JARED SCHLIEPER & MICHAEL TIEMEYER

CALCULUS

Preface

A free and open-source calculus

First and foremost, this text is mostly an adaptation of two very excellent open-source textbooks: *Active Calculus* by Dr. Matt Boelkins and *APEX Calculus* by Drs. Gregory Hartman, Brian Heinold, Troy Siemers, Dimplekumar Chalishajar, and Jennifer Bowen. Both texts can be found at

`http://aimath.org/textbooks/approved-textbooks/`.

The authors of this text have combined sections, examples, and exerices from the above two texts along with some of their content to generate this text. The impetus for the creation of this text was to adopt an open-source textbook for Calculus while maintaining the typical schedule and course content of the calculus sequence at our home institution.

Several fundamental ideas in calculus are more than 2000 years old. As a formal subdiscipline of mathematics, calculus was first introduced and developed in the late 1600s, with key independent contributions from Sir Isaac Newton and Gottfried Wilhelm Leibniz. Mathematicians agree that the subject has been understood rigorously since the work of Augustin Louis Cauchy and Karl Weierstrass in the mid 1800s when the field of modern analysis was developed, in part to make sense of the infinitely small quantities on which calculus rests. Hence, as a body of knowledge calculus has been completely understood by experts for at least 150 years. The discipline is one of our great human intellectual achievements: among many spectacular ideas, calculus models how objects fall under the forces of gravity and wind resistance, explains how to compute areas and volumes of interesting shapes, enables us to work rigorously with infinitely small and infinitely large quantities, and connects the varying rates at which quantities change to the total change in the quantities themselves.

While each author of a calculus textbook certainly offers her own creative perspective on the subject, it is hardly the case that many of the ideas she presents are new. Indeed, the mathematics community broadly agrees on what the main ideas of calculus are, as well as their justification and their importance; the core parts of nearly all calculus textbooks are very similar. As such, it is our opinion that in the 21st century – an age where the internet permits seamless and immediate transmission of information – no one should be required to purchase a calculus text to read, to use for a class, or to find a coherent collection of problems to solve. Calculus belongs to humankind, not any individual author or publishing company. Thus, the main purpose of this work is to present a new calculus text that is *free*. In addition, instructors who are looking for a calculus text should have the opportunity to download the source files and make modifications that they see fit; thus this text is *open-source*.

Because the text is free and open-source, any professor or student may use and/or change the electronic version of the text for no charge. Presently, a .pdf copy of the text and its source files may be obtained by download from Github (insert link here!!) This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 Unported License. The graphic

that appears throughout the text shows that the work is licensed with the Creative Commons, that the work may be used for free by any party so long as attribution is given to the author(s), that the work and its derivatives are used in the spirit of "share and share alike," and that no party may sell this work or any of its derivatives for profit, with the following exception: *it is entirely acceptable for university bookstores to sell bound photocopied copies to students at their standard markup above the copying expense.* Full details may be found by visiting

`http://creativecommons.org/licenses/by-nc-sa/4.0/`

or sending a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Acknowledgments

We would like to thank Affordable Learning Georgia for awarding us a Textbook Transformation Grant, which allotted a two-course release for each of us to generate this text. Please see

`http://affordablelearninggeorgia.org/`

for more information on this initiative to promote student success by providing affordable textbook alternatives.

We will gladly take reader and user feedback to correct them, along with other suggestions to improve the text.

Jared Schlieper & Michael Tiemeyer

Contents

Chapter 0

Introduction to Calculus

0.1 Why do we study calculus?

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- What is the behavior of a function arbitrarily close to, but not necessarily at, a specific point?
- Given a specific input in the domain of a function, what is the instantaneous rate of change of the function at that point?
- Given a function that represents the rate of change of some quantity over a specific time interval, how much of that quantity has accumulated over that time interval?
- How are instantaneous rate of change and accumulation related?

Introduction

Calculus is all about answering these Motivating Questions for all functions in general, but let's first consider piece-wise linear functions so that we may illustrate the ideas of questions two and three.

Let's start with a car that's driving on a long, flat, straight road. Instead of having a speedometer and odometer, the car is equipped with a *velocitometer* and *positometer*. A velocitometer measures velocity, which is positive when the car moves forward, negative when the car moves backward, and zero when the car is at rest. A positometer measures position away from some starting point–typically the origin, which may be positive, negative, or zero.

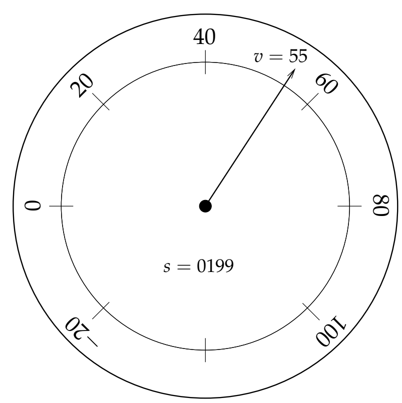


Figure 1: A velocitometer with positometer.

We'll denote the velocity with v, which will have units measured in miles per hour, and we'll denote the position with s, which will have units measured in miles.

Suppose the position of the car increases linearly such that at time $t = 2$, the position is 110 miles, and at time $t = 4$, the position is 220 miles. Notice that since this function is a straight

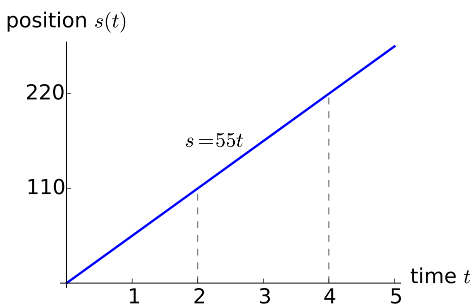


Figure 2: A linear position function.

line, or linear, we can compute its slope, which is the rate of change of a linear function. The slope is

$$\begin{aligned} v &= \frac{\text{change in position}}{\text{change in time}} = \frac{220 \text{ miles} - 110 \text{ miles}}{4 \text{ hours} - 2 \text{ hours}} \\ &= \frac{110 \text{ miles}}{2 \text{ hours}} = 55 \text{ miles per hour.} \end{aligned}$$

As you can see in the equation above, we used v to denote the slope instead of the usual m. That's because velocity is defined exactly to be the ratio of the change in position to the change of time.

Also notice that with linear functions, the slope is constant, meaning the slope is the same at every point and over every interval. So if we wish to know the average or instantaneous rate of change of a linear function, we simply need to calculate the slope of the line!

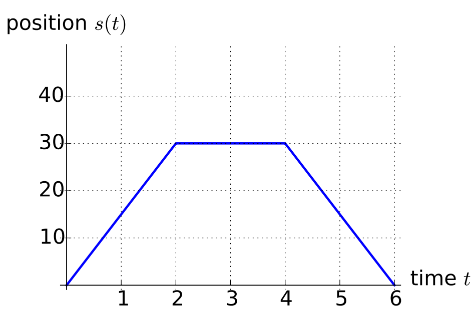


Figure 3: A piece-wise linear position function.

Example 1

Use Figure 3 to determine the instantaneous rate of change, or velocity, of the car at $t = 1$, $t = 3$, and $t = 5$.

Solution. When $t = 1$, the position of the car is determined by the line segment passing through the points $(0, 0)$ and $(2, 3)$. The slope of that segment is then

$$\text{slope} = \frac{30 - 0}{2 - 0} = \frac{30}{2} = 15,$$

and so the instantaneous rate of change of the car is 15 miles per hour at $t = 1$.

When $t = 3$, the position of the car is determined by the line segment passing through the points $(2, 30)$ and $(4, 30)$. The slope of that segment is

$$\text{slope} = \frac{30 - 30}{4 - 2} = 0,$$

and so the instantaneous rate of change of the car is 0 miles per hour at $t = 3$.

Finally, when $t = 5$, the position of the car is determined by the line segment passing through the points $(4, 30)$ and $(6, 0)$. The slope of that segment is then

$$\text{slope} = \frac{30 - 0}{4 - 6} = \frac{30}{-2} = -15,$$

and so the instantaneous rate of change of the car is -15 miles per hour at $t = 5$.

Now suppose the velocity of the car is constant at $v = 55$ miles per hour starting at time $t = 0$. If we wanted to find the position of the car after $t = 3$ hours, then we would simply use the equation

$$\begin{aligned}\text{position} &= \text{rate} \times \text{time}\\ &= 55 \text{ miles per hour} \times 3 \text{ hours}\\ &= 165 \text{ miles.}\end{aligned}$$

After 5 hours, the postion would be

$$\begin{aligned}\text{position} &= \text{rate} \times \text{time}\\ &= 55 \text{ miles per hour} \times 5 \text{ hours}\\ &= 275 \text{ miles.}\end{aligned}$$

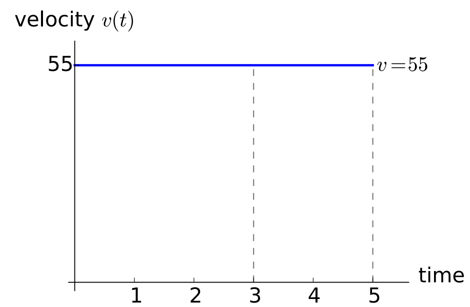


Figure 4: The constant velocity function.

Geometrically, we can represent the position of the car after 3 hours by the area of the rectangle created by the constant function $v(t) = 55$, the t-axis, and the vertical lines $t = 0$ and $t = 3$. Similarly, we can represent the position after 5 hours geometrically as the area of the region bounded by the t-axis and $v(t) = 55$ from $t = 0$ to $t = 5$. Notice that region is also a rectangle with height equal to 55 and width equal to $5 - 0 = 5$.

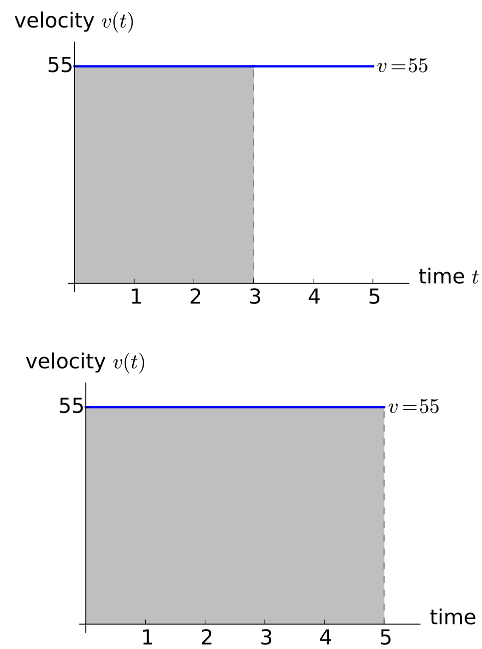


Figure 5: Position at $t = 3$ and $t = 5$ represented by the area of a rectangle.

So in general, if we have a piece-wise constant function–a function consisting of segments of horizontal lines–that represents the velocity of the car (or any other object!), then we can find the position of the car at time t by summing the areas of the rectangles underneath the piece-wise constant function from time 0 to time t.

Example 2

Use Figure 6 to determine the position of the car at time $t = 6$.

Solution. In the figure, there are four horizontal line segments that comprise the piece-wise constant function. So there will be four rectangles whose areas we must calculate, and then we will sum those areas to determine the position of the car at $t = 6$.

From $t = 0$ to $t = 2$, the rectangle has height 10 and width 2; therefore, its area is $10 \times 2 = 20$.

From $t = 2$ to $t = 3$, the rectangle has height 40 and width 1; therefore, its area is $40 \times 1 = 40$.

From $t = 3$ to $t = 5$, the rectangle has height 30 and width 2; therefore, its area is $30 \times 2 = 60$.

And from $t = 5$ to $t = 6$, the rectangle has height 20 and width 1; therefore, its area is $20 \times 1 = 20$.

So the position of the car at $t = 6$ is $20 + 40 + 60 + 20 = 140$ miles.

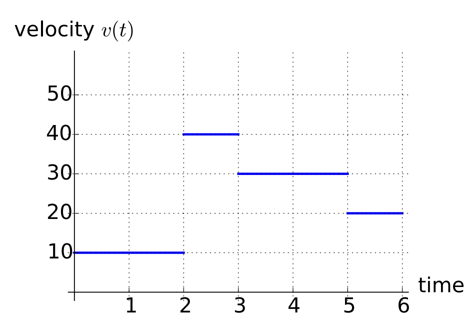


Figure 6: A piece-wise constant velocity function.

What if the velocity function is a piece-wise linear function instead of a piece-wise constant function as seen in Figure 7? Nothing changes with respect to *how* we find the position–we still must find the area underneath the function, but the region(s) may not be rectangular. They may be triangles, trapezoids, or rectangles, which means we may need to remember

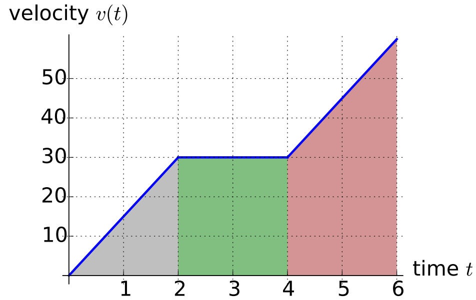


Figure 7: A piece-wise linear velocity function.

the area formulas for those polygons. But it is certainly possible to calculate the position from a piece-wise linear velocity function using only algebra.

Slope and Area

The slope of the position graph s at some point gives the velocity v at that point. The area of the region underneath the velocity graph v from time 0 to time t gives the position s at time t.

So what if we wish to find the instantaneous rate of change, or slope, of a function that is not piece-wise linear? Or what if we wish to find the total accumulated amount of some quantity, or the area of a region underneath a function, when that function is not piece-wise linear? Continue reading...

0.2 *How do we measure velocity?*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- How is the average velocity of a moving object connected to the values of its position function?
- How do we interpret the average velocity of an object geometrically with regard to the graph of its position function?
- How is the notion of instantaneous velocity connected to average velocity?

Introduction

We begin with a familiar problem: a ball being tossed straight up in the air from an initial height. From this elementary scenario, we will ask questions about how the ball is moving. These questions will lead us to begin investigating ideas that will be central throughout our study of differential calculus and that have wide-ranging consequences. In a great deal of our thinking about calculus, we will be well-served by remembering this first example and asking ourselves how the various (sometimes abstract) ideas we are considering are related to the simple act of tossing a ball straight up in the air.

Preview Activity 0.2

Suppose that the height s of a ball (in feet) at time t (in seconds) is given by the formula $s(t) = 64 - 16(t-1)^2$.

(a) Construct an accurate graph of $y = s(t)$ on the time interval $0 \le t \le 3$. Label at least six distinct points on the graph, including the three points that correspond to when the ball was released, when the ball reaches its highest point, and when the ball lands.

(b) In everyday language, describe the behavior of the ball on the time interval $0 < t < 1$ and on time interval $1 < t < 3$. What occurs at the instant $t = 1$?

(c) Consider the expression

$$AV_{[0.5,1]} = \frac{s(1) - s(0.5)}{1 - 0.5}.$$

Compute the value of $AV_{[0.5,1]}$. What does this value measure geometrically? What does this value measure physically? In particular, what are the units on $AV_{[0.5,1]}$?

Position and average velocity

Any moving object has a *position* that can be considered a function of *time*. When this motion is along a straight line, the position is given by a single variable, and we usually let this position be denoted by $s(t)$, which reflects the fact that position is a function of time. For example, we might view $s(t)$ as telling the mile marker of a car traveling on a straight highway at time t in hours; similarly, the function s described in Preview Activity 0.2 is a position function, where position is measured vertically relative to the ground.

Not only does such a moving object have a position associated with its motion, but on any time interval, the object has an *average velocity*. Think, for example, about driving from one location to another: the vehicle travels some number of miles over a certain time interval (measured in hours), from which we can compute the vehicle's average velocity. In this situation, average velocity is the number of miles traveled divided by the time elapsed, which of course is given in *miles per hour*. Similarly, the calculation of $A_{[0.5,1]}$ in Preview Activity 0.2 found the average velocity of the ball on the time interval $[0.5, 1]$, measured in feet per second.

In general, we make the following definition: for an object moving in a straight line whose position at time t is given by the function $s(t)$, the *average velocity of the object on the interval from $t = a$ to $t = b$*, denoted $AV_{[a,b]}$, is given by the formula

$$AV_{[a,b]} = \frac{s(b) - s(a)}{b - a}.$$

Note well: the units on $AV_{[a,b]}$ are "units of s per unit of t," such as "miles per hour" or "feet per second."

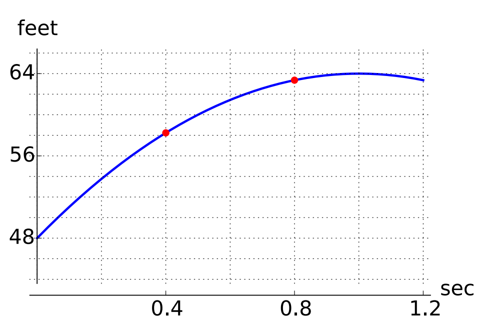


Figure 8: A partial plot of $s(t) = 64 - 16(t-1)^2$.

Activity 0.2–1

The following questions concern the position function given by $s(t) = 64 - 16(t-1)^2$, which is the same function considered in Preview Activity 0.2.

(a) Compute the average velocity of the ball on each of the following time intervals: $[0.4, 0.8]$, $[0.7, 0.8]$, $[0.79, 0.8]$, $[0.799, 0.8]$, $[0.8, 1.2]$, $[0.8, 0.9]$, $[0.8, 0.81]$, $[0.8, 0.801]$. Include units for each value.

(b) On the provided graph in Figure 8, sketch the line that passes through the points $A = (0.4, s(0.4))$ and $B = (0.8, s(0.8))$. What is the meaning of the slope of this line? In light of this meaning, what is a geometric way to interpret each of the values computed in the preceding question?

(c) Use a graphing utility to plot the graph of $s(t) = 64 - 16(t-1)^2$ on an interval containing the value $t = 0.8$. Then, zoom in repeatedly on the point $(0.8, s(0.8))$. What do you observe about how the graph appears as you view it more and more closely?

(d) What do you conjecture is the velocity of the ball at the instant $t = 0.8$? Why?

Instantaneous Velocity

Whether driving a car, riding a bike, or throwing a ball, we have an intuitive sense that any moving object has a velocity at any given moment – a number that measures how fast the object is moving *right now*. For instance, a car's speedometer tells the driver what appears to be the car's velocity at any given instant. In fact, the posted velocity on a speedometer is really an average velocity that is computed over a very small time interval (by computing how many revolutions the tires have undergone to compute distance traveled), since velocity fundamentally comes from considering a change in position divided by a change in time. But if we let the time interval over which average velocity is computed become shorter and shorter, then we can progress from average velocity to *instantaneous* velocity.

Informally, we define the *instantaneous velocity* of a moving object at time $t = a$ to be the value that the average velocity approaches as we take smaller and smaller intervals of time containing $t = a$ to compute the average velocity. We will develop a more formal definition of this momentarily, one that will end up being the foundation of much of our work in first semester calculus. For now, it is fine to think of instantaneous velocity this way: take average velocities on smaller and smaller time intervals, and if those average velocities approach a single number, then that number will be the instantaneous velocity at that point.

Activity 0.2–2

Each of the following questions concern $s(t) = 64 - 16(t-1)^2$, the position function from Preview Activity 0.2.

(a) Compute the average velocity of the ball on the time interval $[1.5, 2]$. What is different between this value and the average velocity on the interval $[0, 0.5]$?

(b) Use appropriate computing technology to estimate the instantaneous velocity of the ball at $t = 1.5$. Likewise, estimate the instantaneous velocity of the ball at $t = 2$. Which value is greater?

(c) How is the sign of the instantaneous velocity of the ball related to its behavior at a given point in time? That is, what does positive instantaneous velocity tell you the ball is doing? Negative instantaneous velocity?

(d) Without doing any computations, what do you expect to be the instantaneous velocity of the ball at $t = 1$? Why?

At this point we have started to see a close connection between average velocity and instantaneous velocity, as well as how each is connected not only to the physical behavior of the moving object but also to the geometric behavior of the graph of the position function. In order to make the link between average and instantaneous velocity more formal, we will introduce the notion of *limit* in Section 1.1. As a preview of that concept, we look at a way to consider the limiting value of average velocity through the introduction of a parameter. Note that if we desire to know the instantaneous velocity at $t = a$ of a moving object with position function s, we are interested in computing average velocities on the interval $[a, b]$ for smaller and smaller intervals. One way to visualize this is to think of the value b as being $b = a + h$, where h is a small number that is allowed to vary. Thus, we observe that the average velocity of the object on the interval $[a, a + h]$ is

$$AV_{[a,a+h]} = \frac{s(a+h) - s(a)}{h},$$

with the denominator being simply h because $(a + h) - a = h$. Initially, it is fine to think of h being a small positive real number; but it is important to note that we allow h to be a small negative number, too, as this enables us to investigate the average velocity of the moving object on intervals prior to $t = a$, as well as following $t = a$. When $h < 0$, $AV_{[a,a+h]}$ measures the average velocity on the interval $[a + h, a]$.

To attempt to find the instantaneous velocity at $t = a$, we investigate what happens as the value of h approaches zero. We consider this further in the following example.

Example 1

For a falling ball whose position function is given by $s(t) = 16 - 16t^2$ where s is measured in feet and t in seconds, find an expression for the average velocity of the ball on a time interval of the form $[0.5, 0.5 + h]$ where $-0.5 < h < 0.5$ and $h \neq 0$. Use this expression to compute the average velocity on $[0.5, 0.75]$ and $[0.4, 0.5]$, as well as to make a conjecture about the instantaneous velocity at $t = 0.5$.

Solution. We make the assumptions that $-0.5 < h < 0.5$ and $h \neq 0$ because h cannot be zero (otherwise there is no interval on which to compute average velocity) and because the function only makes sense on the time interval $0 \leq t \leq 1$, as this is the duration of time during which the ball is falling. Observe that we want to compute and simplify

$$AV_{[0.5,0.5+h]} = \frac{s(0.5+h) - s(0.5)}{(0.5+h) - 0.5}.$$

The most unusual part of this computation is finding $s(0.5 + h)$. To do so, we follow the rule that defines the function s. In particular, since

$s(t) = 16 - 16t^2$, we see that

$$\begin{aligned} s(0.5+h) &= 16 - 16(0.5+h)^2 \\ &= 16 - 16(0.25 + h + h^2) \\ &= 16 - 4 - 16h - 16h^2 \\ &= 12 - 16h - 16h^2. \end{aligned}$$

Now, returning to our computation of the average velocity, we find that

$$\begin{aligned} AV_{[0.5,0.5+h]} &= \frac{s(0.5+h) - s(0.5)}{(0.5+h) - 0.5} \\ &= \frac{(12 - 16h - 16h^2) - (16 - 16(0.5)^2)}{0.5 + h - 0.5} \\ &= \frac{12 - 16h - 16h^2 - 12}{h} \\ &= \frac{-16h - 16h^2}{h}. \end{aligned}$$

At this point, we note two things: first, the expression for average velocity clearly depends on h, which it must, since as h changes the average velocity will change. Further, we note that since h can never equal zero, we may further simplify the most recent expression. Removing the common factor of h from the numerator and denominator, it follows that

$$AV_{[0.5,0.5+h]} = -16 - 16h.$$

Now, for any small positive or negative value of h, we can compute the average velocity. For instance, to obtain the average velocity on $[0.5, 0.75]$, we let $h = 0.25$, and the average velocity is $-16 - 16(0.25) = -20$ ft/sec. To get the average velocity on $[0.4, 0.5]$, we let $h = -0.1$, which tells us the average velocity is $-16 - 16(-0.1) = -14.4$ ft/sec. Moreover, we can even explore what happens to $AV_{[0.5,0.5+h]}$ as h gets closer and closer to zero. As h approaches zero, $-16h$ will also approach zero, and thus it appears that the instantaneous velocity of the ball at $t = 0.5$ should be -16 ft/sec.

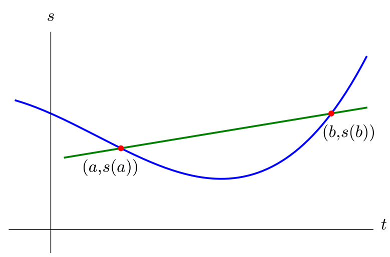
Activity 0.2–3

For the function given by $s(t) = 64 - 16(t-1)^2$ from Preview Activity 0.2, find the most simplified expression you can for the average velocity of the ball on the interval $[2, 2+h]$. Use your result to compute the average velocity on $[1.5, 2]$ and to estimate the instantaneous velocity at $t = 2$. Finally, compare your earlier work in Activity 0.2–1.

Summary

In this section, we encountered the following important ideas:

- The average velocity on $[a, b]$ can be viewed geometrically as the slope of the line between the points $(a, s(a))$ and $(b, s(b))$ on the graph of $y = s(t)$, as shown below.



- Given a moving object whose position at time t is given by a function s, the average velocity of the object on the time interval $[a, b]$ is given by $AV_{[a,b]} = \frac{s(b)-s(a)}{b-a}$. Viewing the interval $[a, b]$ as having the form $[a, a + h]$, we equivalently compute average velocity by the formula $AV_{[a,a+h]} = \frac{s(a+h)-s(a)}{h}$.
- The instantaneous velocity of a moving object at a fixed time is estimated by considering average velocities on shorter and shorter time intervals that contain the instant of interest.

Exercises

1. A bungee jumper dives from a tower at time $t = 0$. Her height h (measured in feet) at time t (in seconds) is given by the graph in Figure 9.

   In this problem, you may base your answers on estimates from the graph or use the fact that the jumper's height function is given by $s(t) = 100\cos(0.75t) \cdot e^{-0.2t} + 100$.

   (a) What is the change in vertical position of the bungee jumper between $t = 0$ and $t = 15$?

   (b) Estimate the jumper's average velocity on each of the following time intervals: $[0, 15]$, $[0, 2]$, $[1, 6]$, and $[8, 10]$. Include units on your answers.

   (c) On what time interval(s) do you think the bungee jumper achieves her greatest average velocity? Why?

   (d) Estimate the jumper's instantaneous velocity at $t = 5$. Show your work and explain your reasoning, and include units on your answer.

   (e) Among the average and instantaneous velocities you computed in earlier questions, which are positive and which are negative? What does negative velocity indicate?

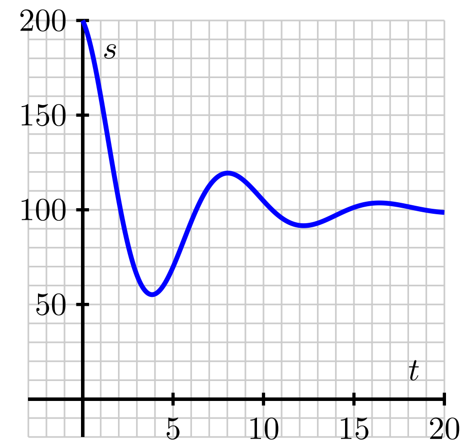


Figure 9: A bungee jumper's height function.

2. A diver leaps from a 3 meter springboard. His feet leave the board at time $t = 0$, he reaches his maximum height of 4.5 m at $t = 1.1$ seconds, and enters the water at $t = 2.45$. Once in the water, the diver coasts to the bottom of the pool (depth 3.5 m), touches bottom at $t = 7$, rests for one second, and then pushes off the bottom. From there he coasts to the surface, and takes his first breath at $t = 13$.

   (a) Let $s(t)$ denote the function that gives the height of the diver's feet (in meters) above the water at time t. (Note that the "height" of the bottom of the pool is -3.5 meters.) Sketch a carefully labeled graph of $s(t)$ on the provided axes in Figure 10. Include scale and units on the vertical axis. Be as detailed as possible.

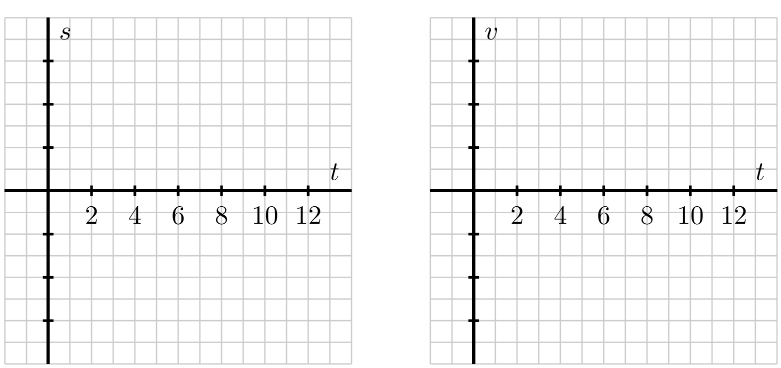


Figure 10: Axes for plotting $s(t)$ in part (a) and $v(t)$ in part (c) of the diver problem.

   (b) Based on your graph in (a), what is the average velocity of the diver between $t = 2.45$ and $t = 7$? Is his average velocity the same on every time interval within $[2.45, 7]$?

   (c) Let the function $v(t)$ represent the *instantaneous vertical velocity* of the diver at time t (i.e. the speed at which the height function $s(t)$ is changing; note that velocity in the upward direction is positive, while the velocity of a falling object is negative). Based on your understanding of the diver's behavior, as well as your graph of the position function, sketch a carefully labeled graph of $v(t)$ on the axes provided in Figure 10. Include scale and units on the vertical axis. Write several sentences that explain how you constructed your graph, discussing when you expect $v(t)$ to be zero, positive, negative, relatively large, and relatively small.

(d) Is there a connection between the two graphs that you can describe? What can you say about the velocity graph when the height function is increasing? decreasing? Make as many observations as you can.

3. According to the U.S. census, the population of the city of Grand Rapids, MI, was 181,843 in 1980; 189,126 in 1990; and 197,800 in 2000.

(a) Between 1980 and 2000, by how many people did the population of Grand Rapids grow?

(b) In an average year between 1980 and 2000, by how many people did the population of Grand Rapids grow?

(c) Just like we can find the average velocity of a moving body by computing change in position over change in time, we can compute the average rate of change of any function f. In particular, the *average rate of change* of a function f over an interval $[a, b]$ is the quotient

$$\frac{f(b) - f(a)}{b - a}.$$

What does the quantity $\frac{f(b)-f(a)}{b-a}$ measure on the graph of $y = f(x)$ over the interval $[a, b]$?

(d) Let $P(t)$ represent the population of Grand Rapids at time t, where t is measured in years from January 1, 1980. What is the average rate of change of P on the interval $t = 0$ to $t = 20$? What are the units on this quantity?

(e) If we assume the the population of Grand Rapids is growing at a rate of approximately 4% per decade, we can model the population function with the formula

$$P(t) = 181843(1.04)^{t/10}.$$

Use this formula to compute the average rate of change of the population on the intervals $[5, 10]$, $[5, 9]$, $[5, 8]$, $[5, 7]$, and $[5, 6]$.

(f) How fast do you think the population of Grand Rapids was changing on January 1, 1985? Said differently, at what rate do you think people were being added to the population of Grand Rapids as of January 1, 1985? How many additional people should the city have expected in the following year? Why?

Chapter 1
Limits

1.1 The notion of limit

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- What is the behavior of a function arbitrarily close to, but not necessarily at, a specific point?
- What is the mathematical notion of *limit* and what role do limits play in the study of functions?
- What is the meaning of the notation $\lim_{x \to a} f(x) = L$?
- How do we go about determining the value of the limit of a function at a point?
- How does the notion of limit allow us to move from average velocity to instantaneous velocity?

Introduction

Functions are at the heart of mathematics: a function is a process or rule that associates each individual input to exactly one corresponding output. Students learn in courses prior to calculus that there are many different ways to represent functions, including through formulas, graphs, tables, and even words. For example, the squaring function can be thought of in any of these ways. In words, the squaring function takes any real number x and computes its square. The formulaic and graphical representations go hand in hand, as $y = f(x) = x^2$ is one of the simplest curves to graph. Finally, we can also partially represent this function through a table of values, essentially by listing some of the ordered pairs that lie on the curve, such as $(-2,4)$, $(-1,1)$, $(0,0)$, $(1,1)$, and $(2,4)$.

