taylor series centered at a, and the radius of convergence.

1. $\cos x$ ⇒

2. e^x ⇒

3. $1/x$, $a = 5$ ⇒

4. $\ln x$, $a = 1$ ⇒

5. $\ln x$, $a = 2$ ⇒

6. $1/x^2$, $a = 1$ ⇒

7. $1/\sqrt{1-x}$ ⇒

8. Find the first four terms of the Maclaurin series for $\tan x$ (up to and including the x^3 term). ⇒

9. Use a combination of Maclaurin series and algebraic manipulation to find a series centered at zero for $x\cos(x^2)$. ⇒

10. Use a combination of Maclaurin series and algebraic manipulation to find a series centered at zero for xe^{-x}. ⇒

11.11 Taylor's Theorem

One of the most important uses of infinite series is the potential for using an initial portion of the series for f to approximate f. We have seen, for example, that when we add up the first n terms of an alternating series with decreasing terms that the difference between this and the true value is at most the size of the next term. A similar result is true of many Taylor series.

THEOREM 11.11.1 Suppose that f is defined on some open interval I around a and suppose $f^{(N+1)}(x)$ exists on this interval. Then for each $x \neq a$ in I there is a value z between x and a so that

$$f(x) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(N+1)}(z)}{(N+1)!}(x-a)^{N+1}.$$

Proof. The proof requires some cleverness to set up, but then the details are quite elementary. We want to define a function $F(t)$. Start with the equation

$$F(t) = \sum_{n=0}^{N} \frac{f^{(n)}(t)}{n!}(x-t)^n + B(x-t)^{N+1}.$$

Here we have replaced a by t in the first $N+1$ terms of the Taylor series, and added a carefully chosen term on the end, with B to be determined. Note that we are temporarily keeping x fixed, so the only variable in this equation is t, and we will be interested only in t between a and x. Now substitute $t = a$:

$$F(a) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + B(x-a)^{N+1}.$$

Set this equal to $f(x)$:

$$f(x) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + B(x-a)^{N+1}.$$

Since $x \neq a$, we can solve this for B, which is a "constant"—it depends on x and a but those are temporarily fixed. Now we have defined a function $F(t)$ with the property that $F(a) = f(x)$. Consider also $F(x)$: all terms with a positive power of $(x-t)$ become zero when we substitute x for t, so we are left with $F(x) = f^{(0)}(x)/0! = f(x)$. So $F(t)$ is a function with the same value on the endpoints of the interval $[a, x]$. By Rolle's theorem (6.5.1), we know that there is a value $z \in (a, x)$ such that $F'(z) = 0$. Let's look at $F'(t)$. Each term in $F(t)$, except the first term and the extra term involving B, is a product, so to take the derivative we use the product rule on each of these terms. It will help to write out the first few terms of the definition:

$$\begin{aligned} F(t) = f(t) + \frac{f^{(1)}(t)}{1!}(x-t)^1 + \frac{f^{(2)}(t)}{2!}(x-t)^2 + \frac{f^{(3)}(t)}{3!}(x-t)^3 + \cdots \\ + \frac{f^{(N)}(t)}{N!}(x-t)^N + B(x-t)^{N+1}. \end{aligned}$$

Now take the derivative:

$$\begin{aligned}
F'(t) = f'(t) &+ \left(\frac{f^{(1)}(t)}{1!}(x-t)^0(-1) + \frac{f^{(2)}(t)}{1!}(x-t)^1\right) \\
&+ \left(\frac{f^{(2)}(t)}{1!}(x-t)^1(-1) + \frac{f^{(3)}(t)}{2!}(x-t)^2\right) \\
&+ \left(\frac{f^{(3)}(t)}{2!}(x-t)^2(-1) + \frac{f^{(4)}(t)}{3!}(x-t)^3\right) + \ldots + \\
&+ \left(\frac{f^{(N)}(t)}{(N-1)!}(x-t)^{N-1}(-1) + \frac{f^{(N+1)}(t)}{N!}(x-t)^N\right) \\
&+ B(N+1)(x-t)^N(-1).
\end{aligned}$$