Functions are especially important in calculus because they often model important phenomena – the location of a moving object at a given time, the rate at which an automobile is consuming gasoline at a certain velocity, the reaction of a patient to the size of a dose of a drug – and calculus can be used to study how these output quantities change in response to changes in the input variable. Moreover, thinking about concepts like average and instantaneous velocity leads us naturally from an initial function to a related, sometimes more complicated function. As one example of this, think about the falling ball whose position function is given by $s(t) = 64 - 16t^2$ and the average velocity of the ball on the interval $[1, x]$. Observe that

$$AV_{[1,x]} = \frac{s(x) - s(1)}{x - 1} = \frac{(64 - 16x^2) - (64 - 16)}{x - 1} = \frac{16 - 16x^2}{x - 1}.$$

Now, two things are essential to note: this average velocity depends on x (indeed, $AV_{[1,x]}$ is a function of x), and our most focused interest in this function occurs near $x = 1$, which is where the function is not defined. Said differently, the function $g(x) = \frac{16-16x^2}{x-1}$ tells us the average velocity of the ball on the interval from $t = 1$ to $t = x$, and if we are interested in the instantaneous velocity of the ball when $t = 1$, we'd like to know what happens to $g(x)$ as x gets closer and closer to 1. At the same time, $g(1)$ is not defined, because it leads to the quotient $0/0$.

This is where the idea of *limits* comes in. By using a limit, we'll be able to allow x to get arbitrarily close, but not equal, to 1 and fully understand the behavior of $g(x)$ near this value. We'll develop key language, notation, and conceptual understanding in what follows, but for now we consider a preliminary activity that uses the graphical interpretation of a function to explore points on a graph where interesting behavior occurs.

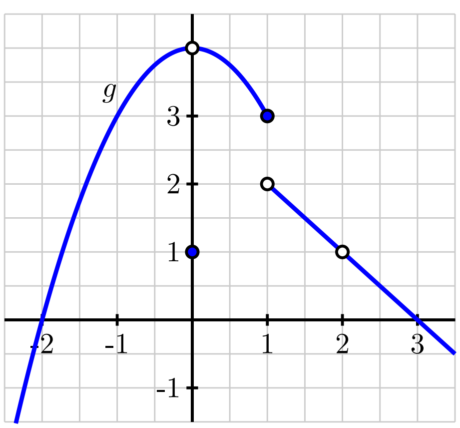


Figure 1.1: Graph of $y = g(x)$ for Preview Activity 1.1.

Preview Activity 1.1

Suppose that g is the function given by the Figure 1.1. Use the graph to answer each of the following questions.

(a) Determine the values $g(-2)$, $g(-1)$, $g(0)$, $g(1)$, and $g(2)$, if defined. If the function value is not defined, explain what feature of the graph tells you this.

(b) For each of the values $a = -1$, $a = 0$, and $a = 2$, complete the following sentence: "As x gets closer and closer (but not equal) to a, $g(x)$ gets as close as we want to _____."

(c) What happens as x gets closer and closer (but not equal) to $a = 1$? Does the function $g(x)$ get as close as we would like to a single value?

The Idea of a Limit

Limits can be thought of as a way to study the tendency or trend of a function as the input variable approaches a fixed value, or even as the input variable increases or decreases without bound. Here, we focus on what it means to say that "a function f has limit L as x approaches a." To begin, we think about a recent example.

In Preview Activity 1.1, you saw that for the given function g, as x gets closer and closer (but not equal) to 0, $g(x)$ gets as close as we want to the value 4. At first, this may feel counterintuitive, because the value of $g(0)$ is 1, not 4. By their very definition, limits regard the behavior of a function *arbitrarily close to* a fixed input, but the value of the function *at* the fixed input does not matter. More formally[1], we say the following.

[1] What follows here is not what mathematicians consider the formal definition of a limit. To be completely precise, it is necessary to quantify both what it means to say "as close to L as we like" and "sufficiently close to a." That can be accomplished through what is traditionally called the epsilon-delta definition of limits, which will be covered in Section 1.4.

The Limit of a Function

Given a function f, a fixed input $x = a$, and a real number L, we say that f *has limit L as x approaches a*, and write

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

provided that we can make $f(x)$ as close to L as we like by taking x sufficiently close (but not equal) to a. If we cannot make $f(x)$ as close to a single value as we would like as x approaches a, then we say that f *does not have a limit as x approaches a.*

For the function g pictured in Figure 1.1, we can make the following observations:

$$\lim_{x \to -1} g(x) = 3, \ \lim_{x \to 0} g(x) = 4, \text{ and } \lim_{x \to 2} g(x) = 1,$$

but g does not have a limit as $x \to 1$. When working graphically, it suffices to ask if the function approaches a single value from each side of the fixed input, while understanding that the function value right at the fixed input is irrelevant. This reasoning explains the values of the first three stated limits. In a situation such as the jump in the graph of g at $x = 1$, the issue is that if we approach $x = 1$ from the left, the function values tend to get as close to 3 as we'd like, but if we approach $x = 1$ from the right, the function values get as close to 2 as we'd like, and there is no single number that all of these function values approach. This is why the limit of g does not exist at $x = 1$.

For any function f, there are typically three ways to answer the question "does f have a limit at $x = a$, and if so, what is the limit?" The first is to reason graphically as we have just done with the example from Preview Activity 1.1. If we have a formula for $f(x)$, there are two additional possibilities: (1) evaluate the function at a sequence of inputs that approach a on either side, typically using some sort of computing technology, and ask if the sequence of outputs seems to approach a single value; (2) use the algebraic form of the function to understand the trend in its output as the input values approach a. The first approach only produces an approximation of the value of the limit, while the latter can often be used to determine the limit exactly. The following examples demonstrate the approaches of using a table and graph to evaluate limits.

x	$f(x)$
-0.9	2.9
-0.99	2.99
-0.999	2.999
-0.9999	2.9999
-1.1	3.1
-1.01	3.01
-1.001	3.001
-1.0001	3.0001

x	$f(x)$
-1.9	3.9
-1.99	3.99
-1.999	3.999
-1.9999	3.9999
-2.1	4.1
-2.01	4.01
-2.001	4.001
-2.0001	4.0001

Table 1.1: Function values near -1 and -2.

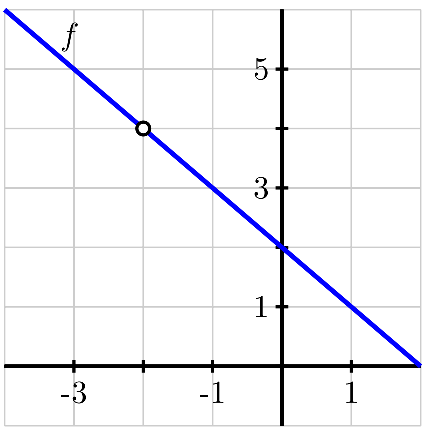


Figure 1.2: The graph of $f(x) = \frac{4-x^2}{x+2}$ near -1 and -2.

Example 1

Use both tables and graphical approaches to investigate and, if possible, estimate or determine the value of the limit for the following function at the specified values.

$$f(x) = \frac{4-x^2}{x+2}; \quad a = -1,\ a = -2$$

Solution. We first construct tables of values near $a = -1$ and $a = -2$, see Table 1.1, along with a graph of f, see Figure 1.2.

From the left table, it appears that we can make f as close as we want to 3 by taking x sufficiently close to -1, which suggests that $\lim\limits_{x\to -1} f(x) = 3$. This is also consistent with the graph of f.

From the right table, it appears that we can make f as close as we want to 4 by taking x sufficiently close, but not equal since $f(-2)$ is not defined, to -2, which suggests that $\lim\limits_{x\to -4} f(x) = 4$. Remember, limits ask, "To where is the function going?", not "What is the value of the function?" Again, this observation is consistent with the graph of f.

x	$g(x)$
2.9	0.84864
2.99	0.86428
2.999	0.86585
2.9999	0.86601
3.1	0.88351
3.01	0.86777
3.001	0.86620
3.0001	0.86604

x	$g(x)$
-0.1	0
-0.01	0
-0.001	0
-0.0001	0
0.1	0
0.01	0
0.001	0
0.0001	0

Table 1.2: Tables for $g(x) = \sin\left(\frac{\pi}{x}\right)$.

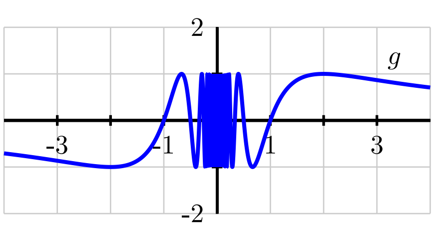


Figure 1.3: The graph of $g(x) = \sin\left(\frac{\pi}{x}\right)$ near 3 and 0.

Example 2

Use both tables and graphical approaches to investigate and, if possible, estimate or determine the value of the limit for the following function at the specified values.

$$g(x) = \sin\left(\frac{\pi}{x}\right); \quad a = 3,\ a = 0$$

Solution. Again, we construct two tables and a graph, see Table 1.2 and Figure 1.3.

First, as $x \to 3$, it appears from the data (and the graph) that the function is approaching approximately 0.866025. To be precise, we have to use the fact that $\frac{\pi}{x} \to \frac{\pi}{3}$, and thus we find that $g(x) = \sin(\frac{\pi}{x}) \to \sin(\frac{\pi}{3})$ as $x \to 3$. The exact value of $\sin(\frac{\pi}{3})$ is $\frac{\sqrt{3}}{2}$, which is approximately 0.8660254038. Thus, we see that

$$\lim_{x\to 3} g(x) = \frac{\sqrt{3}}{2}.$$

As $x \to 0$, we observe that $\frac{\pi}{x}$ does not behave in an elementary way. When x is positive and approaching zero, we are dividing by smaller and smaller positive values, and $\frac{\pi}{x}$ increases without bound. When x is negative and approaching zero, $\frac{\pi}{x}$ decreases without bound. In this sense, as we get close to $x = 0$, the inputs to the sine function are growing rapidly, and this leads to wild oscillations in the graph of g. It is an instructive exercise to plot the function $g(x) = \sin\left(\frac{\pi}{x}\right)$ with a graphing utility and then zoom in on $x = 0$. Doing so shows that the function never settles down to a single value near the origin and suggests that g does not have a limit at $x = 0$.

Activity 1.1–1

Consider the function $f(x) = \dfrac{x^2 - 1}{x - 1}$.

(a) Does the value of $f(1)$ exist? Why/why not?

(b) Use a calculator or spreadsheet to fill in the blanks of Table 1.3. Some of these have already been done for you; your calculations should not be drastically different from the ones already here. For example a calculation equal to 22.5 is very likely incorrect. Based on the table, estimate the value of $\lim_{x\to 1} \dfrac{x^2 - 1}{x - 1}$.

(c) Plot an accurate graph of the function f. It should be a straight line. Using your plot, what is $f(1)$? What is $\lim_{x\to 1} f(x)$?

(d) Now use algebra to simplify the fraction $\dfrac{x^2 - 1}{x - 1}$. Call this new function $g(x)$. What is $\lim_{x\to 1} g(x)$? How is this limit related to the limit of $f(x)$ as x approaches 1?

x	$f(x)$	x	$f(x)$
0.5	1.5	1.5	
0.9		1.1	2.1
0.99	1.99	1.01	
0.999		1.001	2.001

Table 1.3: Tabe of values near $x = 1$.

An important lesson to take from Example 2 is that both graphs and tables can be misleading when determining the value of a limit. While a table of values is useful for investigating the possible value of a limit, we should also use other tools to confirm the value, if we think the table suggests the limit exists.

Identifying When Limits Do Not Exist

As we have seen in Example 2, a function may not have a limit for all values of x. That is, we cannot say $\lim_{x\to a} f(x) = L$ for some numbers L for all values of a, for there may not be a number that $f(x)$ is approaching. There are three ways in which a limit may fail to exist.

1. The function $f(x)$ may approach different values on either side of a.

2. The function may grow without upper or lower bound as x approaches a.

3. The function may oscillate as x approaches a, which we saw in Example 2.

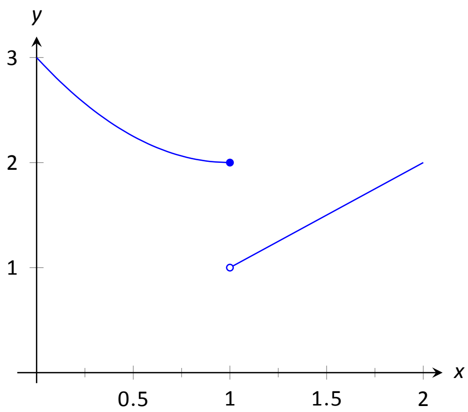


Figure 1.4: Observing no limit as $x \to 1$.

Example 3

Explore why $\lim_{x\to 1} f(x)$ does not exist, where

$$f(x) = \begin{cases} x^2 - 2x + 3 & x \leq 1 \\ x & x > 1 \end{cases}.$$

Solution. A graph of $f(x)$ around $x = 1$ is given in Figure 1.4. It is clear that as x approaches 1, $f(x)$ does not approach a single number. Instead, $f(x)$ approaches two different numbers. When considering values of x less than 1 (approaching 1 from the left), $f(x)$ is approaching 2; when considering values of x greater than 1 (approaching 1 from the right), $f(x)$ is approaching 1. Recognizing this behavior is important; we'll study this in greater depth later. Right now, it suffices to say that the limit does not exist since $f(x)$ is not approaching a single value as x approaches 1.

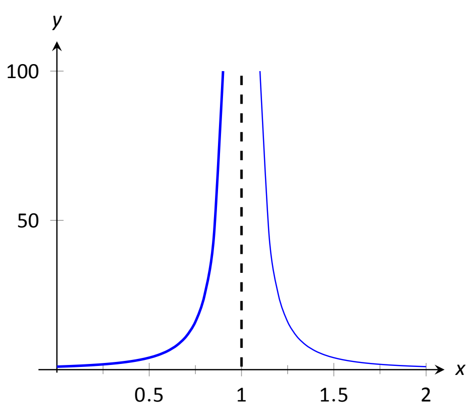


Figure 1.5: Observing no limit as $x \to 1$.

Example 4

Explore why $\lim_{x\to 1} \frac{1}{(x-1)^2}$ does not exist.

Solution. A graph of $f(x) = \frac{1}{(x-1)^2}$ is given in Figure 1.5. It shows that as x approaches 1, $f(x)$ grows without bound in the positive direction.

We can deduce this on our own, without the aid of the graph or table. If x is near 1, then $(x-1)^2$ is very small, and:

$$\frac{1}{\text{very small number}} \to \text{very large number}.$$

Since $f(x)$ is not approaching a single number (growing without bound), we conclude that

$$\lim_{x\to 1} \frac{1}{(x-1)^2}$$

does not exist.

Computing Limits

Thus far, our method of finding a limit is making a really good approximation either graphically or numerically. Later, we will cover the precise method of ϵ–δ proofs[2] which can be quite cumbersome. Now, we will cover a series of laws which allow us to find limits analytically.

Suppose that $\lim_{x\to 2} f(x) = 2$ and $\lim_{x\to 2} g(x) = 3$. What is $\lim_{x\to 2}(f(x) + g(x))$? Intuition tells us that the limit should be 5, as we expect limits to behave in a nice way. The following concept states that already established limits do behave nicely.

[2] See Section 1.4 for more on the precise definition of a limit.

Limit Laws

Let b, c, L and K be real numbers, let n be a positive integer, and let f and g be functions with the following limits:

$$\lim_{x\to c} f(x) = L \text{ and } \lim_{x\to c} g(x) = K.$$

The following limits hold.

1. Constants: $\lim_{x\to c} b = b$
2. Identity: $\lim_{x\to c} x = c$
3. Sums/Differences: $\lim_{x\to c}(f(x) \pm g(x)) = L \pm K$
4. Scalar Multiples: $\lim_{x\to c} b \cdot f(x) = bL$
5. Products: $\lim_{x\to c} f(x) \cdot g(x) = LK$
6. Quotients: $\lim_{x\to c} f(x)/g(x) = L/K$, $(K \neq 0)$
7. Powers: $\lim_{x\to c} f(x)^n = L^n$
8. Roots: $\lim_{x\to c} \sqrt[n]{f(x)} = \sqrt[n]{L}$ (if n is even then L must be greater than 0.)
9. Compositions: Adjust our previously given limit situation to:

   $$\lim_{x\to c} f(x) = L \text{ and } \lim_{x\to L} g(x) = K.$$

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

Example 5

Let $\lim_{x\to 2} f(x) = 2$, $\lim_{x\to 2} g(x) = 3$ and $p(x) = 3x^2 - 5x + 7$. Find the following limits:

1. $\lim_{x\to 2} \big(f(x) + g(x)\big)$
2. $\lim_{x\to 2} \big(5f(x) + g(x)^2\big)$
3. $\lim_{x\to 2} p(x)$

Solution.

1. Using the Sum/Difference rule, we know that $\lim_{x\to 2} \big(f(x) + g(x)\big) = 2 + 3 = 5$.
2. Using the Scalar Multiple and Sum/Difference rules, we find that $\lim_{x\to 2} \big(5f(x) + g(x)^2\big) = 5 \cdot 2 + 3^2 = 19$.
3. Here we combine the Power, Scalar Multiple, Sum/Difference and Constant Rules. We show quite a few steps, but in general these can

be omitted:

$$\begin{aligned}\lim_{x\to 2} p(x) &= \lim_{x\to 2}(3x^2-5x+7)\\ &= \lim_{x\to 2} 3x^2 - \lim_{x\to 2} 5x + \lim_{x\to 2} 7\\ &= 3\cdot 2^2 - 5\cdot 2 + 7\\ &= 9\end{aligned}$$

Notice that in the last part of the previous example, we have

$$\lim_{x\to 2} p(x) = p(2) = 9,$$

meaning that we could have evaluated the limit simply by directly substituting the value of a into the function. That is possible because the function $p(x)$ is *continuous*[3], which means that the limit of the function at any point is equal to its function value.

[3] See Section 1.3 for more on the notion of continuity.

From Example 1, we can evaluate the limit $\lim_{x\to -1} \frac{4-x^2}{x+2}$ by evaluating the function at $x=-1$:

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

Or from Example 2, we can evaluate the limit $\lim_{x\to 3} \sin\left(\frac{\pi}{x}\right)$ by evaluating the function at $x=3$:

$$\lim_{x\to 3} \sin\left(\frac{\pi}{x}\right) = \sin\left(\frac{\pi}{3}\right) = \frac{\sqrt{3}}{2}.$$

This method is quite important and is stated again for emphasis.

Direct Substitution

If f is a continuous function and a is in the domain of f, then

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

For $f(x)$ in Example 1, the situation is more complicated when $x \to -2$ since $f(-2)$ is not defined. If we attempt to evaluate the limit using direct substitution, we observe that

$$\lim_{x\to -2} \frac{4-x^2}{x+2} = \frac{4-(-2)^2}{-2+2} = \frac{4-4}{-2+2} = \frac{0}{0}.$$

We call $0/0$ an *indeterminate form*, or mathematical gibberish, and it tells us nothing about the limit—it could exist, it could not exist, we just don't know. We simply observe that it tells us there is somehow more work to do. From the table and graph of Example 1, it appears that f should have a limit of 4 at $x = -2$. To see algebraically why this is the case, let's work directly with the form of $f(x)$. Observe that

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

At this point, it is important to observe that since we are taking the limit as $x \to -2$, we are considering x values that are close, but not equal, to -2. Since we never actually allow x to equal -2, the quotient $\frac{2+x}{x+2}$ has value 1 for every possible value of x. Thus, we can simplify the most recent expression above, and now find that

$$\lim_{x\to -2} \frac{4-x^2}{x+2} = \lim_{x\to -2} \frac{(2-x)(2+x)}{x+2} = \lim_{x\to -2} (2-x).$$

Because $2-x$ is simply a linear function, this limit is now easy to determine, and its value clearly is 4. Thus, from several points of view we've seen that $\lim_{x\to -2} f(x) = 4$.

Activity 1.1–2

Determine the exact value of the limit by using algebra to simplify the function.

(a) $\displaystyle\lim_{x\to -5} \frac{x^2-25}{x+5}$

(b) $\displaystyle\lim_{x\to 0} \frac{(2+x)^3-8}{x}$

(c) $\displaystyle\lim_{x\to 0} \frac{\sqrt{x+1}-1}{x}$

The section could have been titled "Using Known Limits to Find Unknown Limits." By knowing certain limits of functions, we can find limits involving sums, products, powers, etc., of these functions. We further the development of such comparative tools with the Squeeze Theorem, a clever and intuitive way to find the value of some limits.

Suppose we have functions f, g and h where g always takes on values between f and h; that is, for all x in an interval,

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

If f and h have the same limit at c, and g is always "squeezed" between them, then g must have the same limit as well. That is what the Squeeze Theorem states.

Squeeze Theorem

Let f, g and h be functions on an open interval I containing c such that for all x in I, $f(x) \leq g(x) \leq h(x)$. If

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

then

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

It can take some work to figure out appropriate functions by which to "squeeze" the given function you are trying to evaluate a limit of. However, that is generally the only place work is necessary; the theorem makes the "evaluating the limit part" very simple.

Example 6

Given $4x - 9 \leq f(x) \leq x^2 - 4x + 7$, use the Squeeze Theorem to find $\lim\limits_{x \to 4} f(x)$.

Solution. We begin by finding $\lim\limits_{x \to 4}(4x - 9)$ and $\lim\limits_{x \to 4}(x^2 - 4x + 7)$.

$$\lim_{x \to 4}(4x - 9) = 4(4) - 9 = 7$$

and

$$\lim_{x \to 4}(x^2 - 4x + 7) = (4)^2 - 4(4) + 7 = 7.$$

Therefore,

$$\lim_{x \to 4}(4x - 9) = 7 = \lim_{x \to 4}(x^2 - 4x + 7),$$

and by the Squeeze Theorem, $\lim\limits_{x \to 4} f(x) = 7$.

Activity 1.1–3

In this activity, we prove $\lim\limits_{\theta \to 0} \dfrac{\sin(\theta)}{\theta} = 1$ using the squeeze theorem.

(a) Draw the part of the unit circle that lies in the first quadrant. Place a point on the circle in the first quadrant and label it (x, y).

(b) Draw the line connecting the origin to the point (x, y) on the same graph. Label the angle between the line and the $x - axis$ as θ.

(c) Find the area of the sector of the unit circle subtended by θ that you drew above.

(d) Draw a vertical line from the point (x, y) to the x-axis. This will form a right triangle.

(e) Find the area of the triangle in ((d)) in terms of θ. You will need to use trig functions.

(f) Draw a line on the picture connecting the points $(1, 0)$ and $(1, 1)$. This will also form a right triangle.

(g) Find the area of the triangle in ((f)). You will need to use trig functions.

(h) What is the relationship between the three areas you found? Write this as an inequality.

(i) Write this inequality so that it involves the function $\frac{\sin(\theta)}{\theta}$.

(j) Now apply the squeeze theorem to get the result.

One-Sided Limits

In this section, we explored the three ways in which limits of functions failed to exist:

1. The function approached different values from the left and right,
2. The function grows without bound, and
3. The function oscillates.

Now we explore in depth the concepts behind item #1 by introducing the *one-sided limit*. We begin with definitions that are very similar to the definition of the limit given earlier, but the notation is slightly different.

One-Sided Limits

Left-Sided Limit: Let f be a function defined on an open interval containing a. The notation

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

is read as "the limit of $f(x)$ as x approaches a from the left is L," or "the *left-sided limit of f at a is L*".

Right-Sided Limit: Let f be a function defined on an open interval containing a. The notation

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

is read as "the limit of $f(x)$ as x approaches a from the right is L," or "the *right-sided limit of f at a is L*".

Practically speaking, when evaluating a left-hand limit, we consider only values of x "to the left of a," i.e., where $x < a$. The admittedly imperfect notation $x \to a^-$ is used to imply that we look at values of x to the left of a. The notation has nothing to do with positive or negative values of either x or a.

A similar statement holds for evaluating right-hand limits; there we consider only values of x to the right of a, i.e., $x > a$. We can use the Limit Laws given earlier to help us evaluate these limits; we just restrict our view to one side of a.

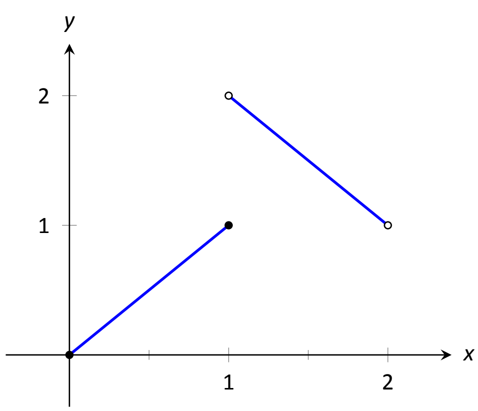


Figure 1.6: The graph of f.

Example 7

Let $f(x) = \begin{cases} x & 0 \le x \le 1 \\ 3 - x & 1 < x < 2 \end{cases}$, as shown in Figure 1.6. Find each of the following:

1. $\lim\limits_{x\to 1^-} f(x)$
2. $\lim\limits_{x\to 1^+} f(x)$
3. $\lim\limits_{x\to 1} f(x)$
4. $f(1)$
5. $\lim\limits_{x\to 0^+} f(x)$
6. $f(0)$
7. $\lim\limits_{x\to 2^-} f(x)$
8. $f(2)$

Solution. For these problems, the visual aid of the graph is likely more effective in evaluating the limits than using f itself. Therefore we will refer often to the graph.

1. As x goes to 1 *from the left*, we see that $f(x)$ is approaching the value of 1. Therefore $\lim\limits_{x\to 1^-} f(x) = 1$.
2. As x goes to 1 *from the right*, we see that $f(x)$ is approaching the value of 2. Recall that it does not matter that there is an "open circle" there; we are evaluating a limit, not the value of the function. Therefore $\lim\limits_{x\to 1^+} f(x) = 2$.
3. The limit of f as x approaches 1 does not exist, as discussed earlier. The function does not approach one particular value, but two different values from the left and the right.
4. Using the definition and by looking at the graph we see that $f(1) = 1$.
5. As x goes to 0 from the right, we see that $f(x)$ is also approaching 0. Therefore $\lim\limits_{x\to 0^+} f(x) = 0$. Note we cannot consider a left-hand limit at 0 as f is not defined for values of $x < 0$.

6. Using the definition and the graph, $f(0) = 0$.
7. As x goes to 2 from the left, we see that $f(x)$ is approaching the value of 1. Therefore $\lim\limits_{x\to 2^-} f(x) = 1$.
8. The graph and the definition of the function show that $f(2)$ is not defined.

Note how the left and right-hand limits were different; this, of course, causes the two-sided limit to not exist. The following theorem states what is fairly intuitive: the two-sided limit exists precisely when the left and right-hand limits are equal.

Limits and One Sided Limits

Let f be a function defined on an open interval I containing a. Then

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

if, and only if,

$$\lim_{x\to a^-} f(x) = L \quad \text{and} \quad \lim_{x\to a^+} f(x) = L.$$

The phrase "if, and only if" means the two statements are *equivalent*: they are either both true or both false. If the limit equals L, then the left and right hand limits both equal L. If the limit is not equal to L, then at least one of the left and right-hand limits is not equal to L (it may not even exist).

One thing to consider is that the value of the function may/may not be equal to the value(s) of its left/right-hand limits, even when these limits agree.

Example 8

Let $f(x) = \begin{cases} (x-1)^2 & 0 \le x \le 2, x \ne 1 \\ 1 & x = 1 \end{cases}$, as shown in Figure 1.7. Evaluate the following.

1. $\lim\limits_{x\to 1^-} f(x)$
2. $\lim\limits_{x\to 1^+} f(x)$
3. $\lim\limits_{x\to 1} f(x)$
4. $f(1)$

Solution. It is clear by looking at the graph that both the left and right-hand limits of f, as x approaches 1, is 0. Thus it is also clear that the limit is 0; i.e., $\lim\limits_{x\to 1} f(x) = 0$. It is also clearly stated that $f(1) = 1$.

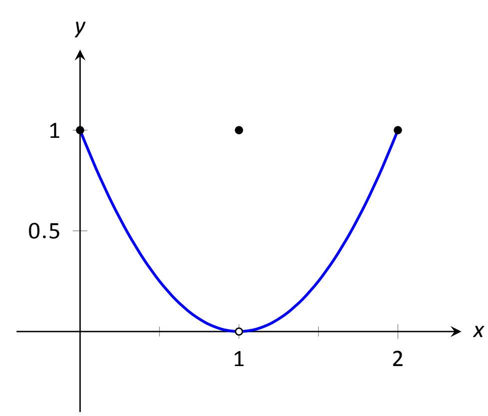


Figure 1.7: The graph of f.

This concludes a rather lengthy introduction to the notion of limits. It is important to remember that our primary motivation for considering limits of functions comes from our interest in studying the rate of change of a function. To that end, we close this section by revisiting our previous work with average and instantaneous velocity and highlighting the role that limits play.

Instantaneous Velocity

Suppose that we have a moving object whose position at time t is given by a function s. We know that the average velocity of the object on the time interval $[a, b]$ is $AV_{[a,b]} = \frac{s(b)-s(a)}{b-a}$. We define the *instantaneous velocity* at a to be the limit of average velocity as b approaches a. Note particularly that as $b \to a$, the length of the time interval gets shorter and shorter (while always including a). In Section **??**, we will introduce a helpful shorthand notation to represent the instantaneous rate of change. For now, we will write $IV_{t=a}$ for the instantaneous velocity at $t = a$, and thus

$$IV_{t=a} = \lim_{b \to a} AV_{[a,b]} = \lim_{b \to a} \frac{s(b) - s(a)}{b - a}.$$

Equivalently, if we think of the changing value b as being of the form $b = a + h$, where h is some small number, then we may instead write

$$IV_{t=a} = \lim_{h \to 0} AV_{[a,a+h]} = \lim_{h \to 0} \frac{s(a+h) - s(a)}{h}.$$

Again, the most important idea here is that to compute instantaneous velocity, we take a limit of average velocities as the time interval shrinks. Two different activities offer the opportunity to investigate these ideas and the role of limits further.

Activity 1.1–4

Consider a moving object whose position function is given by $s(t) = t^2$, where s is measured in meters and t is measured in minutes.

(a) Determine the most simplified expression for the average velocity of the object on the interval $[3, 3 + h]$, where $h > 0$.

(b) Determine the average velocity of the object on the interval $[3, 3.2]$. Include units on your answer.

(c) Determine the instantaneous velocity of the object when $t = 3$. Include units on your answer.

The closing activity of this section asks you to make some connections among average velocity, instantaneous velocity, and slopes of certain lines.

Activity 1.1–5

For the moving object whose position s at time t is given by the graph below, answer each of the following questions. Assume that s is measured in feet and t is measured in seconds.

(a) Use the graph to estimate the average velocity of the object on each of the following intervals: $[0.5, 1]$, $[1.5, 2.5]$, $[0, 5]$. Draw each line whose slope represents the average velocity you seek.

(b) How could you use average velocities or slopes of lines to estimate the instantaneous velocity of the object at a fixed time?

(c) Use the graph to estimate the instantaneous velocity of the object when $t = 2$. Should this instantaneous velocity at $t = 2$ be greater or less than the average velocity on $[1.5, 2.5]$ that you computed in (a)? Why?

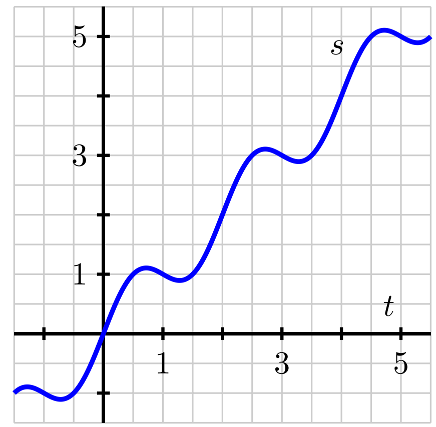


Figure 1.8: Plot of the position function $y = s(t)$ in Activity 1.1–5.

Summary

In this section, we encountered the following important ideas:

- Limits enable us to examine trends in function behavior near a specific point. In particular, taking a limit at a given point asks if the function values nearby tend to approach a particular fixed value.
- When we write $\lim_{x \to a} f(x) = L$, we read this as saying "the limit of f as x approaches a is L," and this means that we can make the value of $f(x)$ as close to L as we want by taking x sufficiently close (but not equal) to a.
- If we desire to know $\lim_{x \to a} f(x)$ for a given value of a and a known function f, we can estimate this value from the graph of f or by generating a table of function values that result from a sequence of x-values that are closer and closer to a. If we want the exact value of the limit, we need to work with the function algebraically and see if we can use familiar properties of known, basic functions to understand how different parts of the formula for f change as $x \to a$.
- The instantaneous velocity of a moving object at a fixed time is found by taking the limit of average velocities of the object over shorter and shorter time intervals that all contain the time of interest.

Exercises

Terms and Concepts

1) In your own words, what does it mean to "find the limit of $f(x)$ as x approaches 3"?

2) An expression of the form $\frac{0}{0}$ is called ___.

3) T/F: The limit of $f(x)$ as x approaches 5 is always $f(5)$.

4) T/F: If $\lim\limits_{x\to 1^-} f(x) = 5$, then $\lim\limits_{x\to 1} f(x) = 5$.

5) T/F: If $\lim\limits_{x\to 1^-} f(x) = 5$, then $\lim\limits_{x\to 1^+} f(x) = 5$.

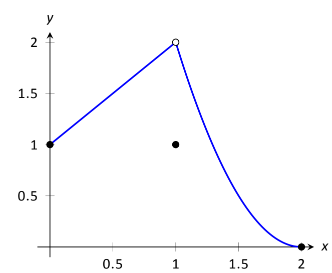
6) T/F: If $\lim\limits_{x\to 1} f(x) = 5$, then $\lim\limits_{x\to 1^-} f(x) = 5$.

7) Describe three situations where $\lim\limits_{x\to c} f(x)$ does not exist.

Problems

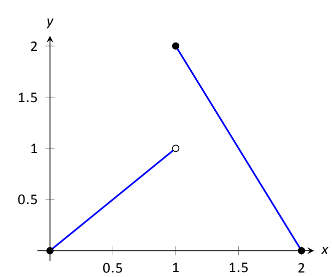
In the following exercises, evaluate each expression using the given graph of $f(x)$.

8)



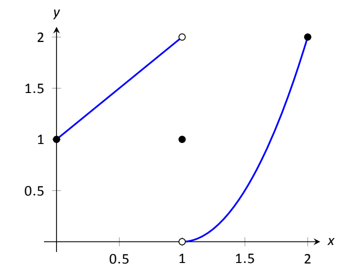
(a) $\lim\limits_{x\to 1^-} f(x)$
(b) $\lim\limits_{x\to 1^+} f(x)$
(c) $\lim\limits_{x\to 1} f(x)$
(d) $f(1)$
(e) $\lim\limits_{x\to 0^-} f(x)$
(f) $\lim\limits_{x\to 0^+} f(x)$

9)



(a) $\lim\limits_{x\to 1^-} f(x)$
(b) $\lim\limits_{x\to 1^+} f(x)$
(c) $\lim\limits_{x\to 1} f(x)$
(d) $f(1)$
(e) $\lim\limits_{x\to 2^-} f(x)$
(f) $\lim\limits_{x\to 2^+} f(x)$

10)



(a) $\lim\limits_{x\to 1^-} f(x)$
(b) $\lim\limits_{x\to 1^+} f(x)$
(c) $\lim\limits_{x\to 1} f(x)$
(d) $f(1)$

Using:

$\lim\limits_{x\to 9} f(x) = 6$ $\quad$ $\lim\limits_{x\to 6} f(x) = 9$

$\lim\limits_{x\to 9} g(x) = 3$ $\quad$ $\lim\limits_{x\to 6} g(x) = 3$

evaluate the limits given in the following exercises, where possible. If it is not possible to know, state so.

11) $\lim\limits_{x\to 9}\big(f(x)+g(x)\big)$

12) $\lim\limits_{x\to 9}\big(3f(x)/g(x)\big)$

13) $\lim\limits_{x\to 9}\left(\dfrac{f(x)-2g(x)}{g(x)}\right)$

14) $\lim\limits_{x\to 6}\left(\dfrac{f(x)}{3-g(x)}\right)$

15) $\lim\limits_{x\to 9} g\big(f(x)\big)$

16) $\lim\limits_{x\to 6} f\big(g(x)\big)$

17) $\lim\limits_{x\to 6} g\big(f(f(x))\big)$

18) $\lim\limits_{x\to 6} f(x)g(x) - f^2(x) + g^2(x)$

Using:

$\lim\limits_{x\to 1} f(x) = 2$ $\quad$ $\lim\limits_{x\to 10} f(x) = 1$

$\lim\limits_{x\to 1} g(x) = 0$ $\quad$ $\lim\limits_{x\to 10} g(x) = \pi$

evaluate the limits given in the following exercises, where possible. If it is not possible to know, state so.

19) $\lim\limits_{x\to 1} f(x)^{g(x)}$

20) $\lim\limits_{x\to 10} \cos\big(g(x)\big)$

21) $\lim\limits_{x\to 1} f(x)g(x)$

22) $\lim\limits_{x\to 1} g\big(5f(x)\big)$

In the following exercises, evaluate the limit.

23) $\lim\limits_{x\to 3} x^2-3x+7$

24) $\lim\limits_{x\to \pi} \dfrac{3x+1}{1-x}$

25) $\lim\limits_{x\to \pi} \dfrac{x^2+3x+5}{5x^2-2x-3}$

26) $\lim\limits_{x\to \pi} \left(\dfrac{x-3}{x-5}\right)^7$

27) $\lim\limits_{x\to \pi/4} \cos x \sin x$

28) $\lim\limits_{x\to 0} \ln x$

29) $\lim_{x\to 3} 4^{x^3-8x}$

30) $\lim_{x\to \pi/6} \csc x$

31) $\lim_{x\to 0} \ln(1+x)$

32) $\lim_{x\to 6^-} \dfrac{x^2-4x-12}{x^2-13x+42}$

33) $\lim_{x\to 0^+} \dfrac{x^2+2x}{x^2-2x}$

34) $\lim_{x\to 2^-} \dfrac{x^2+6x-16}{x^2-3x+2}$

35) $\lim_{x\to 2^+} \dfrac{x^2-10x+16}{x^2-x-2}$

36) $\lim_{x\to -2} \dfrac{x^2-5x-14}{x^2+10x+16}$

37) $\lim_{x\to -1} \dfrac{x^2+9x+8}{x^2-6x-7}$

38) $\lim_{x\to 1} \dfrac{x-1}{\sqrt{x}-1}$

39) $\lim_{x\to 0} \dfrac{\sqrt{x+4}-2}{x}$

40) $\lim_{x\to 0} \dfrac{\sqrt{16+x}-4}{x}$

Use the Squeeze Theorem, where appropriate, to evaluate the following limits.

41) $\lim_{x\to 0} x \sin\left(\dfrac{1}{x}\right)$

42) $\lim_{x\to 0} \sin x \cos\left(\dfrac{1}{x^2}\right)$

43) $\lim_{x\to 3} f(x)$, where $x^2 \le f(x) \le 3x$ on $[0,3]$.

Using the techniques you learned in Activity 1.1–4 and Activity 1.1–5, determine the instantaneous velocity of some object with given position function and specific instant.

44) $s(t) = -7t+2; \quad t=3$

45) $s(t) = 9t+0.06; \quad t=1$

46) $s(t) = t^2+3t-7; \quad t=1$

47) $s(t) = \dfrac{1}{t+1}; \quad t=2$

1.2 *Limits involving infinity*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- What does it mean to say that $\lim_{x \to \infty} f(x) = L$ and $\lim_{x \to a} f(x) = \infty$?
- What is the connection between limits involving infinity and the asymptotes of a function?

Introduction

The concept of infinity, denoted ∞, arises naturally in calculus, like it does in much of mathematics. It is important to note from the outset that ∞ is a concept, but not a number itself. Indeed, the notion of ∞ naturally invokes the idea of limits. Consider, for example, the function $f(x) = \frac{1}{x}$, whose graph is pictured in Figure 1.9.

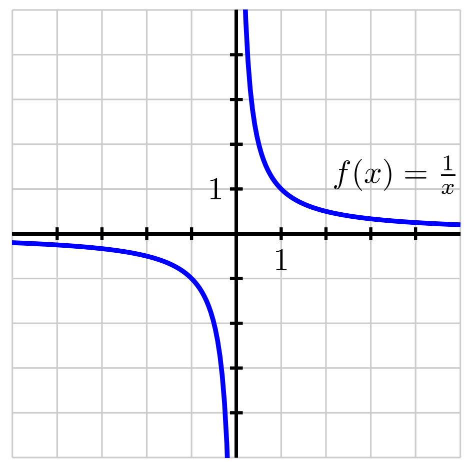


Figure 1.9: The graph of $f(x) = \frac{1}{x}$.

We note that $x = 0$ is not in the domain of f, so we may naturally wonder what happens as $x \to 0$. As $x \to 0^+$, we observe that $f(x)$ *increases without bound*. That is, we can make the value of $f(x)$ as large as we like by taking x closer and closer (but not equal) to 0, while keeping $x > 0$. This is a good way to think about what infinity represents: a quantity is tending to infinity if there is no single number that the quantity is always less than.

Recall that when we write $\lim_{x \to a} f(x) = L$, this means that we can make $f(x)$ as close to L as we'd like by taking x sufficiently close (but not equal) to a. We thus expand this notation and language to include the possibility that either L or a can be ∞. For instance, for $f(x) = \frac{1}{x}$, we now write

$$\lim_{x \to 0^+} \frac{1}{x} = \infty,$$

by which we mean that we can make $\frac{1}{x}$ as large as we like by taking x sufficiently close (but not equal) to 0. When we write the limit is equal to ∞ or $-\infty$, we call that limit an *infinite limit*.

Note that we are not saying the limit exists. It doesn't since the function does not approach a single, unique real number as $x \to 0^+$. We just give a special designation for a limit that does not exist because the function grows without bound in the positive or negative direction.

In a similar way, we naturally write

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

since we can make $\frac{1}{x}$ as close to 0 as we'd like by taking x sufficiently large (i.e., by letting x increase without bound). This limit does exist because 0 is a single, unique real number.

In general, we understand the notation $\lim_{x \to a} f(x) = \infty$ to mean that we can make $f(x)$ as large as we'd like by taking x sufficiently close (but not equal) to a, and the notation $\lim_{x \to \infty} f(x) = L$ to mean that we can make $f(x)$ as close to L as we'd like by taking x sufficiently large. This notation applies to left- and right-hand limits, plus we can also use limits involving $-\infty$. For example, returning to Figure 1.9 and $f(x) = \frac{1}{x}$, we can say that

$$\lim_{x \to 0^-} \frac{1}{x} = -\infty \quad \text{and} \quad \lim_{x \to -\infty} \frac{1}{x} = 0.$$

Finally, we write

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

when we can make the value of $f(x)$ as large as we'd like by taking x sufficiently large. For example,

$$\lim_{x \to \infty} x^2 = \infty.$$

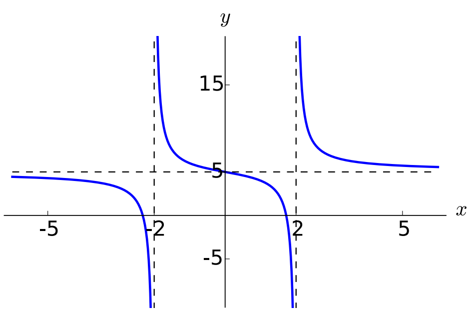


Figure 1.10: The graph of $y = f(x)$.

Preview Activity 1.2

Using the graph of f in Figure 1.10, evaluate the following, being sure to use any special designations if needed.

(a) $\lim_{x \to -\infty} f(x)$

(b) $\lim_{x \to -2^-} f(x)$

(c) $\lim_{x \to -2^+} f(x)$

(d) $\lim_{x \to -2} f(x)$

(e) $f(0)$

(f) $\lim_{x \to 0} f(x)$

(g) $\lim_{x \to 2^-} f(x)$

(h) $\lim_{x \to 2^+} f(x)$

(i) $\lim_{x \to 2} f(x)$

(j) $\lim_{x \to \infty} f(x)$

Note particularly that limits involving infinity identify *vertical* and *horizontal asymptotes* of a function. If $\lim_{x \to a} f(x) = \infty$, then $x = a$ is a vertical asymptote of f, while if $\lim_{x \to \infty} f(x) = L$, then $y = L$ is a horizontal asymptote of f. Similar statements can be made using $-\infty$, as well as with left- and right-hand limits as $x \to a^-$ or $x \to a^+$.

Infinite Limits

We can easily determine if $\lim_{x \to a} f(x)$ is an infinite limit by observing the graph of $f(x)$. But what if we have no graph? How do we determine if $\lim_{x \to a} f(x)$ is an infinite limit or not?

In Section 1.1, we learned the method of Direct Substitution to evaluate limits. Now, we can use that method to determine if a limit is an infinite limit. Certainly, if

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

then we have an infinite limit. However, we may also have an infinite limit if

$$\lim_{x\to a} f(x) = f(a) = \frac{\text{a nonzero real number}}{0}.$$

Consider $f(x) = 1/x^2$ as shown in Figure 1.11 and $\lim_{x\to 0} f(x)$. Note that if evaluate $f(0)$, we have

$$f(0) = \frac{1}{0^2} = \frac{1}{0}.$$

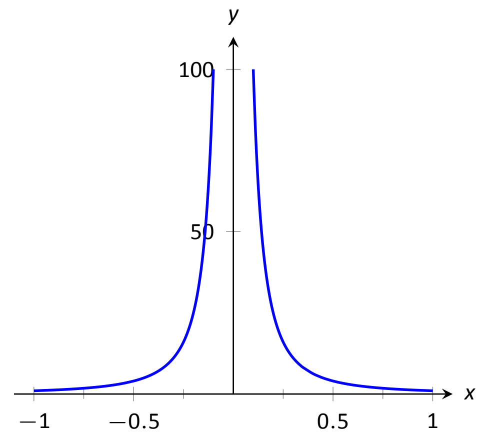


Figure 1.11: Graphing $f(x) = 1/x^2$ for values of x near 0.

But also observe in the graph how, as x approaches 0 from both the left and the right, $f(x)$ grows without bound. So we can state that

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

Consider $f(x) = 1/x$ as shown in Figure 1.9 one more time. Again, if we evaluate $f(0)$, then we have

$$f(0) = \frac{1}{0},$$

but observing the graph, we see that the function grows without bound in different directions from the left and right. Since that happens, we do not give a special designation to this limit and simply say that the limit as $x \to 0$ does not exist, i.e.,

$$\lim_{x\to 0} \frac{1}{x} = \text{ Does Not Exist.}$$

So we can determine if we have an infinite limit when by Direct Substitution, we have $f(a) = \frac{\text{a nonzero real number}}{0}$ and then comparing the individual left- and right-sided limits. If they're the same, then we give the appropriate special designation. If not, then we simply say the limit does not exist.

Example 1

Evaluate the possible infinite limit: $\lim_{x\to 1} \dfrac{1}{(x-1)^2}$.

Solution. By Direct Substitution, we have

$$\lim_{x\to 1} \frac{1}{(x-1)^2} = \frac{1}{(1-1)^2} = \frac{1}{0},$$

so now we need to compare the one-sided limits

$$\lim_{x\to 1^-} \frac{1}{(x-1)^2} \quad \text{and} \quad \lim_{x\to 1^+} \frac{1}{(x-1)^2}.$$

As $x \to 1^-$, the quantity $(x-1)^2 \to 0$ and is positive; therefore, $\lim\limits_{x\to 1^-} \dfrac{1}{(x-1)^2} = \infty$. As $x \to 1^+$, the quantity $(x-1)^2 \to 0$ and is positive; therefore, $\lim\limits_{x\to 1^+} \dfrac{1}{(x-1)^2} = \infty$, and we can say that

$$\lim_{x\to 1} \frac{1}{(x-1)^2} = \infty.$$

Example 2

Evaluate the possible infinite limit: $\lim\limits_{x\to 1} \dfrac{1}{(x-1)^3}$.

Solution. By Direct Substitution, we have

$$\lim_{x\to 1} \frac{1}{(x-1)^3} = \frac{1}{(1-1)^3} = \frac{1}{0},$$

so now we need to compare the one-sided limits

$$\lim_{x\to 1^-} \frac{1}{(x-1)^3} \quad \text{and} \quad \lim_{x\to 1^+} \frac{1}{(x-1)^3}.$$

As $x \to 1^-$, the quantity $(x-1)^3 \to 0$ and is negative; therefore, $\lim\limits_{x\to 1^-} \dfrac{1}{(x-1)^3} = -\infty$. As $x \to 1^+$, the quantity $(x-1)^3 \to 0$ and is positive; therefore, $\lim\limits_{x\to 1^+} \dfrac{1}{(x-1)^3} = \infty$, and we can say that

$$\lim_{x\to 1} \frac{1}{(x-1)^2} = \text{ Does not exist.}$$

Activity 1.2–1

Evaluate the following infinite limits analytically.

(a) i) $\lim\limits_{x\to 2^+} \dfrac{1}{x-2}$ ii) $\lim\limits_{x\to 2^-} \dfrac{1}{x-2}$ iii) $\lim\limits_{x\to 2} \dfrac{1}{x-2}$

(b) i) $\lim\limits_{x\to 3^+} \dfrac{x+2}{(x-3)^3}$ ii) $\lim\limits_{x\to 3^-} \dfrac{x+2}{(x-3)^3}$ iii) $\lim\limits_{x\to 3} \dfrac{x+2}{(x-3)^3}$

(c) i) $\lim\limits_{x\to 4^+} \dfrac{x-5}{(x-4)^2}$ ii) $\lim\limits_{x\to 4^-} \dfrac{x-5}{(x-4)^2}$ iii) $\lim\limits_{x\to 4} \dfrac{x-5}{(x-4)^2}$

Limits at Infinity

In precalculus classes, it is common to study the *end behavior* of certain families of functions, by which we mean the behavior of a function as $x \to \infty$ and as $x \to -\infty$, or the limit of the function as $x \to \infty$ and as $x \to -\infty$.

For polynomial functions of the form

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

the limit at infinity, or end behavior, depends on the sign of a_n and whether the highest power n is even or odd. If n is even and a_n is positive, then $\lim\limits_{x\to\infty} p(x) = \infty$ and $\lim\limits_{x\to-\infty} p(x) = \infty$, as in the plot of g in Figure 1.12. If instead a_n is negative, then $\lim\limits_{x\to\infty} p(x) = -\infty$ and $\lim\limits_{x\to-\infty} p(x) = -\infty$. In the situation where n is odd, then either $\lim\limits_{x\to\infty} p(x) = \infty$ and $\lim\limits_{x\to-\infty} p(x) = \infty$ (which occurs when a_n is positive, as in the graph of f in Figure 1.12), or $\lim\limits_{x\to\infty} p(x) = \infty$ and $\lim\limits_{x\to-\infty} p(x) = \infty$ (when a_n is negative).

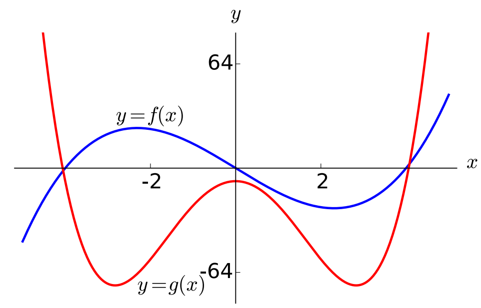


Figure 1.12: The graphs of $f(x) = x^3 - 16x$ and $g(x) = x^4 - 16x^2 - 8$.

A function can fail to have a limit as $x \to \infty$. For example, consider the plot of the sine function at right in Figure 1.13. Because the function continues oscillating between -1 and 1 as $x \to \infty$, we say that $\lim\limits_{x\to\infty} \sin(x)$ does not exist.

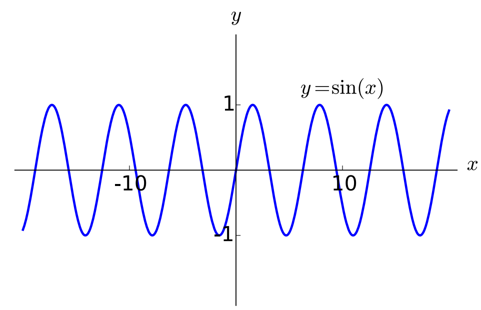


Figure 1.13: The graph of $f(x) = \sin(x)$.

Finally, it is straightforward to analyze the behavior of any rational function as $x \to \infty$. Consider, for example, the function

$$q(x) = \frac{3x^2 - 4x + 5}{7x^2 + 9x - 10}.$$

Note that both $(3x^2 - 4x + 5) \to \infty$ as $x \to \infty$ and $(7x^2 + 9x - 10) \to \infty$ as $x \to \infty$. Here we say that $\lim\limits_{x\to\infty} q(x)$ has indeterminate form $\dfrac{\infty}{\infty}$, much like we did when we encountered limits of the form $\dfrac{0}{0}$. We can determine the value of this limit through a standard algebraic approach. Multiplying the numerator and denominator each by $\frac{1}{x^2}$, we find that

$$\begin{aligned}
\lim_{x\to\infty} q(x) &= \lim_{x\to\infty} \frac{3x^2 - 4x + 5}{7x^2 + 9x - 10} \cdot \frac{\frac{1}{x^2}}{\frac{1}{x^2}} \\
&= \lim_{x\to\infty} \frac{3 - 4\frac{1}{x} + 5\frac{1}{x^2}}{7 + 9\frac{1}{x} - 10\frac{1}{x^2}} \\
&= \frac{3}{7}
\end{aligned}$$

since $\frac{1}{x^2} \to 0$ and $\frac{1}{x} \to 0$ as $x \to \infty$. This shows that the rational function q has a horizontal asymptote at $y = \frac{3}{7}$. A similar approach can be used to determine the limit of any rational function as $x \to \infty$.

Activity 1.2–2

Evaluate $\lim_{x\to\infty}$ and $\lim_{x\to-\infty}$ for each of the rational functions.

(a) $f(x) = \dfrac{4x}{20x+1}$

(b) $f(x) = \dfrac{3x^3-7}{x^4+5x^2}$

(c) $f(x) = \dfrac{4x^2-8}{8x^2+5x+2}$

(d) $f(x) = \dfrac{-x^3+1}{2x+8}$

Summary

In this section, we encountered the following important ideas:

- With respect to the standard notation of a limit

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

  we can let express limits where either the value a or L grows without bound in the positive or negative directions.
- The notions of *vertical and horizontal asymptotoes* are directly connected to infinite limits and limits at infinity respectively.
- The *end behavior* of a function is directly connected to limits at infinity.

Exercises

Terms and Concepts

1) T/F: If $\lim_{x\to 5} f(x) = \infty$, then we are implicitly stating that the limit exists.

2) T/F: If $\lim_{x\to\infty} f(x) = 5$, then we are implicitly stating that the limit exists.

3) T/F: If $\lim_{x\to 1^-} f(x) = -\infty$, then $\lim_{x\to 1^+} f(x) = \infty$

4) T/F: If $\lim_{x\to 5} f(x) = \infty$, then f has a vertical asymptote at $x = 5$.

5) T/F: $\infty/0$ is not an indeterminate form.

6) Construct a function with a vertical asymptote at $x = 5$ and a horizontal asymptote at $y = 5$.

7) Let $\lim_{x\to 7} f(x) = \infty$. Explain how we know that f is/is not continuous at $x = 7$.

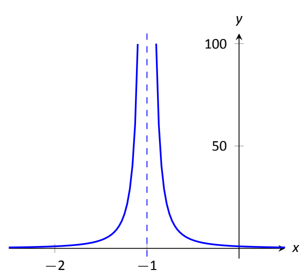
Problems

In exercises 8–13, evaluate each limit using the given graph of $f(x)$.

8) $f(x) = \dfrac{1}{(x+1)^2}$

(a) $\lim_{x\to -1^-} f(x)$

(b) $\lim_{x\to -1^+} f(x)$



9) $f(x) = \dfrac{1}{(x-3)(x-5)^2}$.

(a) $\lim_{x\to 3^-} f(x)$

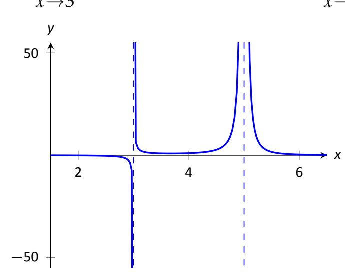
(b) $\lim_{x\to 3^+} f(x)$

(c) $\lim_{x\to 3} f(x)$

(d) $\lim_{x\to 5^-} f(x)$

(e) $\lim_{x\to 5^+} f(x)$

(f) $\lim_{x\to 5} f(x)$



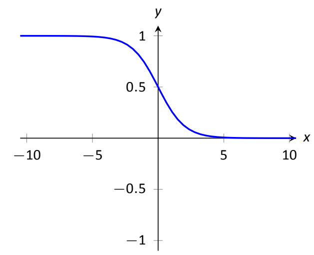
10) $f(x) = \dfrac{1}{e^x + 1}$

(a) $\lim_{x\to -\infty} f(x)$

(b) $\lim_{x\to \infty} f(x)$

(c) $\lim_{x\to 0^-} f(x)$

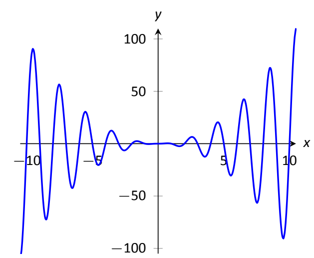
(d) $\lim_{x\to 0^+} f(x)$



11) $f(x) = x^2 \sin(\pi x)$

(a) $\lim_{x\to -\infty} f(x)$

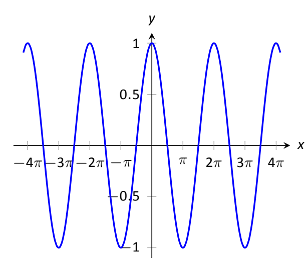
(b) $\lim_{x\to \infty} f(x)$



12) $f(x) = \cos(x)$

(a) $\lim_{x\to -\infty} f(x)$

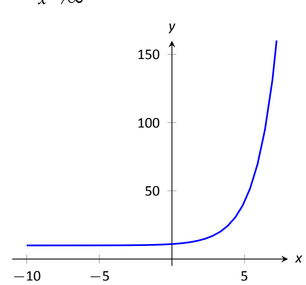
(b) $\lim_{x\to \infty} f(x)$



13) $f(x) = 2^x + 10$

(a) $\lim_{x\to -\infty} f(x)$

(b) $\lim_{x\to \infty} f(x)$



In exercises 14–24, evaluate the limit.

14) $\displaystyle\lim_{x\to -3^+} \frac{x-2}{x+3}$

15) $\displaystyle\lim_{x\to 5^-} \frac{4}{x-5}$

16) $\displaystyle\lim_{x\to 1} \frac{3-x}{(x-1)^2}$

17) $\displaystyle\lim_{x\to \pi^-} \cot(x)$

18) $\displaystyle\lim_{x\to \infty} \frac{x^3+2x^2+1}{5-x}$

19) $\displaystyle\lim_{x\to -\infty} \frac{x^3+2x^2+1}{x^2-5}$

20) $\displaystyle\lim_{x\to -\infty} \frac{x^3+2x^2+1}{5-x^2}$

21) $\displaystyle\lim_{x\to \infty} \left(x-\sqrt{x}\right)$

22) $\displaystyle\lim_{x\to \infty} \frac{x+2}{\sqrt{9x^2+8}}$

23) $\displaystyle\lim_{x\to -\infty} \frac{\sin^2(x)}{x^2}$

24) $\displaystyle\lim_{x\to -\infty} \left(x^4+x^5\right)$

1.3 *Continuity*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- What does it mean to say that a function f is continuous at $x = a$? What role do limits play in determining whether or not a function is continuous at a point?
- What properties do continuous functions have?
- How can we use continuous functions?

Introduction

In Section 1.1, we learned about how the concept of limits can be used to study the trend of a function near a fixed input value. As we study such trends, we are fundamentally interested in knowing how well-behaved the function is at the given point, say $x = a$. In this present section, we aim to expand our perspective and develop language and understanding to quantify how the function acts and how its value changes near a particular point. Beyond thinking about whether or not the function has a limit L at $x = a$, we will also consider the value of the function $f(a)$ and how this value is related to $\lim_{x \to a} f(x)$. Throughout, we will build on and formalize ideas that we have encountered in several settings.

We begin to consider these issues through the following preview activity that asks you to consider the graph of a function with a variety of interesting behaviors.

Preview Activity 1.3

A function f defined on $-4 < x < 4$ is given by the graph in Figure 1.14. Use the graph to answer each of the following questions. Note: to the right of $x = 2$, the graph of f is exhibiting infinite oscillatory behavior similar to the function $\sin(\frac{\pi}{x})$ that we encountered in the key example early in Section 1.1.

(a) For each of the values $a = -3, -2, -1, 0, 1, 2, 3$, determine whether or not $\lim_{x \to a} f(x)$ exists. If the function has a limit L at a given point, state the value of the limit using the notation $\lim_{x \to a} f(x) = L$. If the function does not have a limit at a given point, write a sentence to explain why.

(b) For each of the values of a from part (a) where f has a limit, determine the value of $f(a)$ at each such point.

(c) For each of the values $a = -3, -2, -1, 0, 1, 2, 3$, does $f(a)$ have the same value as $\lim_{x \to a} f(x)$?

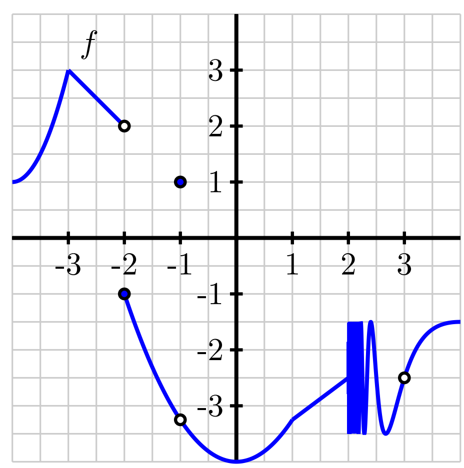


Figure 1.14: The graph of $y = f(x)$.

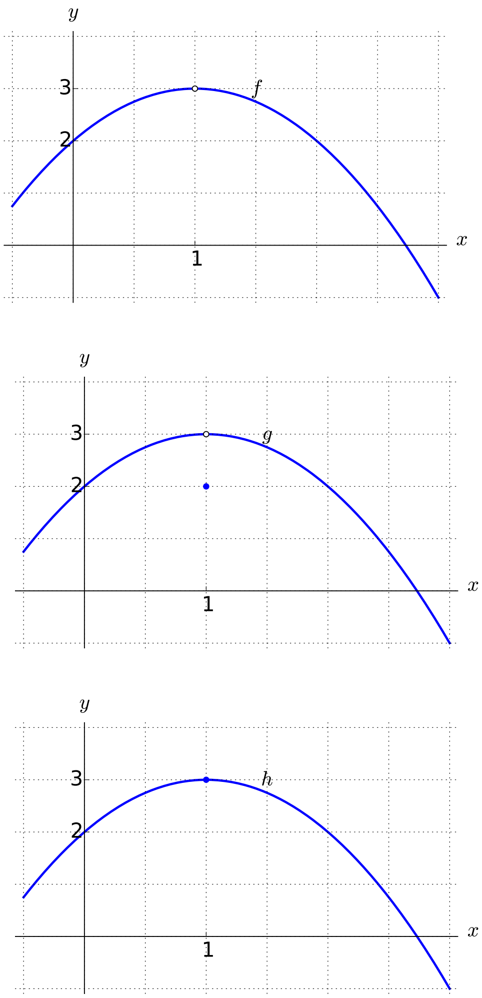


Figure 1.15: Functions f, g, and h that demonstrate subtly different behaviors at $a = 1$.

Being continuous at a point

Intuitively, a function is continuous if we can draw it without ever lifting our pencil from the page. Alternatively, we might say that the graph of a continuous function has no jumps or holes in it. We first consider three specific situations in Figure 1.15 where all three functions have a limit at $a = 1$, and then work to make the idea of continuity more precise.

Note that $f(1)$ is not defined, which leads to the resulting hole in the graph of f at $a = 1$. We will naturally say that f is *not continuous at* $a = 1$. For the next function g in in Figure 1.15, we observe that while $\lim_{x\to 1} g(x) = 3$, the value of $g(1) = 2$, and thus the limit does not equal the function value. Here, too, we will say that g is *not continuous*, even though the function is defined at $a = 1$. Finally, the function h appears to be the most well-behaved of all three, since at $a = 1$ its limit and its function value agree. That is,

$$\lim_{x\to 1} h(x) = 3 = h(1).$$

With no hole or jump in the graph of h at $a = 1$, we desire to say that h is *continuous* there.

More formally, we make the following definition.