Now most of the terms in this expression cancel out, leaving just

$$F'(t) = \frac{f^{(N+1)}(t)}{N!}(x-t)^N + B(N+1)(x-t)^N(-1).$$

At some z, $F'(z) = 0$ so

$$\begin{aligned}
0 &= \frac{f^{(N+1)}(z)}{N!}(x-z)^N + B(N+1)(x-z)^N(-1) \\
B(N+1)(x-z)^N &= \frac{f^{(N+1)}(z)}{N!}(x-z)^N \\
B &= \frac{f^{(N+1)}(z)}{(N+1)!}.
\end{aligned}$$

Now we can write

$$F(t) = \sum_{n=0}^{N} \frac{f^{(n)}(t)}{n!}(x-t)^n + \frac{f^{(N+1)}(z)}{(N+1)!}(x-t)^{N+1}.$$

Recalling that $F(a) = f(x)$ we get

$$f(x) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(N+1)}(z)}{(N+1)!}(x-a)^{N+1},$$

which is what we wanted to show. ■

It may not be immediately obvious that this is particularly useful; let's look at some examples.

EXAMPLE 11.11.2 Find a polynomial approximation for $\sin x$ accurate to ± 0.005.

From Taylor's theorem:

$$\sin x = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(N+1)}(z)}{(N+1)!}(x-a)^{N+1}.$$

What can we say about the size of the term

$$\frac{f^{(N+1)}(z)}{(N+1)!}(x-a)^{N+1}?$$

Every derivative of $\sin x$ is $\pm\sin x$ or $\pm\cos x$, so $|f^{(N+1)}(z)| \le 1$. The factor $(x-a)^{N+1}$ is a bit more difficult, since $x-a$ could be quite large. Let's pick $a=0$ and $|x| \le \pi/2$; if we can compute $\sin x$ for $x \in [-\pi/2, \pi/2]$, we can of course compute $\sin x$ for all x.

We need to pick N so that

$$\left|\frac{x^{N+1}}{(N+1)!}\right| < 0.005.$$

Since we have limited x to $[-\pi/2, \pi/2]$,

$$\left|\frac{x^{N+1}}{(N+1)!}\right| < \frac{2^{N+1}}{(N+1)!}.$$

The quantity on the right decreases with increasing N, so all we need to do is find an N so that

$$\frac{2^{N+1}}{(N+1)!} < 0.005.$$

A little trial and error shows that $N=8$ works, and in fact $2^9/9! < 0.0015$, so

$$\begin{aligned}\sin x &= \sum_{n=0}^{8} \frac{f^{(n)}(0)}{n!} x^n \pm 0.0015\\ &= x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{5040} \pm 0.0015.\end{aligned}$$

Figure 11.11.1 shows the graphs of $\sin x$ and and the approximation on $[0, 3\pi/2]$. As x gets larger, the approximation heads to negative infinity very quickly, since it is essentially acting like $-x^7$. □

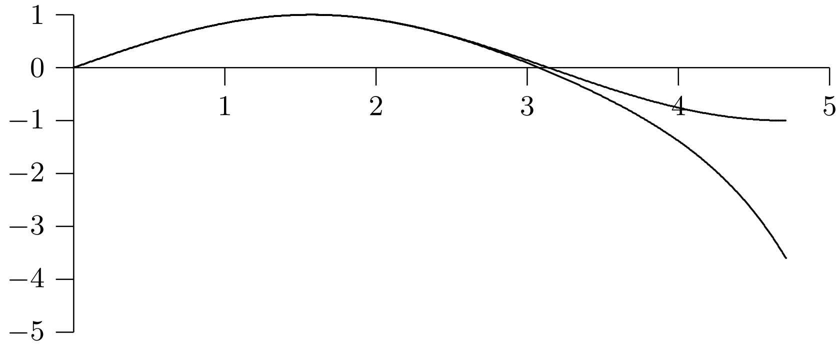


Figure 11.11.1 $\sin x$ and a polynomial approximation. (AP)