Continuity

A function f is *continuous* provided that

- f has a limit as $x \to a$,
- f is defined at $x = a$, and
- $\lim_{x\to a} f(x) = f(a)$.

A function f is *continuous* on an open interval I if f is continuous at c for all c in I.

Conditions (a) and (b) are technically contained implicitly in (c), but we state them explicitly to emphasize their individual importance. In words, (c) essentially says that a function is continuous at $x = a$ provided that its limit as $x \to a$ exists and equals its function value at $x = a$. Thus, continuous functions are particularly nice: to evaluate the limit of a continuous function at a point, all we need to do is evaluate the function.

Example 1

Let f be defined as shown in Figure 1.16. Give the interval(s) on which f is continuous.

Solution. We proceed by examining the three criteria for continuity.

1. The limits $\lim\limits_{x\to a} f(x)$ exists for all a between 0 and 3.
2. $f(a)$ is defined for all a between 0 and 3, *except for* $a = 1$. We know immediately that f cannot be continuous at $x = 1$.
3. The limit $\lim\limits_{x\to a} f(x) = f(a)$ for all a between 0 and 3, except, of course, for $a = 1$.

We conclude that f is continuous at every point of $(0, 3)$ except at $x = 1$. Therefore f is continuous on $(0, 1)$ and $(1, 3)$.

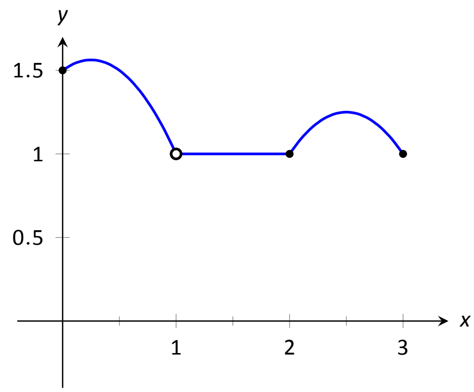


Figure 1.16: A graph of f in Example 1.

Let's now consider $p(x) = x^2 - 2x + 3$. It can be proved that every polynomial is a continuous function at every real number, and thus if we would like to know $\lim_{x\to 2} p(x)$, we simply compute

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

This route of substituting an input value to evaluate a limit works anytime we know function being considered is continuous. Besides polynomial functions, all exponential functions and the sine and cosine functions are continuous at every point, as are many other familiar functions and combinations thereof.

Example 2

For each of the following functions, give the domain of the function and the interval(s) on which it is continuous.

1. $f(x) = 1/x$
2. $f(x) = \sin(x)$
3. $f(x) = \sqrt{x}$
4. $f(x) = \sqrt{1 - x^2}$
5. $f(x) = |x|$

Solution. We examine each in turn.

1. The domain of $f(x) = 1/x$ is $(-\infty, 0) \cup (0, \infty)$. As it is a rational function, we apply the limit laws to recognize that f is continuous on all of its domain.
2. The domain of $f(x) = \sin(x)$ is all real numbers, or $(-\infty, \infty)$. Applying the limit laws shows that $\sin x$ is continuous everywhere.
3. The domain of $f(x) = \sqrt{x}$ is $[0, \infty)$. Applying the limit laws shows that $f(x) = \sqrt{x}$ is continuous on its domain of $[0, \infty)$.
4. The domain of $f(x) = \sqrt{1 - x^2}$ is $[-1, 1]$. Applying the limit laws shows that f is continuous on all of its domain, $[-1, 1]$.

5. The domain of $f(x) = |x|$ is $(-\infty, \infty)$. We can define the absolute value function as $f(x) = \begin{cases} -x & x < 0 \\ x & x \geq 0 \end{cases}$. Each "piece" of this piecewise defined function is continuous on all of its domain, giving that f is continuous on $(-\infty, 0)$ and $[0, \infty)$. As we saw before, we cannot assume this implies that f is continuous on $(-\infty, \infty)$; we need to check that $\lim_{x\to 0} f(x) = f(0)$, as $x = 0$ is the point where f transitions from one "piece" of its definition to the other. It is easy to verify that this is indeed true, hence we conclude that $f(x) = |x|$ is continuous everywhere.

Activity 1.3–1

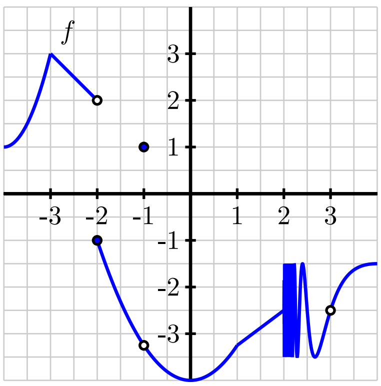


Figure 1.17: The graph of $y = f(x)$ for Activity 1.3–1.

This activity builds on your work in Preview Activity 1.3, using the same function f as given by the graph that is repeated in Figure 1.17.

(a) At which values of a does $\lim_{x\to a} f(x)$ not exist?

(b) At which values of a is $f(a)$ not defined?

(c) At which values of a does f have a limit, but $\lim_{x\to a} f(x) \neq f(a)$?

(d) State all values of a for which f is not continuous at $x = a$.

(e) Which condition is stronger, and hence implies the other: f has a limit at $x = a$ or f is continuous at $x = a$? Explain, and hence complete the following sentence: "If f ________________ at $x = a$, then f ________________ at $x = a$," where you complete the blanks with *has a limit* and *is continuous*, using each phrase once.

Activity 1.3–2

State the interval(s) on which each of the following functions is continuous.

(a) $f(x) = \sqrt{x-1} + \sqrt{5-x}$

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

(c) $f(x) = \tan(x)$

(d) $f(x) = \sqrt{\ln(x)}$

Intermediate Value Theorem

This intuitive notion of continuity does help us understand another important concept as follows. Suppose f is defined on $[1,2]$ and $f(1) = -10$ and $f(2) = 5$. If f is continuous on $[1,2]$ (i.e., its graph can be sketched as a continuous line from $(1,-10)$ to $(2,5)$) then we know intuitively that somewhere on $[1,2]$ f must be equal to -9, and -8, and -7, -6, $\ldots$, 0, $1/2$, etc. In short, f takes on all *intermediate* values between -10 and 5. It may take on more values; f may actually equal 6 at some time, for instance, but we are guaranteed all values between -10 and 5.

While this notion seems intuitive, it is not trivial to prove and its importance is profound. Therefore the concept is stated in the form of a theorem.

Intermediate Value Theorem

Let f be a continuous function on $[a, b]$ and, without loss of generality, let $f(a) < f(b)$. Then for every value y, where $f(a) < y < f(b)$, there is a value c in $[a, b]$ such that $f(c) = y$.

One important application of the Intermediate Value Theorem is root finding. Given a function f, we are often interested in finding values of x where $f(x) = 0$. These roots may be very difficult to find exactly. Good approximations can be found through successive applications of this theorem. Suppose through direct computation we find that $f(a) < 0$ and $f(b) > 0$, where $a < b$. The Intermediate Value Theorem states that there is a c in $[a, b]$ such that $f(c) = 0$. The theorem does not give us any clue as to where that value is in the interval $[a, b]$, just that it exists.

There is a technique that produces a good approximation of c. Let d be the midpoint of the interval $[a, b]$ and consider $f(d)$. There are three possibilities:

1. $f(d) = 0$ – we got lucky and stumbled on the actual value. We stop as we found a root.

2. $f(d) < 0$ Then we know there is a root of f on the interval $[d, b]$ – we have halved the size of our interval, hence are closer to a good approximation of the root.

3. $f(d) > 0$ Then we know there is a root of f on the interval $[a, d]$ – again,we have halved the size of our interval, hence are closer to a good approximation of the root.

Successively applying this technique is called the *Bisection Method* of root finding. We continue until the interval is sufficiently small. We demonstrate this in the following example.

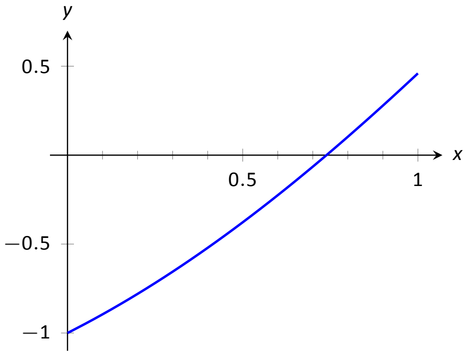


Figure 1.18: Graphing a root of $f(x) = x - \cos(x)$.

Iteration #	Interval	Midpoint Sign
1	$[0.7, 0.9]$	$f(0.8) > 0$
2	$[0.7, 0.8]$	$f(0.75) > 0$
3	$[0.7, 0.75]$	$f(0.725) < 0$
4	$[0.725, 0.75]$	$f(0.7375) < 0$
5	$[0.7375, 0.75]$	$f(0.7438) > 0$
6	$[0.7375, 0.7438]$	$f(0.7407) > 0$
7	$[0.7375, 0.7407]$	$f(0.7391) > 0$
8	$[0.7375, 0.7391]$	$f(0.7383) < 0$
9	$[0.7383, 0.7391]$	$f(0.7387) < 0$
10	$[0.7387, 0.7391]$	$f(0.7389) < 0$
11	$[0.7389, 0.7391]$	$f(0.7390) < 0$
12	$[0.7390, 0.7391]$	

Table 1.4: Iterations of the Bisection Method of Root Finding

Example 3

Approximate the root of $f(x) = x - \cos x$, accurate to three places after the decimal.

Solution. Consider the graph of $f(x) = x - \cos x$, shown in Figure 1.18. It is clear that the graph crosses the x-axis somewhere near $x = 0.8$. To start the Bisection Method, pick an interval that contains 0.8. We choose $[0.7, 0.9]$. Note that all we care about are signs of $f(x)$, not their actual value, so this is all we display.

Iteration 1: $f(0.7) < 0$, $f(0.9) > 0$, and $f(0.8) > 0$. So replace 0.9 with 0.8 and repeat.

Iteration 2: $f(0.7) < 0$, $f(0.8) > 0$, and at the midpoint, 0.75, we have $f(0.75) > 0$. So replace 0.8 with 0.75 and repeat. Note that we don't need to continue to check the endpoints, just the midpoint. Thus we put the rest of the iterations in Table 1.4.

Notice that in the 12$^{\text{th}}$ iteration we have the endpoints of the interval each starting with 0.739. Thus we have narrowed the zero down to an accuracy of the first three places after the decimal. Using a computer, we have

$$f(0.7390) = -0.00014, \quad f(0.7391) = 0.000024.$$

Either endpoint of the interval gives a good approximation of where f is 0. The Intermediate Value Theorem states that the actual zero is still within this interval. While we do not know its exact value, we know it starts with 0.739.

This type of exercise is rarely done by hand. Rather, it is simple to program a computer to run such an algorithm and stop when the endpoints differ by a preset small amount. One of the authors did write such a program and found the zero of f, accurate to 10 places after the decimal, to be 0.7390851332. While it took a few minutes to write the program, it took less than a thousandth of a second for the program to run the necessary 35 iterations. In less than 8 hundredths of a second, the zero was calculated to 100 decimal places (with less than 200 iterations).

It is a simple matter to extend the Bisection Method to solve similar problems to $f(x) = 0$. For instance, we can solve $f(x) = 1$. This may seem obvious, but to many it is not. It actually works very well to define a new function g where $g(x) = f(x) - 1$. Then use the Bisection Method to solve $g(x) = 0$.

Similarly, given two functions f and g, we can use the Bisection Method to solve $f(x) = g(x)$. Once again, create a new function h where $h(x) = f(x) - g(x)$ and solve $h(x) = 0$.

Activity 1.3–3

Use the Bisection Method to approximate, accurate to two decimal places, the value of the root of the given functions and intervals.

(a) $f(x) = x^2 + 2x - 4$; $\quad [1, 1.5]$

(b) $f(x) = e^x - 2$; $\quad [0.65, 0.7]$

Summary

In this section, we encountered the following important ideas:

- A function f is continuous at $x = a$ whenever $f(a)$ is defined, f has a limit as $x \to a$, and the value of the limit and the value of the function agree. This guarantees that there is not a hole or jump in the graph of f at $x = a$.
- We can use the Intermediate Value Theorem to determine if there may be a root of a continuous function on a closed interval.

Exercises

Terms and Concepts

1) In your own words, describe what it means for a function to be continuous.

2) In your own words, describe what the Intermediate Value Theorem states.

3) What is a "root" of a function?

4) Given functions f and g on an interval I, how can the Bisection Method be used to find a value c where $f(c) = g(c)$?

5) T/F: If f is defined on an open interval containing c, and $\lim\limits_{x\to c} f(x)$ exists, then f is continuous at c.

6) T/F: If f is continuous at c, then $\lim\limits_{x\to c} f(x)$ exists.

7) T/F: If f is continuous at c, then $\lim\limits_{x\to c^+} f(x) = f(c)$.

8) T/F: If f is continuous on $[a, b]$, then $\lim\limits_{x\to a^-} f(x) = f(a)$.

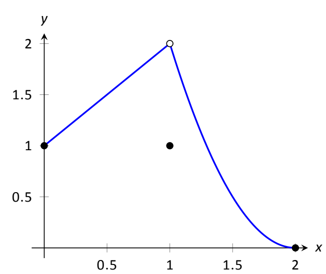
9) T/F: If f is continuous on $[0, 1)$ and $[1, 2)$, then f is continuous on $[0, 2)$.

10) T/F: The sum of continuous functions is also continuous.

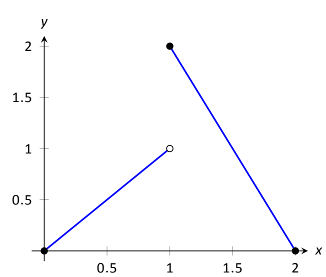
Problems

In exercises 11–17, a graph of f is given along with a value a. Determine if f is continuous at a; if it is not, state why it is not.

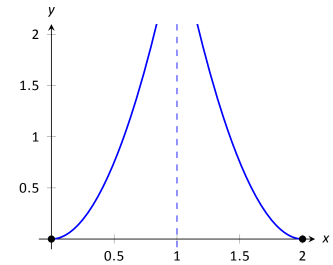
11) $a = 1$



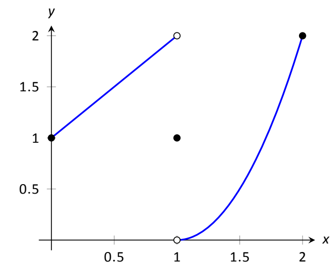
12) $a = 1$



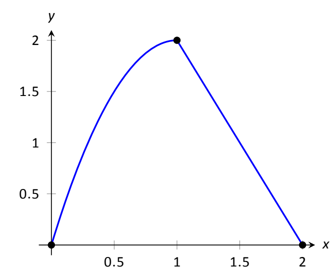
13) $a = 1$



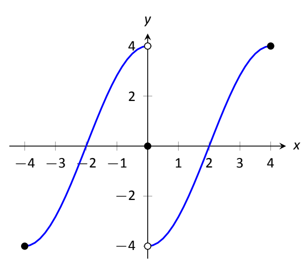
14) $a = 0$



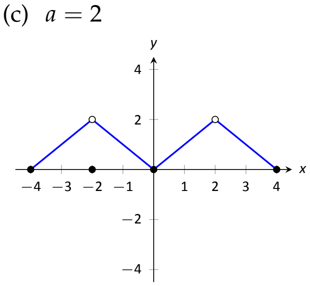
15) $a = 1$



16) $a = 4$



17) (a) $a = -2$
(b) $a = 0$
(c) $a = 2$



In exercises 18–21, determine if f is continuous at the indicated values. If not, explain why.

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

(a) $x = 0$

(b) $x = \pi$

19) $f(x) = \begin{cases} x^3 - x & x < 1 \\ x - 2 & x \geq 1 \end{cases}$

(a) $x = 0$

(b) $x = 1$

20) $f(x) = \begin{cases} \dfrac{x^2 + 5x + 4}{x^2 + 3x + 2} & x \neq -1 \\ 3 & x = -1 \end{cases}$

(a) $x = -1$

(b) $x = 10$

21) $f(x) = \begin{cases} \dfrac{x^2 - 64}{x^2 - 11x + 24} & x \neq 8 \\ 5 & x = 8 \end{cases}$

(a) $x = 0$

(b) $x = 8$

In exercises 22–32, give the intervals on which the given function is continuous.

22) $f(x) = x^2 - 3x + 9$

23) $g(x) = \sqrt{x^2 - 4}$

24) $h(k) = \sqrt{1 - k} + \sqrt{k + 1}$

25) $f(t) = \sqrt{5t^2 - 30}$

26) $g(t) = \dfrac{1}{\sqrt{1 - t^2}}$

27) $g(x) = \dfrac{1}{1 + x^2}$

28) $f(x) = e^x$

29) $g(s) = \ln s$

30) $h(t) = \cos(t)$

31) $f(k) = \sqrt{1 - e^k}$

32) $f(x) = \sin(e^x + x^2)$

In exercises 33–34, use the Bisection Method to approximate, accurate to two decimal places, the value of the root of the given function in the given interval.

33) $f(x) = \sin(x) - \dfrac{1}{2}; \quad [0.5, 0.55]$

34) $f(x) = \cos(x) - \sin(x); \quad [0.7, 0.8]$

1.4 *Epsilon-Delta Definition of a Limit*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- What do we mean by saying "arbitrarily close"?
- What is the precise definition of a *limit*?
- What are ϵ and δ and how do they help in determining the value of the limit of a function at a point?

Introduction

Recall our definition of a limit of a function from Section 1.1.

The Limit of a Function

Given a function f, a fixed input $x = a$, and a real number L, we say that f *has limit* L *as* x *approaches* a, and write

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

provided that we can make $f(x)$ as close to L as we like by taking x sufficiently close (but not equal) to a. If we cannot make $f(x)$ as close to a single value as we would like as x approaches a, then we say that f *does not have a limit as* x *approaches* a.

The problem with this definition is that the words "approaches" and "close" are not exact. In what way does the variable x approach a? How "close" do x and y have to be to a and L, respectively? Finally, what determines "sufficiently close"? The precise definition we describe in this section comes from formalizing our original definition. A quick restatement gets us closer to what we want:

> "If x is within a certain *tolerance level* of a, then the corresponding value $y = f(x)$ is within a certain *tolerance level* of L."

The accepted notation for the x-tolerance is the lowercase Greek letter delta, or δ, and the y-tolerance is lowercase epsilon, or ϵ. One more rephrasing nearly gets us to the actual definition:

> "If x is within δ units of a, then the corresponding value of y is within ϵ units of L."

Note that this means (let the "$\longrightarrow$" represent the word "implies"): $a - \delta < x < a + \delta \longrightarrow L - \epsilon < y < L + \epsilon$ or $|x - a| <$

$\delta \longrightarrow |y - L| < \epsilon$. The point is that δ and ϵ, being tolerances, can be any positive (but typically small) values. Finally, we have the formal definition of the limit with the notation seen in the previous section.

The Limit of a Function f

Let f be a function defined on an open interval containing a. The notation

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

read as "the limit of $f(x)$, as x approaches a, is L," means that given any $\epsilon > 0$, there exists $\delta > 0$ such that whenever $|x - a| < \delta$, we have $|f(x) - L| < \epsilon$.

Mathematicians often enjoy writing ideas without using any words. Here is the wordless definition of the limit:

$$\lim_{x \to a} f(x) = L \iff \forall\, \epsilon > 0, \exists\, \delta > 0 \text{ s.t. } |x - a| < \delta \longrightarrow |f(x) - L| < \epsilon.$$

There is an emphasis here that we may have passed over before. In the definition, the y-tolerance ϵ is given *first* and then the limit will exist *if* we can find an x-tolerance δ that works.

It is time for an example. Note that the explanation is long, but it will take you through all steps necessary to understand the ideas.

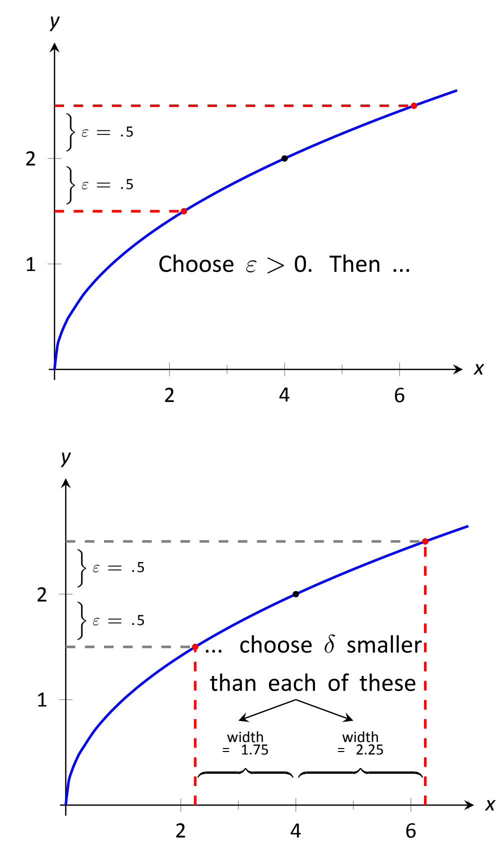


Figure 1.19: Illustrating the $\epsilon - \delta$ process. With $\epsilon = 0.5$, we pick any $\delta < 1.75$.

Example 1

Show that $\lim_{x \to 4} \sqrt{x} = 2$.

Solution. Before we use the formal definition, let's try some numerical tolerances. What if the y tolerance is 0.5, or $\epsilon = 0.5$? How close to 4 does x have to be so that y is within 0.5 units of 2 (or $1.5 < y < 2.5$)? In this case, we can just square these values to get $1.5^2 < x < 2.5^2$, or

$$2.25 < x < 6.25.$$

So, what is the desired x tolerance? Remember, we want to find a symmetric interval of x values, namely $4 - \delta < x < 4 + \delta$. The lower bound of 2.25 is 1.75 units from 4; the upper bound of 6.25 is 2.25 units from 4. We need the smaller of these two distances; we must have $\delta = 1.75$. See Figure 1.19.

Now read it in the correct way: For the y tolerance $\epsilon = 0.5$, we have found an x tolerance, $\delta = 1.75$, so that whenever x is within δ units of 4, then y is within ϵ units of 2. That's what we were trying to find.

Let's try another value of ϵ. What if the y tolerance is 0.01, or $\epsilon = 0.01$? How close to 4 does x have to be in order for y to be within 0.01 units of 2 (or $1.99 < y < 2.01$)? Again, we just square these values to get

$1.99^2 < x < 2.01^2$, or

$$3.9601 < x < 4.0401.$$

So, what is the desired x tolerance? In this case we must have $\delta = 0.0399$. Note that in some sense, it looks like there are two tolerances (below 4 of 0.0399 units and above 4 of 0.0401 units). However, we couldn't use the larger value of 0.0401 for δ since then the interval for x would be $3.9599 < x < 4.0401$ resulting in y values of $1.98995 < y < 2.01$ (which contains values NOT within 0.01 units of 2).

What we have so far: if $\epsilon = 0.5$, then $\delta = 1.75$ and if $\epsilon = 0.01$, then $\delta = 0.0399$. A pattern is not easy to see, so we switch to general ϵ and δ and do the calculations symbolically. We start by assuming $y = \sqrt{x}$ is within ϵ units of 2:

$$\begin{aligned}
|y-2| &< \epsilon & \\
-\epsilon < y-2 &< \epsilon & \text{(Absolute value)} \\
-\epsilon < \sqrt{x}-2 &< \epsilon & (y=\sqrt{x}) \\
2-\epsilon < \sqrt{x} &< 2+\epsilon & \text{(Add 2)} \\
(2-\epsilon)^2 < x &< (2+\epsilon)^2 & \text{(Square all)} \\
4-4\epsilon+\epsilon^2 < x &< 4+4\epsilon+\epsilon^2 & \text{(Expand)} \\
4-(4\epsilon-\epsilon^2) < x &< 4+(4\epsilon+\epsilon^2) & \text{(Rewrite)}
\end{aligned}$$

Since we want this last interval to describe an x tolerance around 4, we have that either $\delta = 4\epsilon + \epsilon^2$ or $\delta = 4\epsilon - \epsilon^2$. However, as we saw in the case when $\epsilon = 0.01$, we want the smaller of the two values for δ. So, to conclude this part, we set δ equal to the minimum of these two values, or $\delta = \min\{4\epsilon + \epsilon^2, 4\epsilon - \epsilon^2\}$. Since $\epsilon > 0$, the minimum will occur when $\delta = 4\epsilon - \epsilon^2$. That's the formula!

We can check this for our previous values. If $\epsilon = 0.5$, the formula gives $\delta = 4(0.5) - (0.5)^2 = 1.75$ and when $\epsilon = 0.01$, the formula gives $\delta = 4(0.01) - (0.01)^2 = 0.399$.

So given any $\epsilon > 0$, we can set $\delta = 4\epsilon - \epsilon^2$ and the limit definition is satisfied. We have shown formally (and finally!) that $\lim\limits_{x\to 4} \sqrt{x} = 2$.

If you are thinking this process is long, you would be right. The previous example is also a bit unsatisfying in that $\sqrt{4} = 2$; why work so hard to prove something so obvious? Many $\epsilon - \delta$ proofs are long and difficult to do. In this section, we will focus on examples where the answer is, frankly, obvious, because the non–obvious examples are even harder. That is why theorems about limits are so useful! After doing a few more ϵ–δ proofs, you will really appreciate the analytical "short cuts" we previously discussed.

Example 2

Show that $\lim\limits_{x\to 2} x^2 = 4$.

Solution. Let's do this example symbolically from the start. Let $\epsilon > 0$ be given; we want $|y-4| < \epsilon$, i.e., $|x^2-4| < \epsilon$. How do we find δ such that when $|x-2|<\delta$, we are guaranteed that $|x^2-4|<\epsilon$?

This is a bit trickier than the previous example, but let's start by noticing that $|x^2-4| = |x-2|\cdot|x+2|$. Consider:

$$|x^2-4|<\epsilon \longrightarrow |x-2|\cdot|x+2|<\epsilon \longrightarrow |x-2|<\frac{\epsilon}{|x+2|}. \tag{1.1}$$

Could we not set $\delta = \dfrac{\epsilon}{|x+2|}$?

We are close to an answer, but the catch is that δ must be a *constant* value (so it can't contain x). There is a way to work around this, but we do have to make an assumption. Remember that ϵ is supposed to be a small number, which implies that δ will also be a small value. In particular, we can (probably) assume that $\delta < 1$. If this is true, then $|x-2|<\delta$ would imply that $|x-2|<1$, giving $1<x<3$.

Now, back to the fraction $\dfrac{\epsilon}{|x+2|}$. If $1<x<3$, then $3<x+2<5$. Taking reciprocals, we have $\dfrac{1}{5}<\dfrac{1}{|x+2|}<\dfrac{1}{3}$ so that, in particular,

$$\frac{\epsilon}{5}<\frac{\epsilon}{|x+2|}. \tag{1.2}$$

This suggests that we set $\delta = \dfrac{\epsilon}{5}$. To see why, let's go back to the equations:

$$\begin{aligned}
|x-2| &< \delta \\
|x-2| &< \frac{\epsilon}{5} \\
|x-2|\cdot|x+2| &< |x+2|\cdot\frac{\epsilon}{5} \\
|x^2-4| &< |x+2|\cdot\frac{\epsilon}{5} \\
|x^2-4| &< |x+2|\cdot\frac{\epsilon}{5} < |x+2|\cdot\frac{\epsilon}{|x+2|} = \epsilon
\end{aligned}$$

We have arrived at $|x^2-4|<\epsilon$ as desired. Note again, in order to make this happen we needed δ to first be less than 1. That is a safe assumption; we want ϵ to be arbitrarily small, forcing δ to also be small.

We have also picked δ to be smaller than "necessary." We could get by with a slightly larger δ, as shown in Figure 1.20. The dashed, red lines show the boundaries defined by our choice of ϵ. The gray, dashed lines show the boundaries defined by setting $\delta = \epsilon/5$. Note how these gray lines are within the red lines. That is perfectly fine; by choosing x within the gray lines we are guaranteed that $f(x)$ will be within ϵ of 4.

In summary, given $\epsilon>0$, set $\delta=\epsilon/5$. Then $|x-2|<\delta$ implies $|x^2-4|<\epsilon$ (i.e. $|y-4|<\epsilon$) as desired. We have shown that $\lim\limits_{x\to 2} x^2 = 4$. Figure 1.20 gives a visualization of this; by restricting x to values within $\delta=\epsilon/5$ of 2, we see that $f(x)$ is within ϵ of 4.

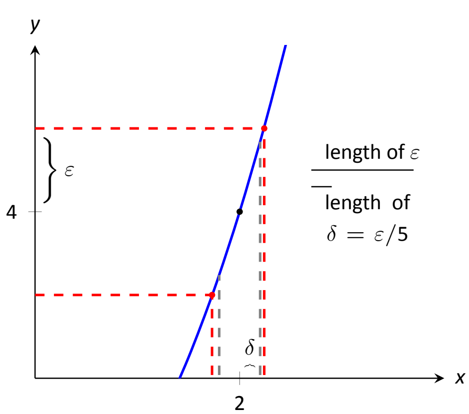


Figure 1.20: Choosing $\delta = \epsilon/5$ in Example 2.

Example 3

Show that $\lim\limits_{x\to 0} e^x = 1$.

Solution. Symbolically, we want to take the equation $|e^x - 1| < \epsilon$ and unravel it to the form $|x - 0| < \delta$. Let's look at some calculations:

$$\begin{aligned} |e^x - 1| &< \epsilon & \\ -\epsilon < e^x - 1 &< \epsilon & \text{(Definition of absolute value)} \\ 1 - \epsilon < e^x &< 1 + \epsilon & \text{(Add 1)} \\ \ln(1 - \epsilon) < x &< \ln(1 + \epsilon) & \text{(Take natural logs)} \end{aligned}$$

Making the safe assumption that $\epsilon < 1$ ensures the last inequality is valid (i.e., so that $\ln(1 - \epsilon)$ is defined). Recall $\ln(1) = 0$ and $\ln(x) < 0$ when $0 < x < 1$. So $\ln(1 - \epsilon) < 0$, hence we consider its absolute value. We can then set δ to be the minimum of $|\ln(1 - \epsilon)|$ and $\ln(1 + \epsilon)$; i.e.,

$$\delta = \min\{|\ln(1 - \epsilon)|, \ln(1 + \epsilon)\}.$$

Now, we work through the actual the proof:

$$\begin{aligned} |x - 0| &< \delta & \\ -\delta < x &< \delta & \text{(Definition of absolute value)} \\ \ln(1 - \epsilon) < x &< \ln(1 + \epsilon) & \text{(By our choice of } \delta) \\ 1 - \epsilon < e^x &< 1 + \epsilon & \text{(Exponentiate)} \\ -\epsilon < e^x - 1 &< \epsilon & \text{(Subtract 1)} \end{aligned}$$

In summary, given $\epsilon > 0$, let $\delta = \min\{|\ln(1 - \epsilon|, \ln(1 + \epsilon)\}$. Then $|x - 0| < \delta$ implies $|e^x - 1| < \epsilon$ as desired. We have shown that $\lim\limits_{x \to 0} e^x = 1$.

We note that we could actually show that $\lim\limits_{x \to a} e^x = e^a$ for any constant a. We do this by factoring out e^a from both sides, leaving us to show $\lim\limits_{x \to a} e^{x-a} = 1$ instead. By using the substitution $y = x - a$, this reduces to showing $\lim_{y \to 0} e^y = 1$ which we just did in the last example. As an added benefit, this shows that in fact the function $f(x) = e^x$ is *continuous* at all values of x.

Limits involving infinity

In the graph of $f(x) = \dfrac{1}{x^2}$ shown in Figure 1.21, we see that near 0, the function explodes, getting larger and larger, heading off to positive infinity.

Recall from Section 1.2 that in a case like this, we write

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

We can make this notion precise as follows:

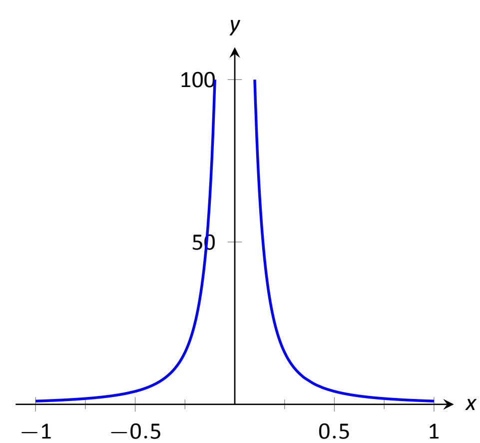


Figure 1.21: Graph of $f(x) = 1/x^2$.

Limit of Infinity, ∞

We say $\lim\limits_{x\to c} f(x) = \infty$ if for every $M > 0$ there exists $\delta > 0$ such that if $0 < |x - c| < \delta$ then $f(x) \geq M$.

This is just like the ϵ–δ definition of a limit above. In that definition, given any (small) value ϵ, if we let x get close enough to c (within δ units of c) then $f(x)$ is guaranteed to be within ϵ of $f(c)$. Here, given any (large) value M, if we let x get close enough to c (within δ units of c), then $f(x)$ will be at least as large as M. In other words, if we get close enough to c, then we can make $f(x)$ as large as we want. We can define limits equal to $-\infty$ in a similar way.

Once again note that by saying $\lim\limits_{x\to c} f(x) = \infty$ we are implicitly stating that *the* limit of $f(x)$, as x approaches c, *does not exist.* A limit only exists when $f(x)$ approaches an actual numeric value. We use the concept of limits that approach infinity because they are helpful and descriptive.

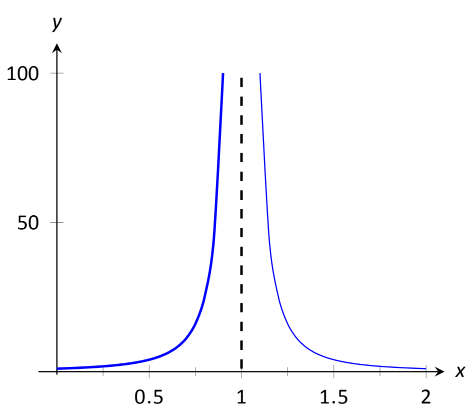


Figure 1.22: Observing infinite limit as $x \to 1$ in Example 4.

Example 4

Find $\lim\limits_{x\to 1} \dfrac{1}{(x-1)^2}$ as shown in Figure 1.22

Solution. In Example 2 of Section 1.2, by inspecting values of x close to 1 we concluded that this limit does not exist. That is, it cannot equal any real number. But the limit could be infinite. And in fact, we see that the function does appear to be growing larger and larger, as $f(.99) = 10^4$, $f(.999) = 10^6$, $f(.9999) = 10^8$. A similar thing happens on the other side of 1. In general, let a "large" value M be given. Let $\delta = 1/\sqrt{M}$. If x is within δ of 1, i.e., if $|x-1| < 1/\sqrt{M}$, then:

$$\begin{aligned} |x-1| &< \frac{1}{\sqrt{M}} \\ (x-1)^2 &< \frac{1}{M} \\ \frac{1}{(x-1)^2} &> M, \end{aligned}$$

which is what we wanted to show. So we may say $\lim\limits_{x\to 1} 1/(x-1)^2 = \infty$.

Limits at infinity

Let's again consider $f(x) = \frac{1}{x^2}$, as shown in Figure 1.21. Note that as x gets very large, $f(x)$ gets very, very close to zero. We represent this concept with notation such as

$$\lim_{x\to\infty} \frac{1}{x^2} = 0,$$

and give the following precise definition.

Limits at Infinity

1. We say $\lim_{x\to\infty} f(x) = L$ if for every $\epsilon > 0$ there exists $N > 0$ such that if $x \geq N$, then $|f(x) - L| < \epsilon$.
2. We say $\lim_{x\to-\infty} f(x) = L$ if for every $\epsilon > 0$ there exists $N < 0$ such that if $x \leq N$, then $|f(x) - L| < \epsilon$.

This says that f is sufficiently close to L whenever x is sufficiently large. In other words, if x is greater than some number N then $f(x)$ is between $L - \epsilon$ and $L + \epsilon$. If a smaller ϵ is chosen, a larger value of N may be required.

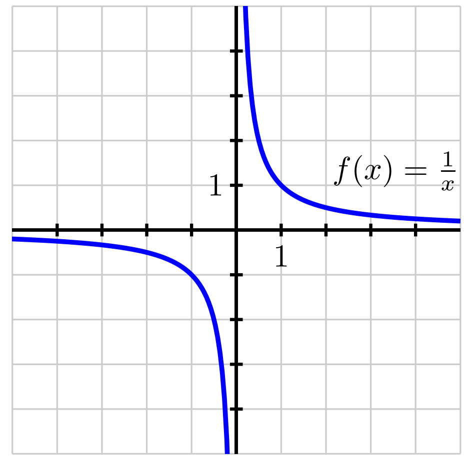


Figure 1.23: Observing limit as $x \to \infty$ in Example 5.

Example 5

Show $\lim_{x\to\infty} \frac{1}{x} = 0$ as shown in Figure 1.23.

Solution. By the precise definition of limits at infinity, given $\epsilon > 0$, we need to find N such that if $x > N$ then $|\frac{1}{x} - 0| < \epsilon$. Since x is approching ∞, we may assume $x > 0$. Then $\frac{1}{x} < \epsilon$ and thus $x > \frac{1}{\epsilon}$. Let $N = \frac{1}{\epsilon}$. If $x > N$ then $|\frac{1}{x} - 0| < \epsilon$, which is what we wanted to show. So we may say $\lim_{x\to\infty} \frac{1}{x} = 0$. If we choose smaller values for ϵ, we will need bigger values for N but the inequality will still hold. Table 1.5 shows different values of ϵ and the corresponding values of N.

ϵ	N
1	1
0.2	5
0.1	10
0.05	20
0.01	100

Table 1.5: Values of ϵ and corresponding values of N.

Summary

In this section, we encountered the following important ideas:

- We now have definitions of limit that do not include arbitrary measures such as "approaches" or "sufficiently close to".
- The precise definitions given in this section are used to prove the Limit Laws we gave in Section 1.1.

Exercises

Terms and Concepts

1) What is wrong with the following "definition" of a limit?

   "The limit of $f(x)$, as x approaches a, is K" means that given any $\delta > 0$ there exists $\epsilon > 0$ such that whenever $|f(x) - K| < \epsilon$, we have $|x - a| < \delta$.

2) Which is given first in establishing a limit, the x–tolerance or the y–tolerance?

3) T/F: δ must always be positive.

4) T/F: ϵ must always be positive.

Problems

In exercises 5–11, prove the given limit using an $\epsilon - \delta$ proof.

5) $\lim_{x\to 2} 5 = 5$

6) $\lim_{x\to 5} 3 - x = -2$

7) $\lim_{x\to 3} x^2 - 3 = 6$

8) $\lim_{x\to 4} x^2 + x - 5 = 15$

9) $\lim_{x\to 2} x^3 - 1 = 7$

10) $\lim_{x\to 0} e^{2x} - 1 = 0$

11) $\lim_{x\to 0} \sin x = 0$ (Hint: use the fact that $|\sin x| \leq |x|$, with equality only when $x = 0$.)

Chapter 2
Derivatives

2.1 Derivatives and Rates of Change

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- How is the average rate of change of a function on a given interval defined, and what does this quantity measure?
- How is the instantaneous rate of change of a function at a particular point defined? How is the instantaneous rate of change linked to average rate of change?
- What is the derivative of a function at a given point? What does this derivative value measure? How do we interpret the derivative value graphically?
- How are limits used formally in the computation of derivatives?
- In contexts other than the position of a moving object, what does the derivative of a function measure?
- What are the units on the derivative function f', and how are they related to the units of the original function f?
- What is a central difference, and how can one be used to estimate the value of the derivative at a point from given function data?
- Given the value of the derivative of a function at a point, what can we infer about how the value of the function changes nearby?

Introduction

An idea that sits at the foundations of calculus is the *instantaneous rate of change* of a function. This rate of change is always considered with respect to change in the input variable, often at a particular fixed input value. This is a generalization of the notion of instantaneous velocity and essentially allows us to consider the question "how do we measure how fast a particular function is changing at a given point?" When the original function represents the position of a moving object, this instantaneous rate of change is precisely velocity, and might be measured in units such as feet per second. But in other contexts, instantaneous rate of change could measure the number of cells added to a bacteria culture per day, the number of additional gallons of gasoline consumed by going one mile per additional mile per hour in a car's velocity, or the number of dollars added to a mortgage payment for each percentage increase in interest rate. Regardless of the presence of a physical or practical interpretation of a function, the instantaneous rate of change may also be interpreted geometrically in connection to the function's graph, and this connection is also foundational to many of the main ideas in calculus.

In what follows, we will introduce terminology and notation that makes it easier to talk about the instantaneous rate of change of a function at a point. In addition, just as instantaneous velocity is defined in terms of average velocity, the more general instantaneous rate of change will be connected to the more general average rate of change. Recall that for a moving object with position function s, its average velocity on the time interval $t = a$ to $t = a + h$ is given by the quotient

$$AV_{[a,a+h]} = \frac{s(a+h) - s(a)}{h}.$$

In a similar way, we make the following definition for an arbitrary function $y = f(x)$.

Average Rate of Change

For a function f, the *average rate of change* of f on the interval $[a, a+h]$ is given by the value

$$AV_{[a,a+h]} = \frac{f(a+h) - f(a)}{h}.$$

Equivalently, if we want to consider the average rate of change of f on $[a, b]$, we compute

$$AV_{[a,b]} = \frac{f(b) - f(a)}{b - a}.$$

It is essential to understand how the average rate of change of f on an interval is connected to its graph.

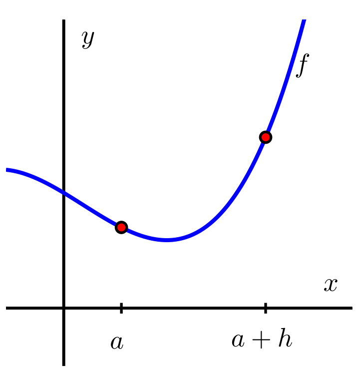


Figure 2.1: Plot of $y = f(x)$ for Preview Activity 2.4.

Preview Activity 2.1

Suppose that f is the function given by the graph in Figure 2.1 and that a and $a + h$ are the input values as labeled on the x-axis. Use the graph in Figure 2.1 to answer the following questions.

(a) Locate and label the points $(a, f(a))$ and $(a+h, f(a+h))$ on the graph.

(b) Construct a right triangle whose hypotenuse is the line segment from $(a, f(a))$ to $(a+h, f(a+h))$. What are the lengths of the respective legs of this triangle?

(c) What is the slope of the line that connects the points $(a, f(a))$ and $(a+h, f(a+h))$?

(d) Write a meaningful sentence that explains how the average rate of change of the function on a given interval and the slope of a related line are connected.

The Derivative of a Function at a Point

Just as we defined instantaneous velocity in terms of average velocity, we now define the instantaneous rate of change of a function at a point in terms of the average rate of change of the function f over related intervals. In addition, we give a special name to "the instantaneous rate of change of f at a," calling this quantity "the *derivative* of f at a," with this value being represented by the shorthand notation $f'(a)$. Specifically, we make the following definition.

Derivative at a Point

Let f be a function and $x = a$ a value in the function's domain. We define the *derivative of f with respect to x evaluated at $x = a$*, denoted $f'(a)$, by the formula

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

provided this limit exists. We say that a function that has a derivative at $x = a$ is *differentiable* at $x = a$.

Aloud, we read the symbol $f'(a)$ as either "f-prime at a" or "the derivative of f evaluated at $x = a$." Much of the next several chapters will be devoted to understanding, computing, applying, and interpreting derivatives. For now, we make the following important notes.

- The derivative of f at the value $x = a$ is defined as the limit of the average rate of change of f on the interval $[a, a+h]$ as $h \to 0$. It is possible for this limit not to exist, so not every function has a derivative at every point.

- The derivative is a generalization of the instantaneous velocity of a position function: when $y = s(t)$ is a position function of a moving body, $s'(a)$ tells us the instantaneous velocity of the body at time $t = a$.

- Because the units on $\frac{f(a+h)-f(a)}{h}$ are "units of f per unit of x," the derivative has these very same units. For instance, if s measures position in feet and t measures time in seconds, the units on $s'(a)$ are feet per second.

- Because the quantity $\frac{f(a+h)-f(a)}{h}$ represents the slope of the line through $(a, f(a))$ and $(a+h, f(a+h))$, when we compute the derivative we are taking the limit of a collection of slopes of lines, and thus the derivative itself represents the slope of a particularly important line.

While all of the above ideas are important and we will add depth and perspective to them through additional time and study, for now it is most essential to recognize how the derivative of a function at a given value represents the slope of a certain line. Thus, we expand upon the last bullet item above.

As we move from an average rate of change to an instantaneous one, we can think of one point as "sliding towards" another. In particular, provided the function has a derivative at $(a, f(a))$, the point $(a+h, f(a+h))$ will approach $(a, f(a))$ as $h \to 0$. Because this process of taking a limit is a dynamic one, it can be helpful to use computing technology to visualize what the limit is accomplishing. While there are many different options[1], one of the best is a java applet in which the user is able to control the point that is moving. See the examples referenced in the footnote here, or consider building your own, perhaps using the fantastic free program Geogebra[2].

[1] For a helpful collection of java applets, consider the work of David Austin of Grand Valley State University at `http://gvsu.edu/s/5r`, and the particularly relevant example at `http://gvsu.edu/s/5s`. For applets that have been built in Geogebra, a nice example is the work of Marc Renault of Shippensburg University at `http://gvsu.edu/s/5p`, with the example at `http://gvsu.edu/s/5q` being especially fitting for our work in this section. There are scores of other examples posted by other authors on the internet.

[2] Available for free download from `http://geogebra.org`.

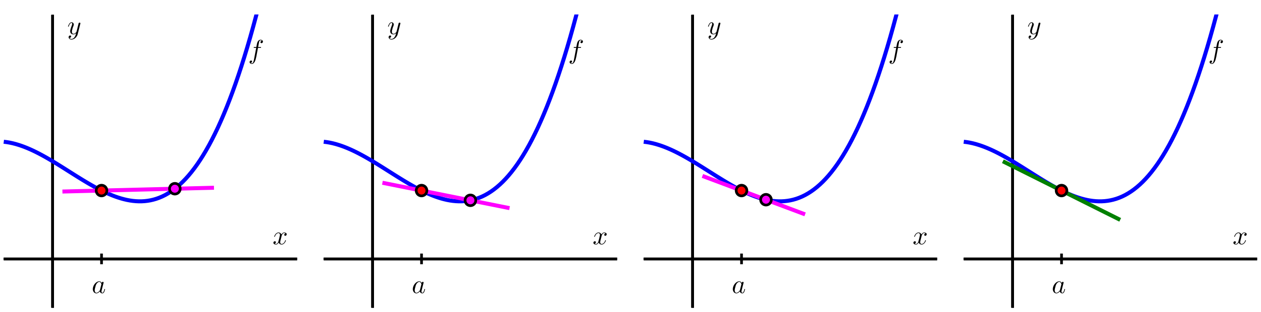


Figure 2.2: A sequence of secant lines approaching the tangent line to f at $(a, f(a))$.

In Figure 2.2, we provide a sequence of figures with several different lines through the points $(a, f(a))$ and $(a+h, f(a+h))$ that are generated by different values of h. These lines (shown in the first three figures in magenta), are often called *secant lines* to the curve $y = f(x)$. A secant line to a curve is simply a line that passes through two points that lie on the curve. For each such line, the slope of the secant line is $m = \frac{f(a+h)-f(a)}{h}$, where the value of h depends on the location of the point we choose. We can see in the diagram how, as $h \to 0$, the secant lines start to approach a single line that passes through the point $(a, f(a))$. In the situation where the limit of the slopes of the secant lines exists, we say that the resulting value is the slope of the *tangent line* to the curve. This tangent line (shown in the right-most figure in green) to the graph of $y = f(x)$ at the point $(a, f(a))$ is the line through $(a, f(a))$ whose slope is $m = f'(a)$.

As we will see in subsequent study, the existence of the tangent line at $x = a$ is connected to whether or not the function f looks like a straight line when viewed up close at $(a, f(a))$, which can also be seen in Figure 2.3, where we combine the four graphs in Figure 2.2 into the single one on the left, and then we zoom in on the box centered at $(a, f(a))$, with that view expanded on the right (with two of the secant lines omitted). Note how the tangent line sits relative to the curve $y = f(x)$ at $(a, f(a))$ and how closely it resembles the curve near $x = a$.

At this time, it is most important to note that $f'(a)$, the instantaneous rate of change of f with respect to x at $x = a$, also measures the slope of the tangent line to the curve $y = f(x)$ at $(a, f(a))$. The following example demonstrates several key ideas involving the derivative of a function.

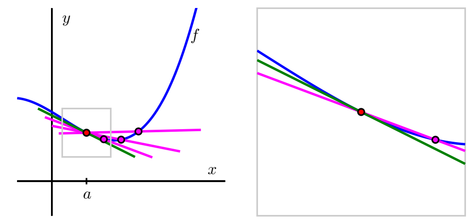


Figure 2.3: A sequence of secant lines approaching the tangent line to f at $(a, f(a))$. At right, we zoom in on the point $(a, f(a))$. The slope of the tangent line (in green) to f at $(a, f(a))$ is given by $f'(a)$.

Example 1

For the function given by $f(x) = x - x^2$, use the limit definition of the derivative to compute $f'(2)$. In addition, discuss the meaning of this value and draw a labeled graph that supports your explanation.

Solution. From the limit definition, we know that

$$f'(2) = \lim_{h \to 0} \frac{f(2+h) - f(2)}{h}.$$

Now we use the rule for f, and observe that $f(2) = 2 - 2^2 = -2$ and $f(2+h) = (2+h) - (2+h)^2$. Substituting these values into the limit definition, we have that

$$f'(2) = \lim_{h \to 0} \frac{(2+h) - (2+h)^2 - (-2)}{h}.$$

Observe that with h in the denominator and our desire to let $h \to 0$, we have to wait to take the limit (that is, we wait to actually let h approach 0). Thus, we do additional algebra. Expanding and distributing in the numerator,

$$f'(2) = \lim_{h \to 0} \frac{2 + h - 4 - 4h - h^2 + 2}{h}.$$

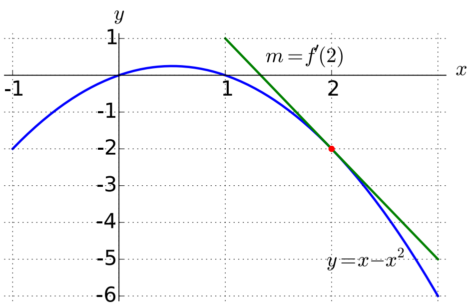


Figure 2.4: The tangent line to $y = x - x^2$ at the point $(2, -2)$.

Combining like terms, we have

$$f'(2) = \lim_{h \to 0} \frac{-3h - h^2}{h}.$$

Next, we observe that there is a common factor of h in both the numerator and denominator, which allows us to simplify and find that

$$f'(2) = \lim_{h \to 0} (-3 - h).$$

Finally, we are able to take the limit as $h \to 0$, and thus conclude that $f'(2) = -3$.

Now, we know that $f'(2)$ represents the slope of the tangent line to the curve $y = x - x^2$ at the point $(2, -2)$; $f'(2)$ is also the instantaneous rate of change of f at the point $(2, -2)$. Graphing both the function and the line through $(2, -2)$ with slope $m = f'(2) = -3$, we indeed see that by calculating the derivative, we have found the slope of the tangent line at this point, as shown in Figure 2.4.

Example 2

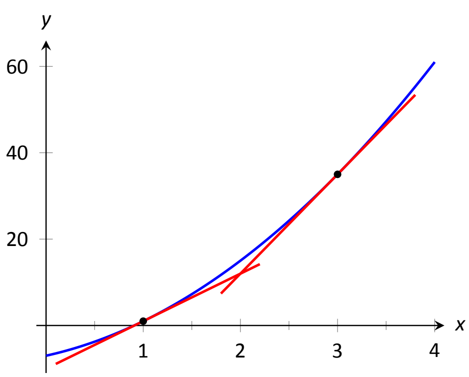


Figure 2.5: A graph of $f(x) = 3x^2 + 5x - 7$ and its tangent lines at $x = 1$ and $x = 3$.

Let $f(x) = 3x^2 + 5x - 7$. Find: 1) $f'(1)$; 2) the equation of the tangent line to the graph of f at $x = 1$; 3) $f'(3)$; and 4) the equation of the tangent line to the graph f at $x = 3$.

Solution.

1) We compute this directly using the definition of the derivative at a point.

$$\begin{aligned} f'(1) &= \lim_{h \to 0} \frac{f(1+h) - f(1)}{h} \\ &= \lim_{h \to 0} \frac{3(1+h)^2 + 5(1+h) - 7 - (3(1)^2 + 5(1) - 7)}{h} \\ &= \lim_{h \to 0} \frac{3h^2 + 11h}{h} \\ &= \lim_{h \to 0} 3h + 11 = 11. \end{aligned}$$

2) The tangent line at $x = 1$ has slope $f'(1)$ and goes through the point $(1, f(1)) = (1, 1)$. Thus the tangent line has equation, in point-slope form, $y = 11(x - 1) + 1$. In slope-intercept form we have $y = 11x - 10$.

3) Again, using the definition,

$$\begin{aligned} f'(3) &= \lim_{h \to 0} \frac{f(3+h) - f(3)}{h} \\ &= \lim_{h \to 0} \frac{3(3+h)^2 + 5(3+h) - 7 - (3(3)^2 + 5(3) - 7)}{h} \\ &= \lim_{h \to 0} \frac{3h^2 + 23h}{h} \\ &= \lim_{h \to 0} 3h + 23 = 23 \end{aligned}$$

4) The tangent line at $x = 3$ has slope 23 and goes through the point

$(3, f(3)) = (3, 35)$. Thus the tangent line has equation $y = 23(x - 3) + 35 = 23x - 34$.

The following activities will help you explore a variety of key ideas related to derivatives.

Activity 2.1–1

Consider the function f whose formula is $f(x) = 3 - 2x$.

(a) What familiar type of function is f? What can you say about the slope of f at every value of x?

(b) Compute the average rate of change of f on the intervals $[1,4]$, $[3,7]$, and $[5,5+h]$; simplify each result as much as possible. What do you notice about these quantities?

(c) Use the limit definition of the derivative to compute the exact instantaneous rate of change of f with respect to x at the value $a = 1$. That is, compute $f'(1)$ using the limit definition. Show your work. Is your result surprising?

(d) Without doing any additional computations, what are the values of $f'(2)$, $f'(\pi)$, and $f'(-\sqrt{2})$? Why?

Activity 2.1–2

A water balloon is tossed vertically in the air from a window. The balloon's height in feet at time t in seconds after being launched is given by $s(t) = -16t^2 + 16t + 32$. Use this function to respond to each of the following questions.

(a) Sketch an accurate, labeled graph of s on the axes provided in Figure 2.6. You should be able to do this without using computing technology.

(b) Compute the average rate of change of s on the time interval $[1,2]$. Include units on your answer and write one sentence to explain the meaning of the value you found.

(c) Use the limit definition to compute the instantaneous rate of change of s with respect to time, t, at the instant $a = 1$. Show your work using proper notation, include units on your answer, and write one sentence to explain the meaning of the value you found.

(d) On your graph in (a), sketch two lines: one whose slope represents the average rate of change of s on $[1,2]$, the other whose slope represents the instantaneous rate of change of s at the instant $a = 1$. Label each line clearly.

(e) For what values of a do you expect $s'(a)$ to be positive? Why? Answer the same questions when "positive" is replaced by "negative" and "zero."

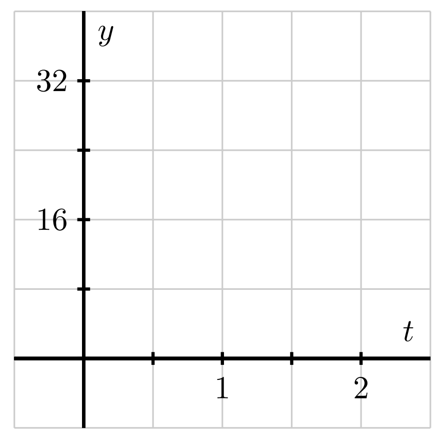


Figure 2.6: Axes for plotting $y = s(t)$ in Activity 2.4–5.

Activity 2.1–3

A rapidly growing city in Arizona has its population P at time t, where t is the number of decades after the year 2010, modeled by the formula $P(t) = 25000e^{t/5}$. Use this function to respond to the following questions.

(a) Sketch an accurate graph of P for $t = 0$ to $t = 5$ on the axes provided in Figure 2.7. Label the scale on the axes carefully.

(b) Compute the average rate of change of P between 2030 and 2050. Include units on your answer and write one sentence to explain the meaning (in everyday language) of the value you found.

(c) Use the limit definition to write an expression for the instantaneous rate of change of P with respect to time, t, at the instant $a = 2$. Explain why this limit is difficult to evaluate exactly.

(d) Estimate the limit in (c) for the instantaneous rate of change of P at the instant $a = 2$ by using several small h values. Once you have determined an accurate estimate of $P'(2)$, include units on your answer, and write one sentence (using everyday language) to explain the meaning of the value you found.

(e) On your graph above, sketch two lines: one whose slope represents the average rate of change of P on $[2, 4]$, the other whose slope represents the instantaneous rate of change of P at the instant $a = 2$.

(f) In a carefully-worded sentence, describe the behavior of $P'(a)$ as a increases in value. What does this reflect about the behavior of the given function P?

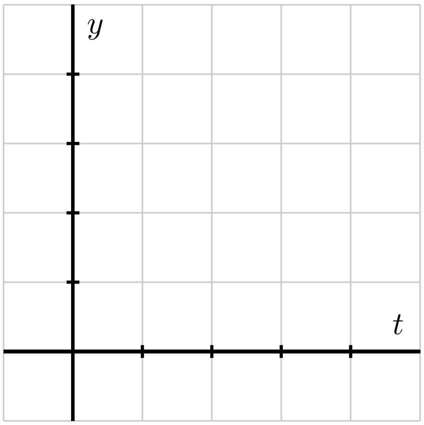


Figure 2.7: Axes for plotting $y = P(t)$ in Activity 2.1–3.

Units of the derivative function

As we now know, the derivative of the function f at a fixed value x is given by

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

and this value has several different interpretations. If we set $x = a$, one meaning of $f'(a)$ is the slope of the tangent line at the point $(a, f(a))$.

In alternate notation, we also sometimes equivalently write $\frac{df}{dx}$ or $\frac{dy}{dx}$ instead of $f'(x)$, and these notations help us to further see the units (and thus the meaning) of the derivative as it is viewed as *the instantaneous rate of change of f with respect to x*. Note that the units on the slope of the secant line, $\frac{f(x+h)-f(x)}{h}$, are "units of f per unit of x." Thus, when we take the limit to get $f'(x)$, we get these same units on the derivative $f'(x)$: units of f per unit of x. Regardless of the function f under consideration (and regardless of the variables being used), it is helpful to remember that the units on the derivative function are "units of output per unit of input," in terms of the input and output of the original function.

For example, say that we have a function $y = P(t)$, where P measures the population of a city (in thousands) at the start of year t (where $t = 0$ corresponds to 2010 AD), and we are told that $P'(2) = 21.37$. What is the meaning of this value? Well, since P is measured in thousands and t is measured in years, we can say that the instantaneous rate of change of the city's population with respect to time at the start of 2012 is 21.37 thousand people per year. We therefore expect that in the coming year, about $21,370$ people will be added to the city's population.

Estimating Derivatives

An interesting and powerful feature of mathematics is that it can often be thought of both in abstract terms and in applied ones. For instance, calculus can be developed almost entirely as an abstract collection of ideas that focus on properties of arbitrary functions. At the same time, calculus can also be very directly connected to our experience of physical reality by considering functions that represent meaningful processes. We have already seen that for a position function $y = s(t)$, say for a ball being tossed straight up in the air, the ball's velocity at time t is given by $v(t) = s'(t)$, the derivative of the position function. Further, recall that if $s(t)$ is measured in feet at time t, the units on $v(t) = s'(t)$ are feet per second.

In what follows, we investigate several different functions, each with specific physical meaning, and think about how the units on the independent variable, dependent variable, and the derivative function add to our understanding. To start, we consider the familiar problem of a position function of a moving object.

Activity 2.1–4

One of the longest stretches of straight (and flat) road in North America can be found on the Great Plains in the state of North Dakota on state highway 46, which lies just south of the interstate highway I-94 and runs through the town of Gackle. A car leaves town (at time $t = 0$) and heads east on highway 46; its position in miles from Gackle at time t in minutes is given by the graph of the function in Figure 2.8. Three important points are labeled on the graph; where the curve looks linear, assume that it is indeed a straight line.

(a) In everyday language, describe the behavior of the car over the provided time interval. In particular, discuss what is happening on the time intervals $[57, 68]$ and $[68, 104]$.

(b) Find the slope of the line between the points $(57, 63.8)$ and $(104, 106.8)$. What are the units on this slope? What does the slope represent?

(c) Find the average rate of change of the car's position on the interval

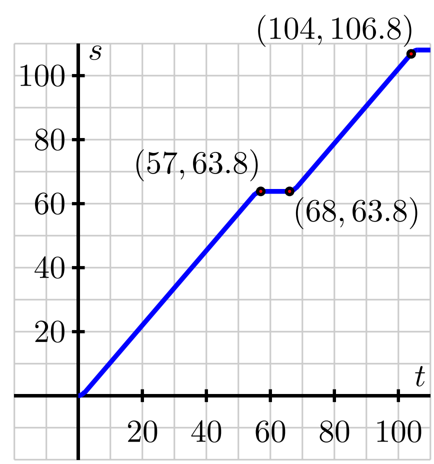


Figure 2.8: The graph of $y = s(t)$, the position of the car along highway 46, which tells its distance in miles from Gackle, ND, at time t in minutes.

[68, 104]. Include units on your answer.

(d) Estimate the instantaneous rate of change of the car's position at the moment $t = 80$. Write a sentence to explain your reasoning and the meaning of this value.

Toward more accurate derivative estimates

It is also helpful to recall that when we want to estimate the value of $f'(x)$ at a given x, we can use the *difference quotient* $\frac{f(x+h)-f(x)}{h}$ with a relatively small value of h. In doing so, we should use both positive and negative values of h in order to make sure we account for the behavior of the function on both sides of the point of interest. To that end, we consider the following brief example to demonstrate the notion of a *central difference* and its role in estimating derivatives.

Example 3

Suppose that $y = f(x)$ is a function for which three values are known: $f(1) = 2.5$, $f(2) = 3.25$, and $f(3) = 3.625$. Estimate $f'(2)$.

Solution. We know that $f'(2) = \lim_{h\to 0} \dfrac{f(2+h)-f(2)}{h}$. But since we don't have a graph for $y = f(x)$ nor a formula for the function, we can neither sketch a tangent line nor evaluate the limit exactly. We can't even use smaller and smaller values of h to estimate the limit. Instead, we have just two choices: using $h = -1$ or $h = 1$, depending on which point we pair with $(2, 3.25)$.

So, one estimate is

$$f'(2) \approx \frac{f(1)-f(2)}{1-2} = \frac{2.5-3.25}{-1} = 0.75.$$

The other is

$$f'(2) \approx \frac{f(3)-f(2)}{3-2} = \frac{3.625-3.25}{1} = 0.375.$$

Since the first approximation looks only backward from the point $(2, 3.25)$ and the second approximation looks only forward from $(2, 3.25)$, it makes sense to average these two values in order to account for behavior on both sides of the point of interest. Doing so, we find that

$$f'(2) \approx \frac{0.75+0.375}{2} = 0.5625.$$

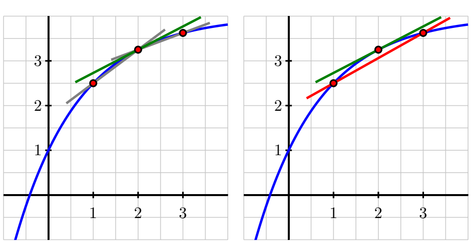


Figure 2.9: At left, the graph of $y = f(x)$ along with the secant line through $(1, 2.5)$ and $(2, 3.25)$, the secant line through $(2, 3.25)$ and $(3, 3.625)$, as well as the tangent line. At right, the same graph along with the secant line through $(1, 2.5)$ and $(3, 3.625)$, plus the tangent line.