We can extract a bit more information from this example. If we do not limit the value of x, we still have

$$\left|\frac{f^{(N+1)}(z)}{(N+1)!}x^{N+1}\right| \le \left|\frac{x^{N+1}}{(N+1)!}\right|$$

so that $\sin x$ is represented by

$$\sum_{n=0}^{N} \frac{f^{(n)}(0)}{n!}\, x^n \pm \left|\frac{x^{N+1}}{(N+1)!}\right|.$$

If we can show that

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

for each x then

$$\sin x = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}\, x^n = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!},$$

that is, the sine function is actually equal to its Maclaurin series for all x. How can we prove that the limit is zero? Suppose that N is larger than $|x|$, and let M be the largest integer less than $|x|$ (if $M = 0$ the following is even easier). Then

$$\begin{aligned}\frac{|x^{N+1}|}{(N+1)!} &= \frac{|x|}{N+1}\frac{|x|}{N}\frac{|x|}{N-1}\cdots\frac{|x|}{M+1}\frac{|x|}{M}\frac{|x|}{M-1}\cdots\frac{|x|}{2}\frac{|x|}{1}\\ &\le \frac{|x|}{N+1}\cdot 1\cdot 1\cdots 1\cdot\frac{|x|}{M}\frac{|x|}{M-1}\cdots\frac{|x|}{2}\frac{|x|}{1}\\ &= \frac{|x|}{N+1}\frac{|x|^M}{M!}.\end{aligned}$$

The quantity $|x|^M/M!$ is a constant, so

$$\lim_{N\to\infty} \frac{|x|}{N+1}\frac{|x|^M}{M!} = 0$$

and by the Squeeze Theorem (11.1.3)

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

as desired. Essentially the same argument works for $\cos x$ and e^x; unfortunately, it is more difficult to show that most functions are equal to their Maclaurin series.

EXAMPLE 11.11.3 Find a polynomial approximation for e^x near $x=2$ accurate to ± 0.005.

From Taylor's theorem:

$$e^x = \sum_{n=0}^{N}\frac{e^2}{n!}(x-2)^n + \frac{e^z}{(N+1)!}(x-2)^{N+1},$$

since $f^{(n)}(x) = e^x$ for all n. We are interested in x near 2, and we need to keep $|(x-2)^{N+1}|$ in check, so we may as well specify that $|x-2|\le 1$, so $x\in[1,3]$. Also

$$\left|\frac{e^z}{(N+1)!}\right| \le \frac{e^3}{(N+1)!},$$

so we need to find an N that makes $e^3/(N+1)! \le 0.005$. This time $N=5$ makes $e^3/(N+1)! < 0.0015$, so the approximating polynomial is

$$e^x = e^2 + e^2(x-2) + \frac{e^2}{2}(x-2)^2 + \frac{e^2}{6}(x-2)^3 + \frac{e^2}{24}(x-2)^4 + \frac{e^2}{120}(x-2)^5 \pm 0.0015.$$

This presents an additional problem for approximation, since we also need to approximate e^2, and any approximation we use will increase the error, but we will not pursue this complication. □

Note well that in these examples we found polynomials of a certain accuracy only on a small interval, even though the series for $\sin x$ and e^x converge for all x; this is typical. To get the same accuracy on a larger interval would require more terms.

Exercises 11.11.

1. Find a polynomial approximation for $\cos x$ on $[0,\pi]$, accurate to $\pm 10^{-3}$ ⇒
2. How many terms of the series for $\ln x$ centered at 1 are required so that the guaranteed error on $[1/2, 3/2]$ is at most 10^{-3}? What if the interval is instead $[1, 3/2]$? ⇒
3. Find the first three nonzero terms in the Taylor series for $\tan x$ on $[-\pi/4, \pi/4]$, and compute the guaranteed error term as given by Taylor's theorem. (You may want to use Sage or a similar aid.) ⇒

4. Show that $\cos x$ is equal to its Taylor series for all x by showing that the limit of the error term is zero as N approaches infinity.

5. Show that e^x is equal to its Taylor series for all x by showing that the limit of the error term is zero as N approaches infinity.