The intuitive approach to average the two estimates found in Example 3 is in fact the best possible estimate to $f'(2)$ when we have just two function values for f on opposite sides of the point of interest. To see why, we think about the diagram in Figure 2.9, which shows a possible function $y = f(x)$ that satisfies the data given in Example 3. On the left, we see the two secant lines with slopes that come from computing the *backward difference* $\frac{f(1)-f(2)}{1-2} = 0.75$ and from the *forward difference* $\frac{f(3)-f(2)}{3-2} = 0.375$. Note how the first such line's slope over-estimates the slope of the tangent line at $(2, f(2))$, while the second line's slope under-estimates $f'(2)$. On the right, however, we see the secant line whose slope is given by the *central difference*

$$\frac{f(3)-f(1)}{3-1} = \frac{3.625-2.5}{2} = \frac{1.125}{2} = 0.5625.$$

Note that this central difference has the exact same value as the average of the forward difference and backward difference (and it is straightforward to explain why this always holds), and moreover that the central difference yields a very good approximation to the derivative's value, in part because the secant line that uses both a point before and after the point of tangency yields a line that is closer to being parallel to the tangent line.

In general, the central difference approximation to the value of the first derivative is given by

$$f'(a) \approx \frac{f(a+h)-f(a-h)}{2h},$$

and this quantity measures the slope of the secant line to $y = f(x)$ through the points $(a-h, f(a-h))$ and $(a+h, f(a+h))$. Anytime we have symmetric data surrounding a point at which we desire to estimate the derivative, the central difference is an ideal choice for so doing.

The following activities will further explore the meaning of the derivative in several different contexts while also viewing the derivative from graphical, numerical, and algebraic perspectives.

Activity 2.1–5

A potato is placed in an oven, and the potato's temperature F (in degrees Fahrenheit) at various points in time is taken and recorded in Table 2.1. Time t is measured in minutes.

(a) Use a central difference to estimate the instantaneous rate of change of the temperature of the potato at $t = 30$. Include units on your answer.

(b) Use a central difference to estimate the instantaneous rate of change of the temperature of the potato at $t = 60$. Include units on your answer.

(c) Without doing any calculation, which do you expect to be greater: $F'(75)$ or $F'(90)$? Why?

(d) Suppose it is given that $F(64) = 330.28$ and $F'(64) = 1.341$. What are the units on these two quantities? What do you expect the temperature of the potato to be when $t = 65$? when $t = 66$? Why?

t	$F(t)$
0	70
15	180.5
30	251
45	296
60	324.5
75	342.8
90	354.5

Table 2.1: The temperature of a potato in an oven at various times.

(e) Write a couple of careful sentences that describe the behavior of the temperature of the potato on the time interval $[0, 90]$, as well as the behavior of the instantaneous rate of change of the temperature of the potato on the same time interval.

Example 4

A company manufactures rope, and the total cost of producing r feet of rope is $C(r)$ dollars. What does it mean to say that $C(2000) = 800$? What are the units of $C'(r)$? Suppose that $C(2000) = 800$ and $C'(2000) = 0.35$. Estimate $C(2100)$, and justify your estimate by writing at least one sentence that explains your thinking. Which of the following statements do you think is true, and why?

- $C'(2000) < C'(3000)$
- $C'(2000) = C'(3000)$
- $C'(2000) > C'(3000)$

Suppose someone claims that $C'(5000) = -0.1$. What would the practical meaning of this derivative value tell you about the approximate cost of the next foot of rope? Is this possible? Why or why not?

Solution. When we say $C(2000) = 800$, we mean that the total cost of producing 2000 feet of rope is 800 dollars. Remember that the units on any derivative are "units of output per unit of input," and the units of $C'(r)$ are "dollars per foot."

If $C(2000) = 800$ and $C'(2000) = 0.35$, then we know once 2000 feet of rope are produced, the total cost function is increasing at \$0.35 per additional foot of rope. Then, if we manufacture an additional 100 feet of rope, the additional total cost will be approximately

$$100 \text{ feet} \cdot 0.35 \, \frac{\text{dollars}}{\text{foot}} = 35 \text{ dollars}.$$

Therefore, we find that $C(2100) \approx C(2000) + 35 = 835$, or that the cost to make 2100 feet of rope is about \$835.

Either $C'(2000) = C'(3000)$ or $C'(2000) > C'(3000)$, since we expect the cost per foot of additional rope to either stay constant or to get smaller as the production volume increases. Said differently, the instantaneous rate of change of the total cost function should either be constant or decrease due to economy of scale.

It is impossible to have $C'(5000) = -0.1$ and indeed to have any negative derivative value for the total cost function. The total cost function $C(r)$ can never decrease, because it doesn't make sense for the total cost of producing 5001 feet of rope to be less than the total cost of producing 5000 feet of rope.

Activity 2.1–6

Researchers at a major car company have found a function that relates gasoline consumption to speed for a particular model of car. In particular, they have determined that the consumption C, in **liters per kilometer**, at a given speed s, is given by a function $C = f(s)$, where s is the car's speed in **kilometers per hour**.

(a) Data provided by the car company tells us that $f(80) = 0.015$, $f(90) = 0.02$, and $f(100) = 0.027$. Use this information to estimate the instantaneous rate of change of fuel consumption with respect to speed at $s = 90$. Be as accurate as possible, use proper notation, and include units on your answer.

(b) By writing a complete sentence, interpret the meaning (in the context of fuel consumption) of "$f(80) = 0.015$."

(c) Write at least one complete sentence that interprets the meaning of the value of $f'(90)$ that you estimated in (a).

Summary

In this section, we encountered the following important ideas:

- The average rate of change of a function f on the interval $[a, b]$ is $\frac{f(b) - f(a)}{b - a}$. The units on the average rate of change are units of f per unit of x, and the numerical value of the average rate of change represents the slope of the secant line between the points $(a, f(a))$ and $(b, f(b))$ on the graph of $y = f(x)$. If we view the interval as being $[a, a + h]$ instead of $[a, b]$, the meaning is still the same, but the average rate of change is now computed by $\frac{f(a+h) - f(a)}{h}$.
- The instantaneous rate of change with respect to x of a function f at a value $x = a$ is denoted $f'(a)$ (read "the derivative of f evaluated at a" or "f-prime at a") and is defined by the formula

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

  provided the limit exists. Note particularly that the instantaneous rate of change at $x = a$ is the limit of the average rate of change on $[a, a + h]$ as $h \to 0$.
- Provided the derivative $f'(a)$ exists, its value tells us the instantaneous rate of change of f with respect to x at $x = a$, which geometrically is the slope of the tangent line to the curve $y = f(x)$ at the point $(a, f(a))$. We even say that $f'(a)$ is the *slope of the curve* $y = f(x)$ at the point $(a, f(a))$.
- Limits are the link between average rate of change and instantaneous rate of change: they allow us to move from the rate of change over an interval to the rate of change at a single point.
- The central difference approximation to the value of the first derivative is given by

$$f'(a) \approx \frac{f(a+h) - f(a-h)}{2h},$$

  and this quantity measures the slope of the secant line to $y = f(x)$ through the points $(a - h, f(a - h))$ and $(a + h, f(a + h))$. The central difference generates a good approximation of the derivative's value any time we have symmetric data surrounding a point of interest.
- Knowing the derivative and function values at a single point enables us to estimate other function values nearby. If, for example, we know that $f'(7) = 2$, then we know that at $x = 7$, the function f is increasing at an instantaneous rate of 2 units of output for every one unit of input. Thus, we expect $f(8)$ to be approximately 2 units greater than $f(7)$. The value is approximate because we don't know that the rate of change stays the same as x changes.

Exercises

Terms and Concepts

1) T/F: Let f be a position function. The average rate of change on $[a, b]$ is the slope of the line through the points $(a, f(a))$ and $(b, f(b))$.

2) T/F: The definition of the derivative of a function at a point involves taking a limit.

3) In your own words, explain the difference between the average rate of change and instantaneous rate of change.

4) What is the instantaneous rate of change of position called?

5) Given a function $y = f(x)$, in your own words describe how to find the units of $f'(x)$.

6) Let $V(x)$ measure the volume, in decibels, measured inside a restaurant with x customers. What are the units of $V'(x)$?

7) Let $v(t)$ measure the velocity, in ft/s, of a car moving in a straight line t seconds after starting. What are the units of $v'(t)$?

Problems

In exercises 8–14, a function and an x-value a are given.

a) Find the derivative of the function at the given point.

b) Find the tangent line to the graph of the function at a.

8) $f(x) = 6; \quad a = -2$

9) $f(x) = 2x; \quad a = 3$

10) $h(x) = 4 - 3x; \quad a = 7$

11) $g(x) = x^2; \quad a = 2$

12) $f(x) = 3x^2 - x + 4; \quad a = -1$

13) $h(x) = \dfrac{1}{x}; \quad a = -2$

14) $r(x) = \dfrac{1}{x-2}; \quad a = 3$

In exercises 15–17, use a central difference to estimate the instantaneous rates of change at the indicated values.

15) The yearly profits $P(t)$, in millions of dollars, of a certain company from 1990 to 1996 are given in the following table.

Year	1990	1991	1992	1993	1994	1995	1996
Profit	0.5	1.0	1.2	1.6	2.5	1.6	2.0

Estimate $P'(1991)$, $P'(1993)$, and $P'(1995)$. Include units in your answers.

16) The position, $s(t)$, of an object moving in a straight line at time t is given by the following table.

t	0	0.5	1	1.5	2
$s(t)$	0	30	52	66	72

Estimate $s'(0.5)$, $s'(1)$, and $s'(1.5)$. Include units in your answers.

17) The average price, $p(t)$, for a ticket to a movie theater in North America for selected years is shown in the following table.

Year	1987	1991	1995	1999	2003	2007	2009
Price ($)	3.91	4.21	4.35	5.06	6.03	6.88	7.50

(Source: National Association of Theater Owners, www.natoonline.org)

(a) Estimate $p'(1991)$ and $p'(2003)$. Include units in your answers.

(b) Are we able to estimate $p'(2007)$ using a central difference? Why or why not?

18) A cup of coffee has its temperature F (in degrees Fahrenheit) at time t given by the function $F(t) = 75 + 110e^{-0.05t}$, where time is measured in minutes.

(a) Use a central difference with $h = 0.01$ to estimate the value of $F'(10)$.

(b) What are the units on the value of $F'(10)$ that you computed in (a)? What is the practical meaning of the value of $F'(10)$?

(c) Which do you expect to be greater: $F'(10)$ or $F'(20)$? Why?

(d) Write a sentence that describes the behavior of the function $y = F'(t)$ on the time interval $0 \leq t \leq 30$. How do you think its graph will look? Why?

19) The temperature change T (in Fahrenheit degrees), in a patient, that is generated by a dose q (in milliliters), of a drug, is given by the function $T = f(q)$.

(a) What does it mean to say $f(50) = 0.75$? Write a complete sentence to explain, using correct units.

(b) A person's sensitivity, s, to the drug is defined by the function $s(q) = f'(q)$. What are the units of sensitivity?

(c) Suppose that $f'(50) = -0.02$. Write a complete sentence to explain the meaning of this value. Include in your response the information given in (a).

20) The velocity of a ball that has been tossed vertically in the air is given by $v(t) = 16 - 32t$, where v is measured in feet per second, and t is measured in seconds. The ball is in the air from $t = 0$ until $t = 2$.

    (a) When is the ball's velocity greatest?

    (b) Determine the value of $v'(1)$. Justify your thinking.

    (c) What are the units on the value of $v'(1)$? What does this value and the corresponding units tell you about the behavior of the ball at time $t = 1$?

    (d) What is the physical meaning of the function $v'(t)$?

21) The value, V, of a particular automobile (in dollars) depends on the number of miles, m, the car has been driven, according to the function $V = h(m)$.

    (a) Suppose that $h(40000) = 15500$ and $h(55000) = 13200$. What is the average rate of change of h on the interval $[40000, 55000]$, and what are the units on this value?

    (b) In addition to the information given in (a), say that $h(70000) = 11100$. Determine the best possible estimate of $h'(55000)$ and write one sentence to explain the meaning of your result, including units on your answer.

    (c) Which value do you expect to be greater: $h'(30000)$ or $h'(80000)$? Why?

    (d) Write a sentence to describe the long-term behavior of the function $V = h(m)$, plus another sentence to describe the long-term behavior of $h'(m)$. Provide your discussion in practical terms regarding the value of the car and the rate at which that value is changing.

2.2 *The derivative function*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- How does the limit definition of the derivative of a function f lead to an entirely new (but related) function f'?
- What is the difference between writing $f'(a)$ and $f'(x)$?
- How is the graph of the derivative function $f'(x)$ connected to the graph of $f(x)$?
- What are some examples of functions f for which f' is not defined at one or more points?
- What does it mean graphically to say that a function f is differentiable at $x = a$? How is this connected to the function being locally linear?

Introduction

Given a function $y = f(x)$, we now know that if we are interested in the instantaneous rate of change of the function at $x = a$, or equivalently the slope of the tangent line to $y = f(x)$ at $x = a$, we can compute the value $f'(a)$. In all of our examples to date, we have arbitrarily identified a particular value of a as our point of interest: $a = 1$, $a = 3$, etc. But it is not hard to imagine that we will often be interested in the derivative value for more than just one a-value, and possibly for many of them. In this section, we explore how we can move from computing simply $f'(1)$ or $f'(3)$ to working more generally with $f'(a)$, and indeed $f'(x)$. Said differently, we will work toward understanding how the so-called process of "taking the derivative" generates a new function that is derived from the original function $y = f(x)$. The following preview activity starts us down this path.

Preview Activity 2.2

Consider the function $f(x) = 4x - x^2$.

(a) Use the limit definition to compute the following derivative values: $f'(0)$, $f'(1)$, $f'(2)$, and $f'(3)$.

(b) Observe that the work to find $f'(a)$ is the same, regardless of the value of a. Based on your work in (a), what do you conjecture is the value of $f'(4)$? How about $f'(5)$? (Note: you should *not* use the limit definition of the derivative to find either value.)

(c) Conjecture a formula for $f'(a)$ that depends only on the value a. That is, in the same way that we have a formula for $f(x)$ (recall $f(x) = 4x - x^2$), see if you can use your work above to guess a formula for $f'(a)$ in terms of a.

How the derivative is itself a function

In your work in Preview Activity 2.2 with $f(x) = 4x - x^2$, you may have found several patterns. One comes from observing that $f'(0) = 4$, $f'(1) = 2$, $f'(2) = 0$, and $f'(3) = -2$. That sequence of values leads us naturally to conjecture that $f'(4) = -4$ and $f'(5) = -6$. Even more than these individual numbers, if we consider the role of 0, 1, 2, and 3 in the process of computing the value of the derivative through the limit definition, we observe that the particular number has very little effect on our work. To see this more clearly, we compute $f'(a)$, where a represents a number to be named later. Following the now standard process of using the limit definition of the derivative,

$$\begin{aligned}
f'(a) &= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \\
&= \lim_{h \to 0} \frac{4(a+h) - (a+h)^2 - (4a - a^2)}{h} \\
&= \lim_{h \to 0} \frac{4a + 4h - a^2 - 2ha - h^2 - 4a + a^2}{h} \\
&= \lim_{h \to 0} \frac{4h - 2ha - h^2}{h} \\
&= \lim_{h \to 0} \frac{h(4 - 2a - h)}{h} \\
&= \lim_{h \to 0} (4 - 2a - h).
\end{aligned}$$

Here we observe that neither 4 nor $2a$ depend on the value of h, so as $h \to 0$, $(4 - 2a - h) \to (4 - 2a)$. Thus, $f'(a) = 4 - 2a$.

This observation is consistent with the specific values we found above: e.g., $f'(3) = 4 - 2(3) = -2$. And indeed, our work with a confirms that while the particular value of a at which we evaluate the derivative affects the value of the derivative, that value has almost no bearing on the process of computing the derivative. We note further that the letter being used is immaterial: whether we call it a, x, or anything else, the derivative at a given value is simply given by "4 minus 2 times the value." We choose to use x for consistency with the original function given by $y = f(x)$, as well as for the purpose of graphing the derivative function, and thus we have found that for the function $f(x) = 4x - x^2$, it follows that $f'(x) = 4 - 2x$.

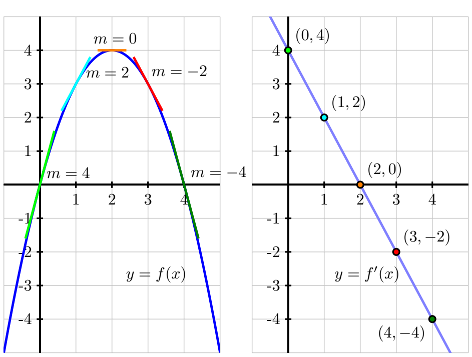


Figure 2.10: The graphs of $f(x) = 4x - x^2$ (at left) and $f'(x) = 4 - 2x$ (at right). Slopes on the graph of f correspond to heights on the graph of f'.

Because the value of the derivative function is so closely linked to the graphical behavior of the original function, it makes sense to look at both of these functions plotted on the same domain. In Figure 2.10, on the left we show a plot of $f(x) = 4x - x^2$ together with a selection of tangent lines at the points we've considered above. On the right, we show a plot of $f'(x) = 4 - 2x$ with emphasis on the heights of the derivative graph at the same selection of points. Notice the connection between colors in the left and right graph: the green tangent line on the original graph is tied to the green point on the right graph in the following way: *the slope of the tangent line* at a point on the left-hand graph is the same as the *height* at the corresponding point on the right-hand graph. That is, at each respective value of x, the slope of the tangent line to the original function at that x-value is the same as the height of the derivative function at that x-value. Do note, however, that the units on the vertical axes are different: in the left graph, the vertical units are simply the output units of f. On the right-hand graph of $y = f'(x)$, the units on the vertical axis are units of f per unit of x.

In Section 2.1 when we first defined the derivative, we wrote the definition in terms of a value a to find $f'(a)$. As we have seen above, the letter a is merely a placeholder, and it often makes more sense to use x instead. For the record, here we restate the definition of the derivative.

Of course, this relationship between the graph of a function $y = f(x)$ and its derivative is a dynamic one. An excellent way to explore how the graph of $f(x)$ generates the graph of $f'(x)$ is through a java applet. See, for instance, the applets at `http://gvsu.edu/s/5C` or `http://gvsu.edu/s/5D`, via the sites of David Austin(`http://gvsu.edu/s/5r`) and Marc Renault(`http://gvsu.edu/s/5p`).

Derivative as a Function

Let f be a function and x a value in the function's domain. We define the *derivative of f with respect to x at the value x*, denoted $f'(x)$, by the formula $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$, provided this limit exists.

Notation: Let $y = f(x)$. The following notation all represents the derivative:

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

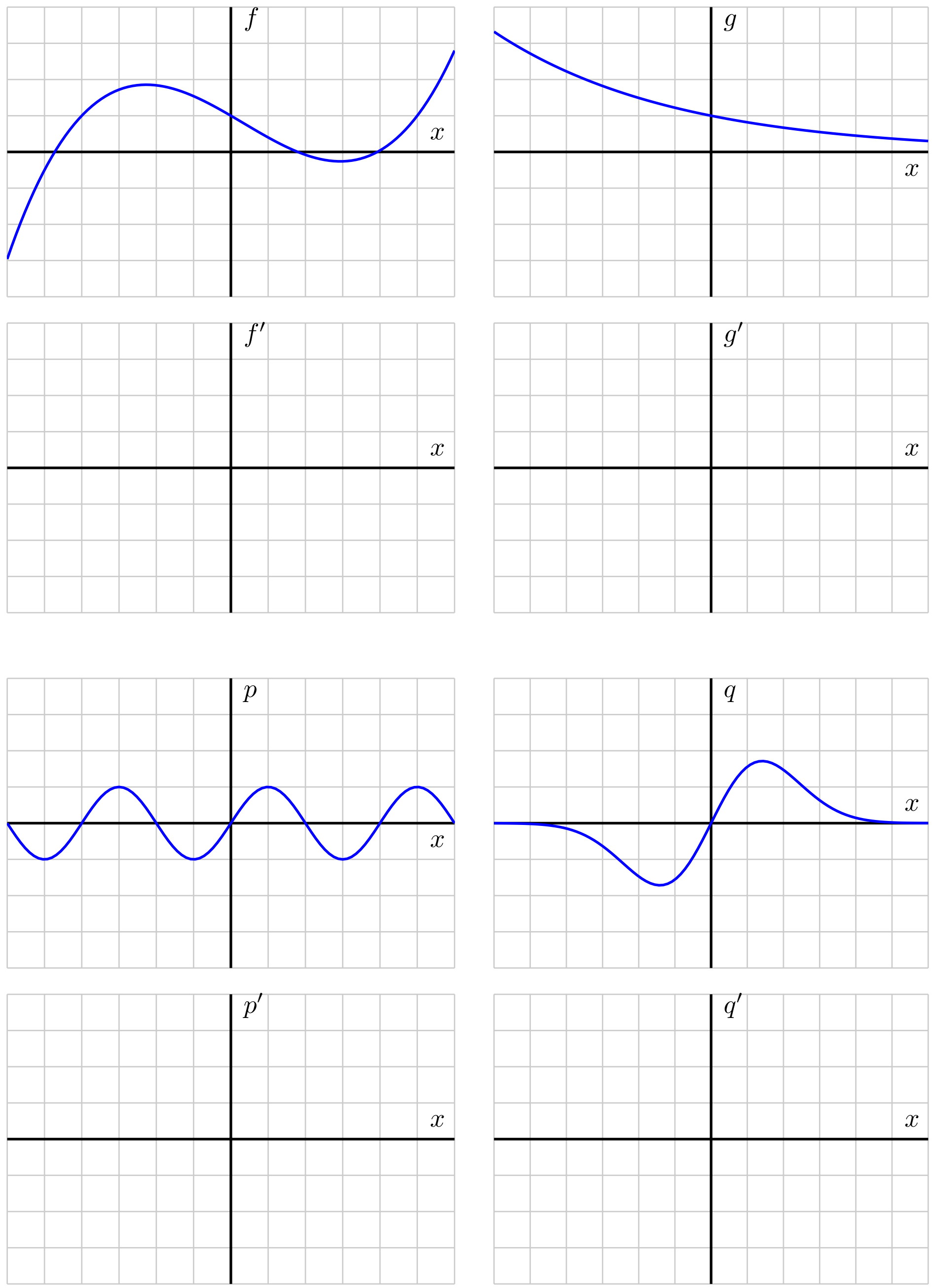
We now may take two different perspectives on thinking about the derivative function: given a graph of $y = f(x)$, how does this graph lead to the graph of the derivative function $y = f'(x)$? and given a formula for $y = f(x)$, how does the limit definition of the derivative generate a formula for $y = f'(x)$? We first explore the graphical relationship in the following activity.

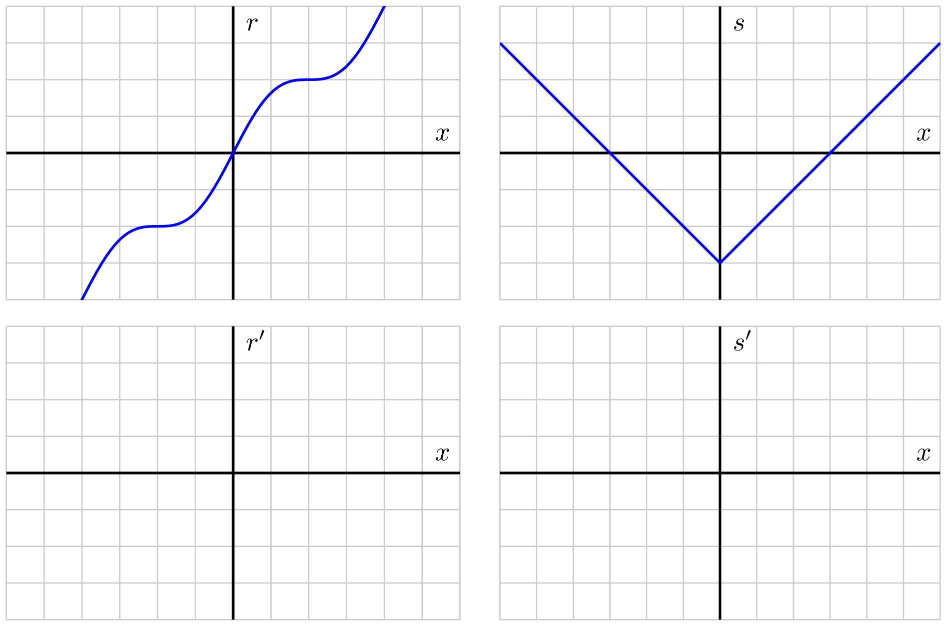
Important: The notation $\frac{dy}{dx}$ is one symbol; it is **not** the fraction "dy/dx". The notation, while somewhat confusing at first, was chosen with care. A fraction–looking symbol was chosen because the derivative has many fraction–like properties. Among other places, we see these properties at work when we talk about the units of the derivative, when we discuss the Chain Rule, and when we learn about integration (topics that appear in later sections and chapters)

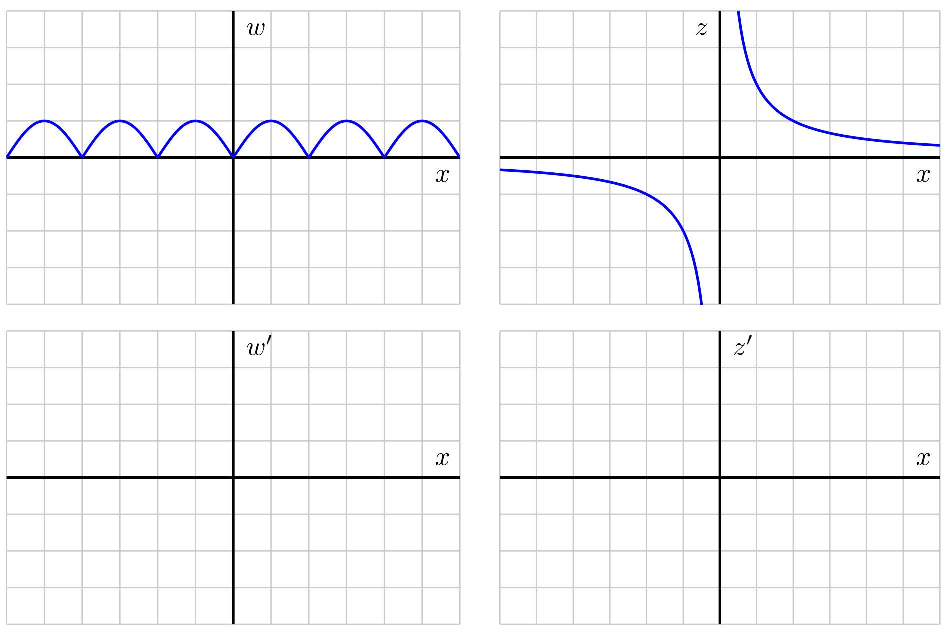
Activity 2.2–1

For each given graph of $y = f(x)$, sketch an approximate graph of its derivative function, $y = f'(x)$, on the axes immediately below. The scale of the grid for the graph of f is 1×1; assume the horizontal scale of the grid for the graph of f' is identical to that for f. If necessary, adjust and label the vertical scale on the axes for the graph of f'.

Write several sentences that describe your overall process for sketching the graph of the derivative function, given the graph the original function. What are the values of the derivative function that you tend to identify first? What do you do thereafter? How do key traits of the graph of the derivative function exemplify properties of the graph of the original function?







Now, recall the opening example of this section: we began with the function $y = f(x) = 4x - x^2$ and used the limit definition of the derivative to show that $f'(a) = 4 - 2a$, or equivalently that $f'(x) = 4 - 2x$. We subsequently graphed the functions f and f' as shown in Figure 2.10. Following Activity 2.2–1, we now understand that we could have constructed a fairly accurate graph of $f'(x)$ *without* knowing a formula for either f or f'. At the same time, it is ideal to know a formula for the derivative function whenever it is possible to find one.

For a dynamic investigation that allows you to experiment with graphing f' when given the graph of f, see `http://gvsu.edu/s/8y`.(Source: Marc Renault, Calculus Applets Using Geogebra.)

Example 1

Let $f(x) = 3x^2 + 5x - 7$. Find $f'(x)$.

Solution. We apply the definition of the derivative.

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h\to 0} \frac{3(x+h)^2 + 5(x+h) - 7 - (3x^2 + 5x - 7)}{h} \\ &= \lim_{h\to 0} \frac{3h^2 + 6xh + 5h}{h} \\ &= \lim_{h\to 0} 3h + 6x + 5 \\ &= 6x + 5 \end{aligned}$$

So $f'(x) = 6x + 5$. Recall in Example 3 of Section 2.1, we found that $f'(1) = 11$ and $f'(3) = 23$. Note our new computation of $f'(x)$ affirm these facts.

Example 2

Let $f(x) = \dfrac{1}{x+1}$. Find $f'(x)$.

Solution. We again apply the definition of the derivative.

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h\to 0} \frac{\frac{1}{x+h+1} - \frac{1}{x+1}}{h} \end{aligned}$$

Now find a common denominator then subtract; pull $1/h$ out front to facilitate reading.

$$\begin{aligned} &= \lim_{h\to 0} \frac{1}{h} \cdot \left(\frac{x+1}{(x+1)(x+h+1)} - \frac{x+h+1}{(x+1)(x+h+1)} \right) \\ &= \lim_{h\to 0} \frac{1}{h} \cdot \left(\frac{x+1-(x+h+1)}{(x+1)(x+h+1)} \right) \\ &= \lim_{h\to 0} \frac{1}{h} \cdot \left(\frac{-h}{(x+1)(x+h+1)} \right) \\ &= \lim_{h\to 0} \frac{-1}{(x+1)(x+h+1)} \\ &= \frac{-1}{(x+1)(x+1)} \\ &= \frac{-1}{(x+1)^2} \end{aligned}$$

So $f'(x) = \dfrac{-1}{(x+1)^2}$. To practice our notation, we could also state

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

In the next activity, we further explore the more algebraic approach to finding $f'(x)$: given a formula for $y = f(x)$, the limit definition of the derivative will be used to develop a formula for $f'(x)$.

Activity 2.2–2

For each of the listed functions, determine a formula for the derivative function. For the first two, determine the formula for the derivative by thinking about the nature of the given function and its slope at various points; do not use the limit definition. For the latter four, use the limit definition. Pay careful attention to the function names and independent variables. It is important to be comfortable with using letters other than f and x. For example, given a function $p(z)$, we call its derivative $p'(z)$.

(a) $(x) = 1$

(b) $g(t) = t$

(c) $p(z) = z^2$

(d) $q(s) = s^3$

(e) $F(t) = \dfrac{1}{t}$

(f) $G(y) = \sqrt{y}$

Differentiability

We recall that a function f is said to be differentiable at $x = a$ whenever $f'(a)$ exists. Moreover, for $f'(a)$ to exist, we know that the function $y = f(x)$ must have a tangent line at the point $(a, f(a))$, since $f'(a)$ is precisely the slope of this line. In order to even ask if f has a tangent line at $(a, f(a))$, it is necessary that f be continuous at $x = a$: if f fails to have a limit at $x = a$, if $f(a)$ is not defined, or if $f(a)$ does not equal the value of $\lim_{x \to a} f(x)$, then it doesn't even make sense to talk about a tangent line to the curve at this point.

Indeed, it can be proved formally that if a function f is differentiable at $x = a$, then it must be continuous at $x = a$. So, if f is not continuous at $x = a$, then it is automatically the case that f is not differentiable there. For example, in Figure 1.15 from our early discussion of continuity, both f and g fail to be differentiable at $x = 1$ because neither function is continuous at $x = 1$. But can a function fail to be differentiable at a point where the function is continuous?

In Figure 2.11, we consider the situation where a function has a sharp corner at a point. For the pictured function f, we observe that f is clearly continuous at $a = 1$, since $\lim_{x \to 1} f(x) = 1 = f(1)$.

But the function f in Figure 2.11 is not differentiable at $a = 1$ because $f'(1)$ fails to exist. One way to see this is to observe that $f'(x) = -1$ for every value of x that is less than 1, while $f'(x) = +1$ for every value of x that is greater than 1. That makes it seem that either $+1$ or -1 would be equally good candidates

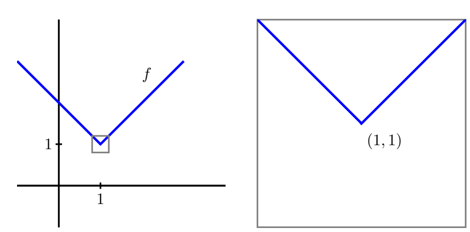


Figure 2.11: A function f that is continuous at $a = 1$ but not differentiable at $a = 1$; at right, we zoom in on the point $(1, 1)$ in a magnified version of the box in the left-hand plot.

for the value of the derivative at $x = 1$. Alternately, we could use the limit definition of the derivative to attempt to compute $f'(1)$, and discover that the derivative does not exist. Finally, we can also see visually that the function f in Figure 2.11 does not have a tangent line. When we zoom in on $(1, 1)$ on the graph of f, no matter how closely we examine the function, it will always look like a "V", and never like a single line, which tells us there is no possibility for a tangent line there.

To make a more general observation, if a function does have a tangent line at a given point, when we zoom in on the point of tangency, the function and the tangent line should appear essentially indistinguishable[3]. Conversely, if we have a function such that when we zoom in on a point the function looks like a single straight line, then the function should have a tangent line there, and thus be differentiable. Hence, a function that is differentiable at $x = a$ will, up close, look more and more like its tangent line at $(a, f(a))$, and thus we say that a function is differentiable at $x = a$ is *locally linear*.

[3] See, for instance, `http://gvsu.edu/s/6J` for an applet (due to David Austin, GVSU) where zooming in shows the increasing similarity between the tangent line and the curve.

Example 3

Find the derivative of the absolute value function—see Figure 2.12.

$$f(x) = |x| = \begin{cases} -x & x < 0 \\ x & x \geq 0 \end{cases}.$$

Solution. We need to evaluate $\lim_{h \to 0} \dfrac{f(x+h) - f(x)}{h}$. As f is piecewise–defined, we need to consider separately the limits when $x < 0$ and when $x > 0$.

When $x < 0$:

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

When $x > 0$, a similar computation shows that $\dfrac{d}{dx}(x) = 1$.

We need to also find the derivative at $x = 0$. By the definition of the derivative at a point, we have

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

Since $x = 0$ is the point where our function's definition switches from one piece to other, we need to consider left and right-hand limits. Consider the following, where we compute the left and right hand limits side by side.

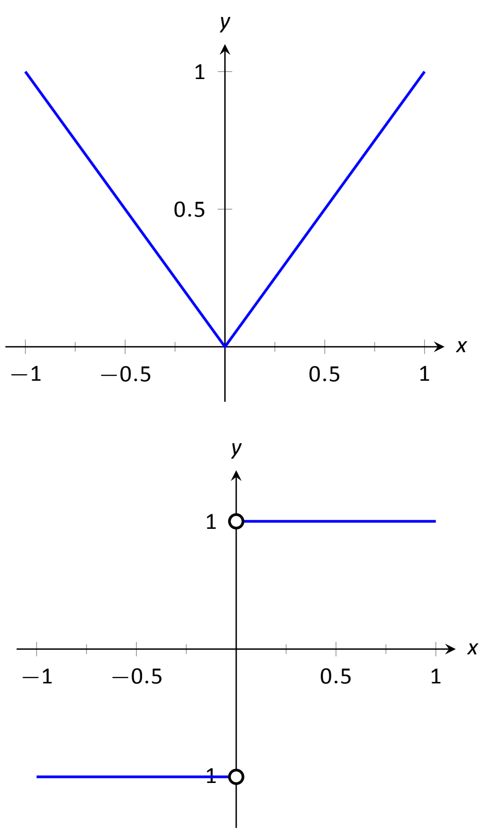


Figure 2.12: Above: The absolute value function, $f(x) = |x|$. Notice how the slope of the lines (and hence the tangent lines) abruptly changes at $x = 0$. Below: A graph of the derivative of $f(x) = |x|$.

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

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

The last lines of each column tell the story: the left and right hand limits are not equal. Therefore the limit does not exist at 0, and f is not differentiable at 0. So we have

$$f'(x) = \begin{cases} -1 & x < 0 \\ 1 & x > 0 \end{cases}.$$

At $x = 0$, $f'(x)$ does not exist; there is a jump discontinuity at 0; see Figure 2.12. So $f(x) = |x|$ is differentiable everywhere except at 0.

Activity 2.2–3

In this activity, we classify the points at which a function does not have a limit and is neither continuous nor differentiable. Let f be the function given in Figure 2.13.

(a) State all values of a for which f does not have a limit at $x = a$. For each, provide a reason for your conclusion.

(b) State all values of a for which f is not continuous at $x = a$. For each, provide a reason for your conclusion.

(c) State all values of a for which f is not differentiable at $x = a$. For each, provide a reason for your conclusion.

(d) True or false: if a function p is differentiable at $x = b$, then $\lim_{x\to b} p(x)$ must exist. Why?

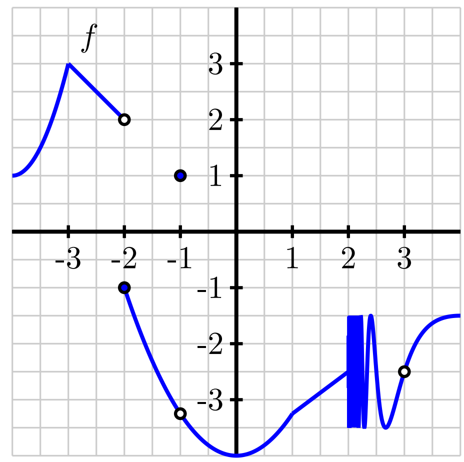


Figure 2.13: The graph of $y = f(x)$ for Activity 2.2–3.

Summary

In this section, we encountered the following important ideas:

- The limit definition of the derivative, $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$, produces a value for each x at which the derivative is defined, and this leads to a new function whose formula is $y = f'(x)$. Hence we talk both about a given function f and its derivative f'.
- There is essentially no difference between writing $f'(a)$ (as we did regularly in Section 2.1) and writing $f'(x)$. In either case, the variable is just a placeholder that is used to define the rule for the derivative function.
- Given the graph of a function $y = f(x)$, we can sketch an approximate graph of its derivative $y = f'(x)$ by observing that *heights* on the derivative's graph correspond to *slopes* on the original function's graph.
- In this section, we encountered some functions that had sharp corners on their graphs, such as the shifted absolute value function. At such points, the derivative fails to exist, and we say that f is not differentiable there. For now, it suffices to understand this as a consequence of the jump that must occur in the derivative function at a sharp corner on the graph of the original function.

Exercises

Terms and Concepts

1) T/F: Let f be a position function. The average rate of change on $[a, b]$ is the slope of the line through the points $(a, f(a))$ and $(b, f(b))$.

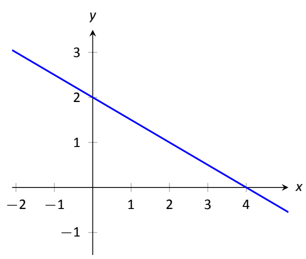
2) T/F: The definition of the derivative of a function at a point involves taking a limit.

3) In your own words, explain the difference between the average rate of change and instantaneous rate of change.

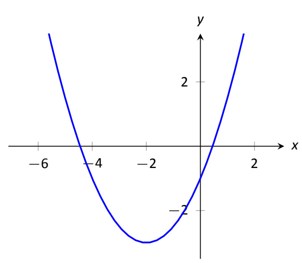
4) Let $y = f(x)$. Give three different notations equialent to "$f'(x)$."

Problems

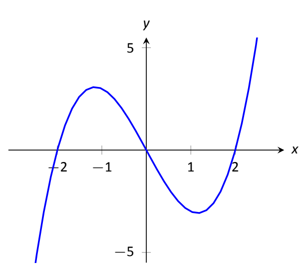
In exercises 5–8, a graph of a function is given. Using the graph, sketch $f'(x)$.



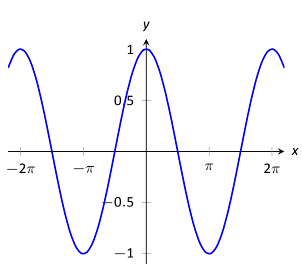
5)



6)



7)



8)

9) Using the graph of $g(x)$ below, answer the following questions.

(a) Where is $g(x) > 0$?

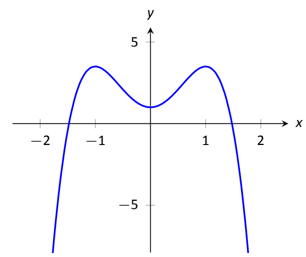
(b) Where is $g(x) < 0$?

(c) Where is $g(x) = 0$?

(d) Where is $g'(x) < 0$?

(e) Where is $g'(x) > 0$?

(f) Where is $g'(x) = 0$?

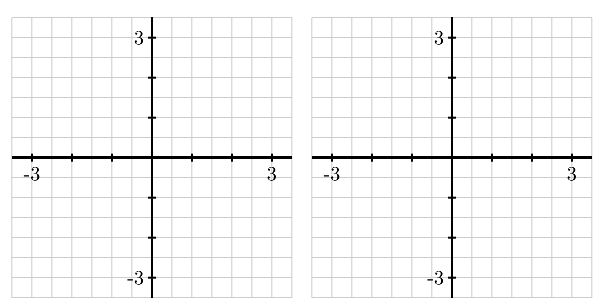


10) Let f be a function with the following properties: f is differentiable at every value of x (that is, f has a derivative at every point), $f(-2) = 1$, and $f'(-2) = -2$, $f'(-1) = -1$, $f'(0) = 0$, $f'(1) = 1$, and $f'(2) = 2$.

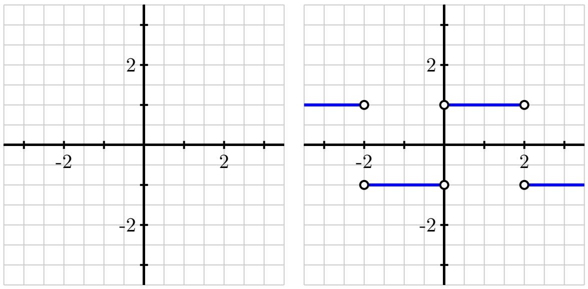
(a) On the axes provided at left below, sketch a possible graph of $y = f(x)$. Explain why your graph meets the stated criteria.

(b) On the axes at right below, sketch a possible graph of $y = f'(x)$. What type of curve does the provided data suggest for the graph of $y = f'(x)$?

(c) Conjecture a formula for the function $y = f(x)$. Use the limit definition of the derivative to determine the corresponding formula for $y = f'(x)$. Discuss both graphical and algebraic evidence for whether or not your conjecture is correct.



11) Let g be a continuous function (that is, one with no jumps or holes in the graph) and suppose that a graph of $y = g'(x)$ is given by the graph on the right below.



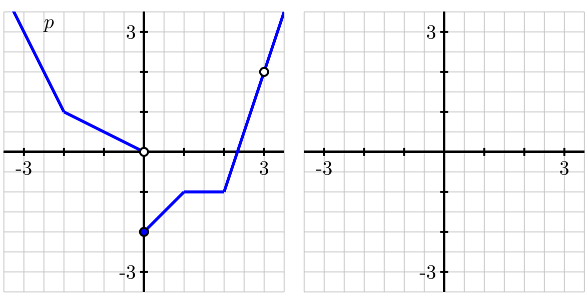
(a) Observe that for every value of x that satisfies $0 < x < 2$, the value of $g'(x)$ is constant. What does this tell you about the behavior of the graph of $y = g(x)$ on this interval?

(b) On what intervals other than $0 < x < 2$ do you expect $y = g(x)$ to be a linear function? Why?

(c) At which values of x is $g'(x)$ not defined? What behavior does this lead you to expect to see in the graph of $y = g(x)$?

(d) Suppose that $g(0) = 1$. On the axes provided at left, sketch an accurate graph of $y = g(x)$.

12) Consider the graph of the function $y = p(x)$ that is provided below on the left. Assume that each portion of the graph of p is a straight line, as pictured.



(a) State all values of a for which $\lim_{x \to a} p(x)$ does not exist.

(b) State all values of a for which p is not continuous at a.

(c) State all values of a for which p is not differentiable at $x = a$.

(d) On the axes provided on the right, sketch an accurate graph of $y = p'(x)$.

13) For each of the following prompts, give an example of a function that satisfies the stated criteria. A formula or a graph, with reasoning, is sufficient for each. If no such example is possible, explain why.

(a) A function f that is continuous at $a = 2$ but not differentiable at $a = 2$.

(b) A function g that is differentiable at $a = 3$ but does not have a limit at $a = 3$.

(c) A function h that has a limit at $a = -2$, is defined at $a = -2$, but is not continuous at $a = -2$.

(d) A function p that satisfies all of the following:

- $p(-1) = 3$ and $\lim_{x \to -1} p(x) = 2$
- $p(0) = 1$ and $p'(0) = 0$
- $\lim_{x \to 1} p(x) = p(1)$ and $p'(1)$ does not exist

For exercises 14–22, use the limit definition of the derivative to compute the derivative function.

14) $f(x) = 8x$

15) $y = x^2$

16) $g(x) = 2x^2 + 3x$

17) $s(t) = \dfrac{1}{\sqrt{t}}$

18) $r(x) = \dfrac{1}{x^2}$

19) $f(x) = \sqrt{3x - 1}$

20) $f(x) = \sqrt{8x}$

21) $s(t) = \dfrac{1}{t+5}$

22) $y = \dfrac{1}{2x - 1}$

2.3 *The second derivative*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- How does the derivative of a function tell us whether the function is increasing or decreasing at a point or on an interval?
- What can we learn by taking the derivative of the derivative (to achieve the *second* derivative) of a function f?
- What does it mean to say that a function is concave up or concave down? How are these characteristics connected to certain properties of the derivative of the function?
- What are the units on the second derivative? How do they help us understand the rate of change of the rate of change?

Introduction

Given a differentiable function $y = f(x)$, we know that its derivative, $y = f'(x)$, is a related function whose output at a value $x = a$ tells us the slope of the tangent line to $y = f(x)$ at the point $(a, f(a))$. That is, heights on the derivative graph tell us the values of slopes on the original function's graph. Therefore, the derivative tells us important information about the function f.

At any point where $f'(x)$ is positive, it means that the slope of the tangent line to f is positive, and therefore the function f is increasing (or rising) at that point. Similarly, if $f'(a)$ is negative, we know that the graph of f is decreasing (or falling) at that point.

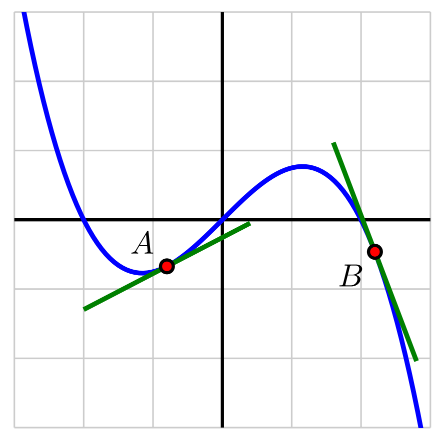


Figure 2.14: Two tangent lines on a graph demonstrate how the slope of the tangent line tells us whether the function is rising or falling, as well as whether it is doing so rapidly or slowly.

In the next part of our study, we work to understand not only *whether* the function f is increasing or decreasing at a point or on an interval, but also *how* the function f is increasing or decreasing. Comparing the two tangent lines shown in Figure 2.14, we see that at point A, the value of $f'(x)$ is positive and relatively close to zero, which coincides with the graph rising slowly. By contrast, at point B, the derivative is negative and relatively large in absolute value, which is tied to the fact that f is decreasing rapidly at B. It also makes sense to not only ask whether the value of the derivative function is positive or negative and whether the derivative is large or small, but also to ask "how is the derivative changing?"

We also now know that the derivative, $y = f'(x)$, is itself a function. This means that we can consider taking its derivative – the derivative of the derivative – and therefore ask questions like "what does the derivative of the derivative tell us about how the original function behaves?" As we have done regularly in our work to date, we start with an investigation of a familiar problem in the context of a moving object.

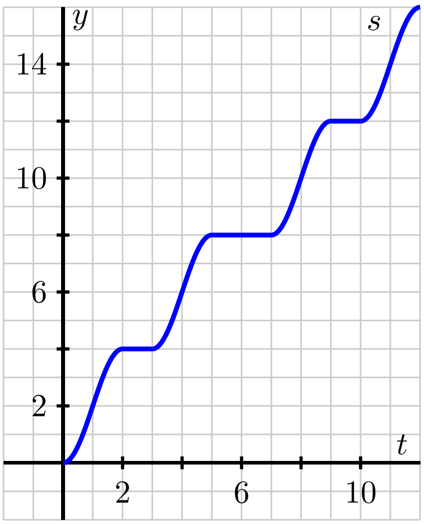


Figure 2.15: The graph of $y = s(t)$, the position of the car (measured in thousands of feet from its starting location) at time t in minutes.

Preview Activity 2.3

The position of a car driving along a straight road at time t in minutes is given by the function $y = s(t)$ that is pictured in Figure 2.15. The car's position function has units measured in thousands of feet. For instance, the point $(2, 4)$ on the graph indicates that after 2 minutes, the car has traveled 4000 feet.

(a) In everyday language, describe the behavior of the car over the provided time interval. In particular, you should carefully discuss what is happening on each of the time intervals $[0, 1]$, $[1, 2]$, $[2, 3]$, $[3, 4]$, and $[4, 5]$, plus provide commentary overall on what the car is doing on the interval $[0, 12]$.

(b) On the lefthand axes provided in Figure 2.16, sketch a careful, accurate graph of $y = s'(t)$.

(c) What is the meaning of the function $y = s'(t)$ in the context of the given problem? What can we say about the car's behavior when $s'(t)$ is positive? when $s'(t)$ is zero? when $s'(t)$ is negative?

(d) Rename the function you graphed in (b) to be called $y = v(t)$. Describe the behavior of v in words, using phrases like "v is increasing on the interval ..." and "v is constant on the interval ...."

(e) Sketch a graph of the function $y = v'(t)$ on the righthand axes provide in Figure 2.16. Write at least one sentence to explain how the behavior of $v'(t)$ is connected to the graph of $y = v(t)$.

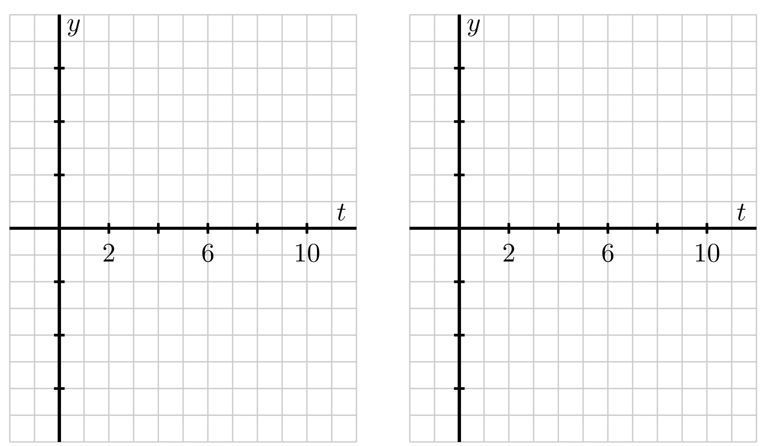


Figure 2.16: Axes for plotting $y = v(t) = s'(t)$ and $y = v'(t)$.

Increasing, decreasing, or neither

When we look at the graph of a function, there are features that strike us naturally, and common language can be used to name these features. In many different settings so far, we have intuitively used the words *increasing* and *decreasing* to describe a function's graph. Here we connect these terms more formally to a function's behavior on an interval of input values.

Increasing/Decreasing

Given a function $f(x)$ defined on the interval (a, b), we say that f *is increasing on* (a, b) provided that for all x, y in the interval (a, b), if $x < y$, then $f(x) < f(y)$. Similarly, we say that f *is decreasing on* (a, b) provided that for all x, y in the interval (a, b), if $x < y$, then $f(x) > f(y)$.

Simply put, an increasing function is one that is rising as we move from left to right along the graph, and a decreasing function is one that falls as the value of the input increases. For a function that has a derivative at a point, we will also talk about whether or not the function is increasing or decreasing *at that point*. Moreover, the fact of whether or not the function is increasing, decreasing, or neither at a given point depends precisely on the value of the derivative at that point.

Increasing/Decreasing at a Point

Let f be a function that is differentiable at $x = a$. Then f is increasing at $x = a$ if and only if $f'(a) > 0$ and f is decreasing at $x = a$ if and only if $f'(a) < 0$. If $f'(a) = 0$, then we say f is neither increasing nor decreasing at $x = a$.

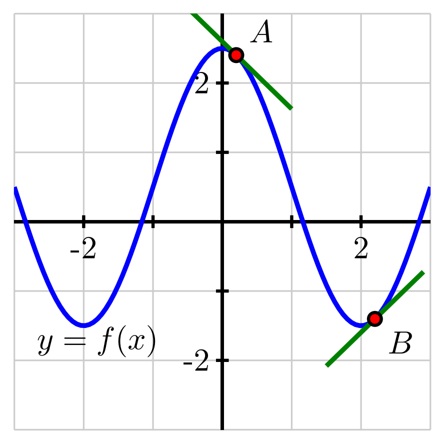


Figure 2.17: A function that is decreasing at A, increasing at B, and more generally, decreasing on the intervals $-3 < x < -2$ and $0 < x < 2$ and increasing on $-2 < x < 0$ and $2 < x < 3$.

For example, the function pictured in Figure 2.17 is increasing at any point at which $f'(x)$ is positive, and hence is increasing on the entire interval $-2 < x < 0$. Note that at both $x = \pm 2$ and $x = 0$, we say that f is neither increasing nor decreasing, because $f'(x) = 0$ at these values.

Example 1

Let $f(x) = x^3 + x^2 - x + 1$. Find intervals on which f is increasing or decreasing.

Solution. We first find the values where f is neither increasing or decreasing. Given $f'(x) = 3x^2 + 2x - 1 = (3x - 1)(x + 1)$, then $f'(x) = 0$ when $x = -1$ and when $x = 1/3$. Notice that f' is never undefined.

Since an interval was not specified for us to consider, we consider the entire domain of f, which is $(-\infty, \infty)$. We thus break the whole real line into three subintervals based on the two values where the derivative is

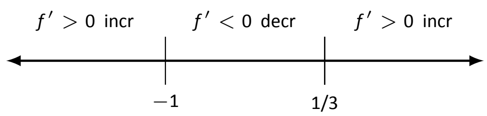


Figure 2.18: Number line for f in Example 1.

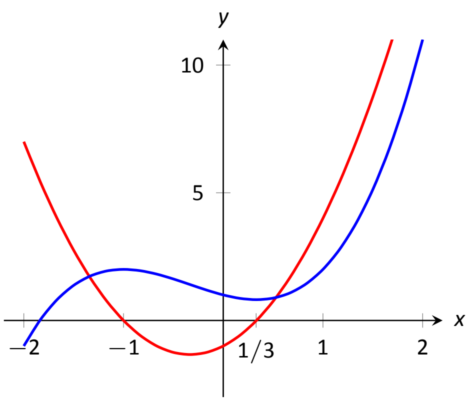


Figure 2.19: A graph of $f(x)$ in Example 1, showing where f is increasing and decreasing.

zero: $(-\infty,-1)$, $(-1,1/3)$ and $(1/3,\infty)$. This is shown in Figure 2.18.

We now pick a value p in each subinterval and find the sign of $f'(p)$. All we care about is the sign, so we do not actually have to fully compute $f'(p)$; pick "nice" values that make this simple.

Subinterval 1, $(-\infty,-1)$: We (arbitrarily) pick $p=-2$. We can compute $f'(-2)$ directly: $f'(-2)=3(-2)^2+2(-2)-1=7>0$. We conclude that f is increasing on $(-\infty,-1)$.

Note we can arrive at the same conclusion without computation. For instance, we could choose $p=-100$. The first term in $f'(-100)$, i.e., $3(-100)^2$ is clearly positive and very large. The other terms are small in comparison, so we know $f'(-100)>0$. All we need is the sign.

Subinterval 2, $(-1,1/3)$: We pick $p=0$ since that value seems easy to deal with. $f'(0)=-1<0$. We conclude f is decreasing on $(-1,1/3)$.

Subinterval 3, $(1/3,\infty)$: Pick an arbitrarily large value for $p>1/3$ and note that $f'(p)=3p^2+2p-1>0$. We conclude that f is increasing on $(1/3,\infty)$.

We can verify our calculations by considering Figure 2.19, where f is graphed in blue. The graph also presents f' in red; note how $f'>0$ when f is increasing and $f'<0$ when f is decreasing.

The Second Derivative

For any function, we are now accustomed to investigating its behavior by thinking about its derivative. Given a function f, its derivative is a new function, one that is given by the rule

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

Because f' is itself a function, it is perfectly feasible for us to consider the derivative of the derivative, which is the new function $y=[f'(x)]'$. We call this resulting function *the second derivative* of $y=f(x)$, and denote the second derivative by $y=f''(x)$. Due to the presence of multiple possible derivatives, we will sometimes call f' “the first derivative” of f, rather than simply “the derivative” of f. Formally, the second derivative is defined by the limit definition of the derivative of the first derivative:

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

We note that all of the established meaning of the derivative function still holds, so when we compute $y=f''(x)$, this new function measures slopes of tangent lines to the curve $y=f'(x)$, as well as the instantaneous rate of change of $y=f'(x)$. In other words, just as the first derivative measures the rate at which the original function changes, the second derivative measures the rate at which the first derivative changes. This means that the second derivative tracks the instantaneous rate of change of the instantaneous rate of change of f. That is, the second derivative will help us to understand how the rate of change of the original function is itself changing.

Concavity

In addition to asking *whether* a function is increasing or decreasing, it is also natural to inquire *how* a function is increasing or decreasing. To begin, there are three basic behaviors that an increasing function can demonstrate on an interval, as pictured in Figure 2.20: the function can increase more and more rapidly, increase at the same rate, or increase in a way that is slowing down. Fundamentally, we are beginning to think about how a particular curve bends, with the natural comparison being made to lines, which don't bend at all. More than this, we want to understand how the bend in a function's graph is tied to behavior characterized by the first derivative of the function.

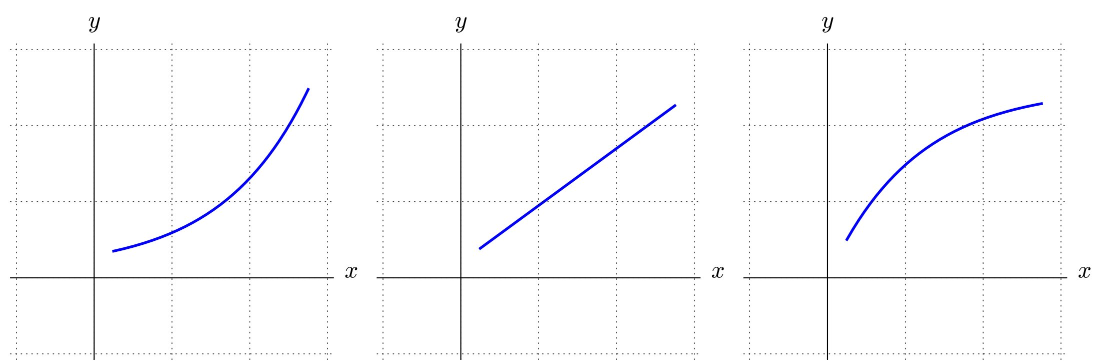


Figure 2.20: Three functions that are all increasing, but doing so at an increasing rate, at a constant rate, and at a decreasing rate, respectively.

For the leftmost curve in Figure 2.20, picture a sequence of tangent lines to the curve. As we move from left to right, the slopes of those tangent lines will increase. Therefore, the rate of change of the pictured function is increasing, and this explains why we say this function is *increasing at an increasing rate*. For the rightmost graph in Figure 2.20, observe that as x increases, the function increases but the slope of the tangent line decreases, hence this function is *increasing at a decreasing rate*.

Of course, similar options hold for how a function can decrease. Here we must be extra careful with our language, since decreasing functions involve negative slopes, and negative numbers present an interesting situation in the tension between common language and mathematical language. For example, it can be tempting to say that "-100 is bigger than -2." But we must remember that when we say one number is greater than another, this describes how the numbers lie on a number line: $x < y$ provided that x lies to the left of y. So of course, -100 is less than -2. Informally, it might be helpful to say that "-100 is more negative than -2." This leads us to note particularly that when a function's values are negative, and those values subsequently get more negative, the function must be decreasing.

Now consider the three graphs shown in Figure 2.21. Clearly the middle graph demonstrates the behavior of a function decreasing at a constant rate. If we think about a sequence of tangent lines to the first curve that progress from left to right, we see that the slopes of these lines get less and less negative as we move from left to right. That means that the values of the first derivative, while all negative, are increasing, and thus we say that the leftmost curve is *decreasing at an increasing rate*.

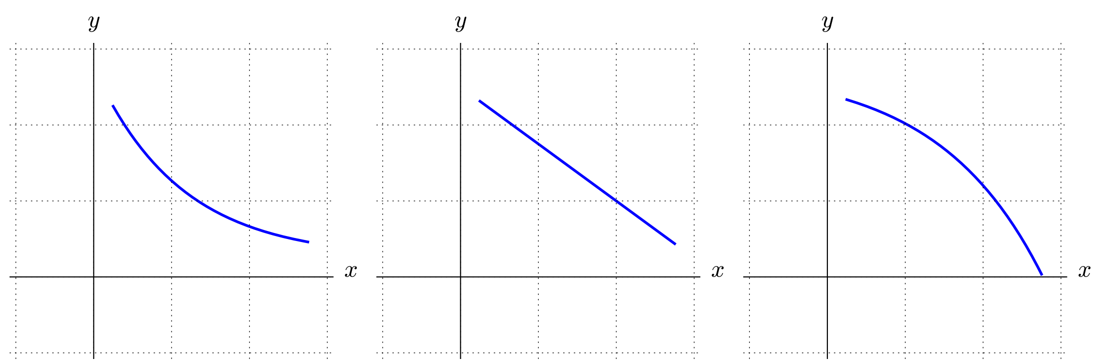


Figure 2.21: From left to right, three functions that are all decreasing, but doing so in different ways.

This leaves only the rightmost curve in Figure 2.21 to consider. For that function, the slope of the tangent line is negative throughout the pictured interval, but as we move from left to right, the slopes get more and more negative. Hence the slope of the curve is decreasing, and we say that the function is *decreasing at a decreasing rate*.

This leads us to introduce the notion of *concavity* which provides simpler language to describe some of these behaviors. Informally, when a curve opens up on a given interval, like the upright parabola $y = x^2$ or the exponential growth function $y = e^x$, we say that the curve is *concave up* on that interval. Likewise, when a curve opens down, such as the parabola $y = -x^2$ or the opposite of the exponential function $y = -e^x$, we say that the function is *concave down*. This behavior is linked to both the first and second derivatives of the function.

In Figure 2.22, we see two functions along with a sequence of tangent lines to each. On the lefthand plot where the function

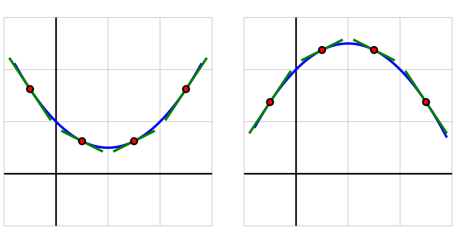


Figure 2.22: At left, a function that is concave up; at right, one that is concave down.

is concave up, observe that the tangent lines to the curve always lie below the curve itself and that, as we move from left to right, the slope of the tangent line is increasing. Said differently, the function f is concave up on the interval shown because its derivative, f', is increasing on that interval. Similarly, on the righthand plot in Figure 2.22, where the function shown is concave down, there we see that the tangent lines alway lie above the curve and that the value of the slope of the tangent line is decreasing as we move from left to right. Hence, what makes f concave down on the interval is the fact that its derivative, f', is decreasing.

We state these most recent observations formally as the definitions of the terms *concave up* and *concave down*.

Concave Up/Down

Let f be a differentiable function on an interval (a, b). Then f is *concave up* on (a, b) if and only if f' is increasing on (a, b); f is *concave down* on (a, b) if and only if f' is decreasing on (a, b).

Example 2

Let $f(x) = x^3 - 3x + 1$. Find the intervals on which f is concave up/down.

Solution. We first find the values where f is neither increasing or decreasing. Given $f'(x) = 3x^2 - 3$ and $f''(x) = 6x$, then $f''(x) = 0$ only when $x = 0$. Notice that f'' is never undefined.