11.12 Additional exercises

These problems require the techniques of this chapter, and are in no particular order. Some problems may be done in more than one way.

Determine whether the series converges.

1. $\displaystyle\sum_{n=0}^{\infty} \frac{n}{n^2+4}$ ⇒

2. $\displaystyle\frac{1}{1\cdot 2}+\frac{1}{3\cdot 4}+\frac{1}{5\cdot 6}+\frac{1}{7\cdot 8}+\cdots$ ⇒

3. $\displaystyle\sum_{n=0}^{\infty} \frac{n}{(n^2+4)^2}$ ⇒

4. $\displaystyle\sum_{n=0}^{\infty} \frac{n!}{8^n}$ ⇒

5. $\displaystyle 1-\frac{3}{4}+\frac{5}{8}-\frac{7}{12}+\frac{9}{16}+\cdots$ ⇒

6. $\displaystyle\sum_{n=0}^{\infty} \frac{1}{\sqrt{n^2+4}}$ ⇒

7. $\displaystyle\sum_{n=0}^{\infty} \frac{\sin^3(n)}{n^2}$ ⇒

8. $\displaystyle\sum_{n=0}^{\infty} \frac{n}{e^n}$ ⇒

9. $\displaystyle\sum_{n=0}^{\infty} \frac{n!}{1\cdot 3\cdot 5\cdots(2n-1)}$ ⇒

10. $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n\sqrt{n}}$ ⇒

11. $\displaystyle\frac{1}{2\cdot 3\cdot 4}+\frac{2}{3\cdot 4\cdot 5}+\frac{3}{4\cdot 5\cdot 6}+\frac{4}{5\cdot 6\cdot 7}+\cdots$ ⇒

12. $\displaystyle\sum_{n=1}^{\infty} \frac{1\cdot 3\cdot 5\cdots(2n-1)}{(2n)!}$ ⇒

13. $\displaystyle\sum_{n=0}^{\infty} \frac{6^n}{n!}$ ⇒

14. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{\sqrt{n}}$ ⇒

15. $\displaystyle\sum_{n=1}^{\infty} \frac{2^n 3^{n-1}}{n!}$ ⇒

16. $\displaystyle 1 + \frac{5^2}{2^2} + \frac{5^4}{(2\cdot 4)^2} + \frac{5^6}{(2\cdot 4\cdot 6)^2} + \frac{5^8}{(2\cdot 4\cdot 6\cdot 8)^2} + \cdots$ ⇒

17. $\displaystyle\sum_{n=1}^{\infty} \sin(1/n)$ ⇒

Find the interval and radius of convergence; you need not check the endpoints of the intervals.

18. $\displaystyle\sum_{n=0}^{\infty} \frac{2^n}{n!} x^n$ ⇒

19. $\displaystyle\sum_{n=0}^{\infty} \frac{x^n}{1+3^n}$ ⇒

20. $\displaystyle\sum_{n=1}^{\infty} \frac{x^n}{n3^n}$ ⇒

21. $\displaystyle x + \frac{1}{2}\frac{x^3}{3} + \frac{1\cdot 3}{2\cdot 4}\frac{x^5}{5} + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\frac{x^7}{7} + \cdots$ ⇒

22. $\displaystyle\sum_{n=1}^{\infty} \frac{n!}{n^2} x^n$ ⇒

23. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2 3^n} x^{2n}$ ⇒

24. $\displaystyle\sum_{n=0}^{\infty} \frac{(x-1)^n}{n!}$ ⇒

Find a series for each function, using the formula for Maclaurin series and algebraic manipulation as appropriate.

25. 2^x ⇒

26. $\ln(1+x)$ ⇒

27. $\displaystyle\ln\left(\frac{1+x}{1-x}\right)$ ⇒

28. $\sqrt{1+x}$ ⇒

29. $\displaystyle\frac{1}{1+x^2}$ ⇒

30. $\arctan(x)$ ⇒

31. Use the answer to the previous problem to discover a series for a well-known mathematical constant. ⇒