Since an interval was not specified for us to consider, we consider the entire domain of f, which is $(-\infty, \infty)$. We use a process similar to the one used in Example 1 to determine the intervals where f is increasing/decreasing. Thus break the whole real line into two subintervals based on the value where the second derivative is zero: $(-\infty, 0)$ and $(0, \infty)$.

Subinterval 1, $(-\infty, 0)$: Picking any value $c < 0$, we can easily see that $f''(c) < 0$; so f is concave down on $(-\infty, 0)$.

Subinterval 2, $(0, \infty)$: Picking any value $c > 0$, we can easily see that $f''(c) > 0$; so f is concave up on $(0, \infty)$.

The number line in Figure 2.23 illustrates the process of determining concavity; Figure 2.23 shows a graph of f in blue and $f''(x)$ in red, confirming our results. Notice how f is concave down precisely when $f''(x) < 0$ and concave up when $f''(x) > 0$.

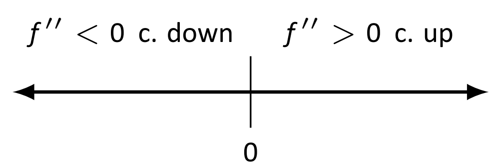


Figure 2.23: A number line determining the concavity of f in Example 2.

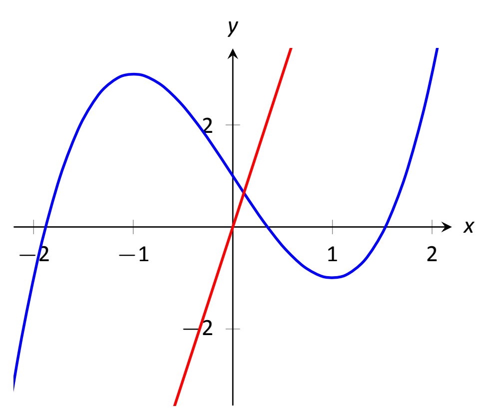


Figure 2.24: A graph of $f(x)$ used in Example 2.

The following activities lead us to further explore how the first and second derivatives of a function determine the behavior and shape of its graph. We begin by revisiting Preview Activity 2.3.

Activity 2.3–1

The position of a car driving along a straight road at time t in minutes is given by the function $y = s(t)$ that is pictured in Figure 2.25. The car's position function has units measured in thousands of feet. Remember that you worked with this function and sketched graphs of $y = v(t) = s'(t)$ and $y = v'(t)$ in Preview Activity 2.3.

(a) On what intervals is the position function $y = s(t)$ increasing? decreasing? Why?

(b) On which intervals is the velocity function $y = v(t) = s'(t)$ increasing? decreasing? neither? Why?

(c) *Acceleration* is defined to be the instantaneous rate of change of velocity, as the acceleration of an object measures the rate at which the velocity of the object is changing. Say that the car's acceleration function is named $a(t)$. How is $a(t)$ computed from $v(t)$? How is $a(t)$ computed from $s(t)$? Explain.

(d) What can you say about s'' whenever s' is increasing? Why?

(e) Using only the words *increasing*, *decreasing*, *constant*, *concave up*, *concave down*, and *linear*, complete the following sentences. For the position function s with velocity v and acceleration a,

- on an interval where v is positive, s is ____________________.
- on an interval where v is negative, s is ____________________.
- on an interval where v is zero, s is ____________________.
- on an interval where a is positive, v is ____________________.
- on an interval where a is negative, v is ____________________.
- on an interval where a is zero, v is ____________________.
- on an interval where a is positive, s is ____________________.
- on an interval where a is negative, s is ____________________.
- on an interval where a is zero, s is ____________________.

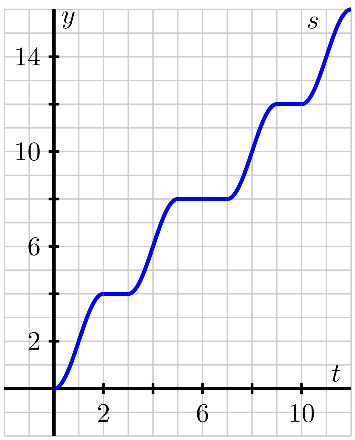


Figure 2.25: The graph of $y = s(t)$, the position of the car (measured in thousands of feet from its starting location) at time t in minutes.

The context of position, velocity, and acceleration is an excellent one in which to understand how a function, its first derivative, and its second derivative are related to one another. In Activity 2.5–1, we can replace s, v, and a with an arbitrary function f and its derivatives f' and f'', and essentially all the same observations hold. In particular, note that f' is increasing if and only if both f is concave up, and similarly f' is increasing if and only if f'' is positive. Likewise, f' is decreasing if and only if both f is concave down, and f' is decreasing if and only if f'' is negative.

Activity 2.3–2

A potato is placed in an oven, and the potato's temperature F (in degrees Fahrenheit) at various points in time is taken and recorded in Table 2.2. Time t is measured in minutes. In Activity 2.1–5, we computed approximations to $F'(30)$ and $F'(60)$ using central differences. Those values and more are also provided in Table 2.2, along with several others computed in the same way.

(a) What are the units on the values of $F'(t)$?

(b) Use a central difference to estimate the value of $F''(30)$.

(c) What is the meaning of the value of $F''(30)$ that you have computed in (b) in terms of the potato's temperature? Write several careful sentences that discuss, with appropriate units, the values of $F(30)$, $F'(30)$, and $F''(30)$, and explain the overall behavior of the potato's temperature at this point in time.

(d) Overall, is the potato's temperature increasing at an increasing rate, increasing at a constant rate, or increasing at a decreasing rate? Why?

t	$F(t)$
0	70
15	180.5
30	251
45	296
60	324.5
75	342.8
90	354.5

t	$F'(t)$
0	NA
15	6.03
30	3.85
45	2.45
60	1.56
75	1.00
90	NA

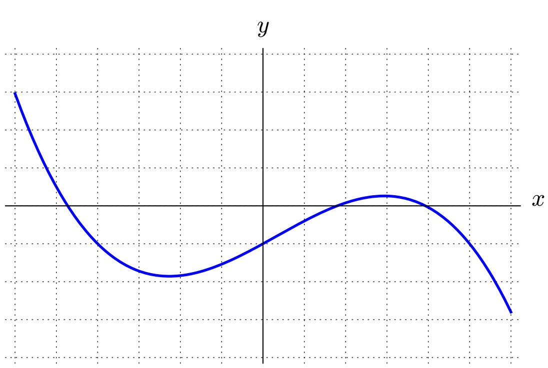
Table 2.2: The temperatures and rates of change of the temperature of a potato in an oven.

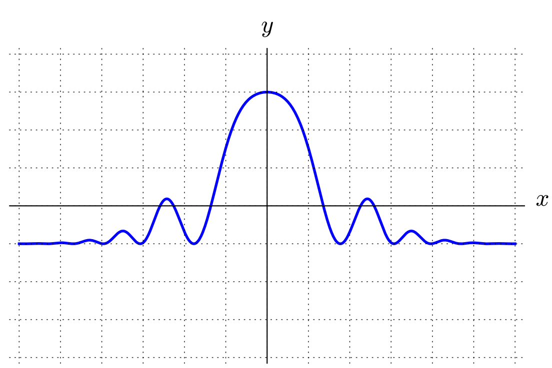
Activity 2.3–3

This activity builds on our experience and understanding of how to sketch the graph of f' given the graph of f. Below, given the graph of a function f, sketch f' and then sketch f''. In addition, for each, write several careful sentences in the spirit of those in Activity 2.5–1 that connect the behaviors of f, f', and f''. For instance, write something such as

> "f' is ____________________ on the interval _______, which is connected to the fact that f is ____________________ on the same interval _______, and f'' is ____________________ on the interval as well..."

but of course with the blanks filled in. Throughout, view the scale of the grid for the graph of f as being 1×1, and assume the horizontal scale of the grid for the graph of f' is identical to that for f. If you need to adjust the vertical scale on the axes for the graph of f' or f'', you should label that accordingly.





Higher Derivatives

So far in this section we have discussed how the derivative of a function f is itself a function, therefore we can take its derivative. We can repeat this process as long as the corresponding limits exist. The following concept introduces the notation we use for higher derivatives.

Higher Order Derivatives

Let $y = f(x)$ be a differentiable function on I.

1) The *second derivative* of f is:

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

2) The *third derivative* of f is:

$$f'''(x) = \frac{d}{dx}\Big(f''(x)\Big) = \frac{d}{dx}\left(\frac{d^2y}{dx^2}\right) = \frac{d^3y}{dx^3} = y'''.$$

3) The n^{th} *derivative* of f is:

$$f^{(n)}(x) = \frac{d}{dx}\left(f^{(n-1)}(x)\right) = \frac{d}{dx}\left(\frac{d^{n-1}y}{dx^{n-1}}\right) = \frac{d^ny}{dx^n} = y^{(n)}.$$

In general, when finding the fourth derivative and on, we resort to the $f,^{(4)}(x)$ notation, not $f''''(x)$; after a while, too many ticks is too confusing. Let's practice using this new concept.

Example 3

Find the first four derivatives of $f(x) = 4x^2$.

Solution. Using the limit definition of the derivative, we compute the first derivative as

$$\begin{aligned}
f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}\\
&= \lim_{h\to 0}\frac{4(x+h)^2-4x^2}{h} = \lim_{h\to 0}\frac{4(x^2+2xh+h^2)-4x^2}{h}\\
&= \lim_{h\to 0}\frac{4x^2+8xh+4h^2-4x^2}{h} = \lim_{h\to 0}\frac{8xh+4h^2}{h}\\
&= \lim_{h\to 0}\frac{h(8x+4h)}{h} = \lim_{h\to 0} 8x+4h = 8x.
\end{aligned}$$

We compute the second derivative as

$$\begin{aligned}
f''(x) &= \lim_{h\to 0} \frac{f'(x+h)-f'(x)}{h} \\
&= \lim_{h\to 0} \frac{8(x+h)-8x}{h} = \lim_{h\to 0} \frac{8x+8h-8x}{h} \\
&= \lim_{h\to 0} \frac{8h}{h} = \lim_{h\to 0} 8 = 8.
\end{aligned}$$

We compute the third derivative as

$$\begin{aligned}
f'''(x) &= \lim_{h\to 0} \frac{f''(x+h)-f''(x)}{h} \\
&= \lim_{h\to 0} \frac{8-8}{h} = \lim_{h\to 0} \frac{0}{h} = 0.
\end{aligned}$$

We compute the fourth derivative as

$$\begin{aligned}
f^{(4)}(x) &= \lim_{h\to 0} \frac{f'''(x+h)-f'''(x)}{h} \\
&= \lim_{h\to 0} \frac{0-0}{h} = \lim_{h\to 0} \frac{0}{h} = 0.
\end{aligned}$$

It can be difficult to consider the meaning of the third, and higher order, derivatives. The third derivative is "the rate of change of the rate of change of the rate of change of f." That is essentially meaningless to the uninitiated. In the context of our position/velocity/acceleration example, the third derivative is the "rate of change of acceleration," commonly referred to as "jerk."

Make no mistake: higher order derivatives have great importance even if their practical interpretations are hard (or "impossible") to understand. The mathematical topic of *series* makes extensive use of higher order derivatives.

Summary

In this section, we encountered the following important ideas:

- A differentiable function f is increasing at a point or on an interval whenever its first derivative is positive, and decreasing whenever its first derivative is negative.
- By taking the derivative of the derivative of a function f, we arrive at the second derivative, f''. The second derivative measures the instantaneous rate of change of the first derivative, and thus the sign of the second derivative tells us whether or not the slope of the tangent line to f is increasing or decreasing.
- A differentiable function is concave up whenever its first derivative is increasing (or equivalently whenever its second derivative is positive), and concave down whenever its first derivative is decreasing (or equivalently whenever its second derivative is negative).
- The units on the second derivative are "units of output per unit of input per unit of input." They tell us how the value of the derivative function is changing in response to changes in the input. In other words, the second derivative tells us the rate of change of the rate of change of the original function.

Exercises

Terms and Concepts

1) Explain in your own words what the second derivative "means."

2) If $f(x)$ describes a position function, then $f'(x)$ describes what kind of function? What kind of function is $f'(x)$?

3) Let $f(x)$ be a function measured in pounds, where x is measured in feet. What are the units of $f''(x)$?

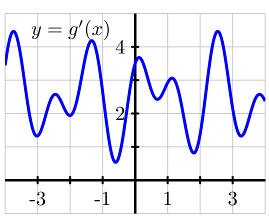
4) Explain in your own words how to find the third derivative of a function $f(x)$.

Problems

In exercises 5–8, a graph of a function is given. Using the graph, sketch $f'(x)$.

5) Suppose that $y = f(x)$ is a differentiable function for which the following information is known: $f(2) = -3$, $f'(2) = 1.5$, $f''(2) = -0.25$.
   (a) Is f increasing or decreasing at $x = 2$? Is f concave up or concave down at $x = 2$?
   (b) Do you expect $f(2.1)$ to be greater than -3, equal to -3, or less than -3? Why?
   (c) Do you expect $f'(2.1)$ to be greater than 1.5, equal to 1.5, or less than 1.5? Why?
   (d) Sketch a graph of $y = f(x)$ near $(2, f(2))$ and include a graph of the tangent line.

6) For a certain function $y = g(x)$, its derivative is given by the function pictured below.



   (a) What is the approximate slope of the tangent line to $y = g(x)$ at the point $(2, g(2))$?
   (b) How many real number solutions can there be to the equation $g(x) = 0$? Justify your conclusion fully and carefully by explaining what you know about how the graph of g must behave based on the given graph of g'.
   (c) On the interval $-3 < x < 3$, how many times does the concavity of g change? Why?
   (d) Use the provided graph to estimate the value of $g''(2)$.

7) For each prompt that follows, sketch a possible graph of a function on the interval $-3 < x < 3$ that satisfies the stated properties.
   (a) $y = f(x)$ such that f is increasing on $-3 < x < 3$, f is concave up on $-3 < x < 0$, and f is concave down on $0 < x < 3$.
   (b) $y = g(x)$ such that g is increasing on $-3 < x < 3$, g is concave down on $-3 < x < 0$, and g is concave up on $0 < x < 3$.
   (c) $y = h(x)$ such that h is decreasing on $-3 < x < 3$, h is concave up on $-3 < x < -1$, neither concave up nor concave down on $-1 < x < 1$, and h is concave down on $1 < x < 3$.
   (d) $y = p(x)$ such that p is decreasing and concave down on $-3 < x < 0$ and p is increasing and concave down on $0 < x < 3$.

8) A bungee jumper's height h (in feet) at time t (in seconds) is given in the table below.
   (a) Use the given data to estimate $h'(4.5)$, $h'(5)$, and $h'(5.5)$. At which of these times is the bungee jumper rising most rapidly?
   (b) Use the given data and your work in (a) to estimate $h''(5)$.
   (c) What physical property of the bungee jumper does the value of $h''(5)$ measure? What are its units?
   (d) Based on the data, on what approximate time intervals is the function $y = h(t)$ concave down? What is happening to the velocity of the bungee jumper on these time intervals?

t	$h(t)$
0.0	200
0.5	184.2
1.0	159.9
1.5	131.9
2.0	104.7
2.5	81.8
3.0	65.5
3.5	56.8
4.0	55.5
4.5	60.4
5.0	69.8

t	$h(t)$
5.5	81.6
6.0	93.7
6.5	104.4
7.0	112.6
7.5	117.7
8.0	119.4
8.5	118.2
9.0	114.8
9.5	110.0
10.0	104.7

2.4 *Basic derivative rules*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- What are alternate notations for the derivative?
- How can we sometimes use the algebraic structure of a function $f(x)$ to easily compute a formula for $f'(x)$?
- What is the derivative of a power function of the form $f(x) = x^n$? What is the derivative of an exponential function of form $f(x) = a^x$?
- If we know the derivative of $y = f(x)$, how is the derivative of $y = kf(x)$ computed, where k is a constant?
- If we know the derivatives of $y = f(x)$ and $y = g(x)$, how is the derivative of $y = f(x) + g(x)$ computed?
- What do the graphs of $y = \sin(x)$ and $y = \cos(x)$ suggest as formulas for their respective derivatives?

Introduction

So far in this chapter, we developed the concept of the derivative of a function. We now know that the derivative f' of a function f measures the instantaneous rate of change of f with respect to x as well as the slope of the tangent line to $y = f(x)$ at any given value of x. To date, we have focused primarily on interpreting the derivative graphically or, in the context of functions in a physical setting, as a meaningful rate of change. To actually calculate the value of the derivative at a specific point, we have typically relied on the limit definition of the derivative.

In this section, we will investigate how the limit definition of the derivative

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

leads to interesting patterns and rules that enable us to quickly find a formula for $f'(x)$ based on the formula for $f(x)$ *without* using the limit definition directly. For example, we already know that if $f(x) = x$, then it follows that $f'(x) = 1$. While we could use the limit definition of the derivative to confirm this, we know it to be true because $f(x)$ is a linear function with slope 1 at every value of x. One of our goals is to be able to take standard functions, say ones such as $g(x) = 4x^7 - \sin(x) + 3e^x$, and, based on the algebraic form of the function, be able to apply shortcuts to almost immediately determine the formula for $g'(x)$.

Preview Activity 2.4

Functions of the form $f(x) = x^n$, where $n = 1, 2, 3, \ldots$, are often called *power functions*. The first two questions below revisit work we did earlier, and the following questions extend those ideas to higher powers of x.

(a) Use the limit definition of the derivative to find $f'(x)$ for $f(x) = x^2$.

(b) Use the limit definition of the derivative to find $f'(x)$ for $f(x) = x^3$.

(c) Use the limit definition of the derivative to find $f'(x)$ for $f(x) = x^4$. (Hint: $(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$. Apply this rule to $(x+h)^4$ within the limit definition.)

(d) Based on your work in (a), (b), and (c), what do you conjecture is the derivative of $f(x) = x^5$? Of $f(x) = x^{13}$?

(e) Conjecture a formula for the derivative of $f(x) = x^n$ that holds for any positive integer n. That is, given $f(x) = x^n$ where n is a positive integer, what do you think is the formula for $f'(x)$?

Constant, Power, and Exponential Functions

So far, we know the derivative formula for two important classes of functions: constant functions and power functions. For the first kind, observe that if $f(x) = c$ is a constant function, then its graph is a horizontal line with slope zero at every point. Thus, $\frac{d}{dx}[c] = 0$. We summarize this with the following rule.

Constant Rule

For any real number c, if $f(x) = c$, then

$$f'(x) = 0.$$

The proof of the Constant Rule is found by applying the definition of the derivative, see Exercise 28).

Thus, if $f(x) = 7$, then $f'(x) = 0$. Similarly, $\frac{d}{dx}[\sqrt{3}] = 0$.

For power functions, from your work in Preview Activity 2.4, you have conjectured that for any positive integer n, if $f(x) = x^n$, then $f'(x) = nx^{n-1}$. Not only can this rule be formally proved to hold for any positive integer n (see Activity 2.4–1), but also for any nonzero real number (see Section **??**).

Power Rule

For any nonzero real number, if $f(x) = x^n$, then

$$f'(x) = nx^{n-1}.$$

This rule for power functions allows us to find derivatives such as the following: if $g(z) = z^{-3}$, then $g'(z) = -3z^{-4}$. Similarly, if $h(t) = t^{7/5}$, then $\frac{dh}{dt} = \frac{7}{5}t^{2/5}$; likewise, $\frac{d}{dq}[q^{\pi}] = \pi q^{\pi-1}$.

Activity 2.4–1

Let $f(x) = x^n$, where n is a positive integer. The goal of this problem is to prove the Power Rule for derivatives, that we used in earlier work in this section.

(a) State the Binomial Theorem.

(b) Using the Binomial Theorem, expand $(x+h)^n$ for some positive integer n.

(c) Use the limit definition of the derivative along with the Binomial Theorem to show that

$$f'(x) = \lim_{h \to 0} \frac{x^n + nx^{n-1}h + \frac{n(n-1)}{2}x^{n-2}h^2 \cdots h^n - x^n}{h}.$$

(d) Simplify the numerator of the derivative so that each term is multiplied by some expression in terms of h.

(e) Reduce and compute the limit using algebra.

As we next turn to thinking about derivatives of combinations of basic functions, it will be instructive to have one more type of basic function whose derivative formula we know. For now, we simply state this rule without explanation or justification; we will explore why this rule is true in Activity 2.4–3, first we will encounter graphical reasoning for why the rule is plausible in Activity 2.4–2.

Activity 2.4–2

Consider the function $g(x) = 2^x$, which is graphed in Figure 2.26.

(a) At each of $x = -2, -1, 0, 1, 2$, use a straightedge to sketch an accurate tangent line to $y = g(x)$.

(b) Use the provided grid to estimate the slope of the tangent line you drew at each point in (a).

(c) Use the limit definition of the derivative to estimate $g'(0)$ by using small values of h, and compare the result to your visual estimate for the slope of the tangent line to $y = g(x)$ at $x = 0$ in (b).

(d) Based on your work in (a), (b), and (c), sketch an accurate graph of $y = g'(x)$ on the axes adjacent to the graph of $y = g(x)$.

(e) Write at least one sentence that explains why it is reasonable to think that $g'(x) = cg(x)$, where c is a constant. In addition, calculate $\ln(2)$, and then discuss how this value, combined with your work above, reasonably suggests that $g'(x) = 2^x \ln(2)$.

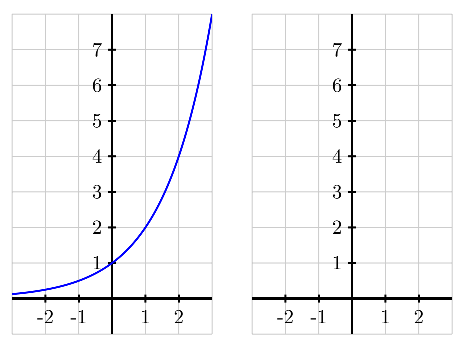


Figure 2.26: At left, the graph of $y = g(x) = 2^x$. At right, axes for plotting $y = g'(x)$.

Exponential Rule

For any positive real number a, if $f(x) = a^x$, then

$$f'(x) = a^x \ln(a).$$

For instance, this rule tells us that if $f(x) = 2^x$, then $f'(x) = 2^x \ln(2)$. Similarly, for $p(t) = 10^t$, $p'(t) = 10^t \ln(10)$. It is especially important to note that when $a = e$, where e is the base of the natural logarithm function, we have that

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

since $\ln(e) = 1$. This is an extremely important property of the function e^x: its derivative function is itself!

Activity 2.4–3

Let $f(x) = a^x$. The goal of this problem is to explore how the value of a affects the derivative of $f(x)$, without assuming we know the rule for $\frac{d}{dx}[a^x]$ that we have stated and used in earlier work in this section.

(a) Use the limit definition of the derivative to show that

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

(b) Explain why it is also true that

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

(c) Use computing technology and small values of h to estimate the value of

$$L = \lim_{h \to 0} \frac{a^h - 1}{h}$$

when $a = 2$. Do likewise when $a = 3$.

(d) Note that it would be ideal if the value of the limit L was 1, for then f would be a particularly special function: its derivative would be simply a^x, which would mean that its derivative is itself. By experimenting with different values of a between 2 and 3, try to find a value for a for which

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

(e) Compute $\ln(2)$ and $\ln(3)$. What does your work in (b) and (c) suggest is true about $\frac{d}{dx}[2^x]$ and $\frac{d}{dx}[2^x]$.

(f) How do your investigations in (d) lead to a particularly important fact about the number e?

Finally, note carefully the distinction between power functions and exponential functions: in power functions, the variable is in the base, as in x^2, while in exponential functions, the variable is in the power, as in 2^x. As we can see from the rules, this makes a big difference in the form of the derivative.

The following activity will check your understanding of the derivatives of the three basic types of functions noted above.

Activity 2.4–4

Use the three rules above to determine the derivative of each of the following functions. For each, state your answer using full and proper notation, labeling the derivative with its name. For example, if you are given a function $h(z)$, you should write "$h'(z) =$" or "$\frac{dh}{dz} =$" as part of your response.

(a) $f(t) = \pi$

(b) $g(z) = 7^z$

(c) $h(w) = w^{3/4}$

(d) $p(x) = 3^{1/2}$

(e) $r(t) = (\sqrt{2})^t$

(f) $\dfrac{d}{dq}[q^{-1}]$

(g) $m(t) = \dfrac{1}{t^3}$

Constant Multiples and Sums of Functions

Of course, most of the functions we encounter in mathematics are more complicated than being simply constant, a power of a variable, or a base raised to a variable power. In this section and several following, we will learn how to quickly compute the derivative of a function constructed as an algebraic combination of basic functions. For instance, we'd like to be able to understand how to take the derivative of a polynomial function such as $p(t) = 3t^5 - 7t^4 + t^2 - 9$, which is a function made up of constant multiples and sums of powers of t. To that end, we develop two new rules: the Constant Multiple Rule and the Sum Rule.

Say we have a function $y = f(x)$ whose derivative formula is known. How is the derivative of $y = kf(x)$ related to the derivative of the original function? Recall that when we multiply a function by a constant k, we vertically stretch the graph by a factor of $|k|$ (and reflect the graph across $y = 0$ if $k < 0$). This vertical stretch affects the slope of the graph, making the slope of the function $y = kf(x)$ be k times as steep as the slope of $y = f(x)$. In terms of the derivative, this is essentially saying that when we multiply a function by a factor of k, we change the value of its derivative by a factor of k as well. Thus, the Constant Multiple Rule holds:

The Constant Multiple Rule

For any real number k, if $f(x)$ is a differentiable function with derivative $f'(x)$, then

$$\frac{d}{dx}[kf(x)] = kf'(x).$$

The Constant Multiple Rule can be formally proved as a consequence of properties of limits, using the limit definition of the derivative, see Exercise 29)

In words, this rule says that "the derivative of a constant times a function is the constant times the derivative of the function." For example, if $g(t) = 3 \cdot 5^t$, we have $g'(t) = 3 \cdot 5^t \ln(5)$. Similarly, $\frac{d}{dz}[5z^{-2}] = 5(-2z^{-3})$.

Next we examine what happens when we take a sum of two functions. If we have $y = f(x)$ and $y = g(x)$, we can compute a new function $y = (f+g)(x)$ by adding the outputs of the two functions: $(f+g)(x) = f(x) + g(x)$. Not only does this result in the value of the new function being the sum of the values of the two known functions, but also the slope of the new function is the sum of the slopes of the known functions. Therefore, we arrive at the following Sum Rule for derivatives:

The Sum Rule:

If $f(x)$ and $g(x)$ are differentiable functions with derivatives $f'(x)$ and $g'(x)$ respectively, then

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

Like the Constant Multiple Rule, the Sum Rule can be formally proved as a consequence of properties of limits, using the limit definition of the derivative, see Exercise 30).

In words, the Sum Rule tells us that "the derivative of a sum is the sum of the derivatives." It also tells us that any time we take a sum of two differentiable functions, the result must also be differentiable. Furthermore, because we can view the difference function $y = (f-g)(x) = f(x) - g(x)$ as $y = f(x) + (-1 \cdot g(x))$, the Sum Rule and Constant Multiple Rules together tell us that $\frac{d}{dx}[f(x) + (-1 \cdot g(x))] = f'(x) - g'(x)$, or that "the derivative of a difference is the difference of the derivatives." Hence we can now compute derivatives of sums and differences of elementary functions. For instance, $\frac{d}{dw}(2^w + w^2) = 2^w \ln(2) + 2w$, and if $h(q) = 3q^6 - 4q^{-3}$, then $h'(q) = 3(6q^5) - 4(-3q^{-4}) = 18q^5 + 12q^{-4}$.

Activity 2.4–5

Use only the rules for constant, power, and exponential functions, together with the Constant Multiple and Sum Rules, to compute the derivative of each function below with respect to the given independent variable. Note well that we do not yet know any rules for how to differentiate the product or quotient of functions. This means that you may have to do some algebra first on the functions below before you can actually use existing rules to compute the desired derivative formula. In each case, label the derivative you calculate with its name using proper notation such as $f'(x)$, $h'(z)$, dr/dt, etc.

(a) $f(x) = x^{5/3} - x^4 + 2^x$

(b) $g(x) = 14e^x + 3x^5 - x$

(c) $h(z) = \sqrt{z} + \frac{1}{z^4} + 5^z$

(d) $r(t) = \sqrt{53}\, t^7 - \pi e^t + e^4$

(e) $s(y) = (y^2+1)(y^2-1)$

(f) $q(x) = \dfrac{x^3 - x + 2}{x}$

(g) $p(a) = 3a^4 - 2a^3 + 7a^2 - a + 12$

In the same way that we have shortcut rules to help us find derivatives, we introduce some language that is simpler and shorter. Often, rather than say "take the derivative of f," we'll instead say simply "differentiate f." This phrasing is tied to the notion of having a derivative to begin with: if the derivative exists at a point, we say "f is differentiable," which is tied to the fact that f can be differentiated.

As we work more and more with the algebraic structure of functions, it is important to strive to develop a big picture view of what we are doing. Here, we can note several general observations based on the rules we have so far. One is that the derivative of any polynomial function will be another polynomial function, and that the degree of the derivative is one less than the degree of the original function. For instance, if $p(t) = 7t^5 - 4t^3 + 8t$, p is a degree 5 polynomial, and its derivative, $p'(t) = 35t^4 - 12t^2 + 8$, is a degree 4 polynomial. Additionally, the derivative of any exponential function is another exponential function: for example, if $g(z) = 7 \cdot 2^z$, then $g'(z) = 7 \cdot 2^z \ln(2)$, which is also exponential.

Furthermore, while our current emphasis is on learning shortcut rules for finding derivatives without directly using the limit definition, we should be certain not to lose sight of the fact that all of the meaning of the derivative still holds that we developed earlier in this chapter. That is, anytime we compute a derivative, that derivative measures the instantaneous rate of change of the original function, as well as the slope of the tangent line at any selected point on the curve. The following activity asks you to combine the just-developed derivative rules with some key perspectives that we studied earlier.

Activity 2.4–6

Each of the following questions asks you to use derivatives to answer key questions about functions. Be sure to think carefully about each question and to use proper notation in your responses.

(a) Find the slope of the tangent line to $h(z) = \sqrt{z} + \frac{1}{z}$ at the point where $z = 4$.

(b) A population of cells is growing in such a way that its total number in millions is given by the function $P(t) = 2(1.37)^t + 32$, where t is measured in days.

i. Determine the instantaneous rate at which the population is growing on day 4, and include units on your answer.

ii. Is the population growing at an increasing rate or growing at a decreasing rate on day 4? Explain.

(c) Find an equation for the tangent line to the curve $p(a) = 3a^4 - 2a^3 + 7a^2 - a + 12$ at the point where $a = -1$.

(d) What is the difference between being asked to find the *slope* of the tangent line (asked in (a)) and the *equation* of the tangent line (asked in (c))?

The sine and cosine functions

The sine and cosine functions are among the most important functions in all of mathematics. Sometimes called the *circular* functions due to their genesis in the unit circle, these periodic functions play a key role in modeling repeating phenomena such as the location of a point on a bicycle tire, the behavior of an oscillating mass attached to a spring, tidal elevations, and more. Like polynomial and exponential functions, the sine and cosine functions are considered basic functions, ones that are often used in the building of more complicated functions. As such, we would like to know formulas for $\frac{d}{dx}[\sin(x)]$ and $\frac{d}{dx}[\cos(x)]$, and the next two activities lead us to that end.

Activity 2.4–7

Consider the function $f(x) = \sin(x)$, which is graphed in Figure 2.27. Note carefully that the grid in the diagram does not have boxes that are 1×1, but rather approximately 1.57×1, as the horizontal scale of the grid is $\pi/2$ units per box.

(a) At each of $x = -2\pi, -\frac{3\pi}{2}, -\pi, -\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}, 2\pi$, use a straightedge to sketch an accurate tangent line to $y = f(x)$.

(b) Use the provided grid to estimate the slope of the tangent line you drew at each point. Pay careful attention to the scale of the grid.

(c) Use the limit definition of the derivative to estimate $f'(0)$ by using small values of h, and compare the result to your visual estimate for the slope of the tangent line to $y = f(x)$ at $x = 0$ in (b). Using periodicity, what does this result suggest about $f'(2\pi)$? about $f'(-2\pi)$?

(d) Based on your work in (a), (b), and (c), sketch an accurate graph of $y = f'(x)$ on the axes adjacent to the graph of $y = f(x)$.

(e) What familiar function do you think is the derivative of $f(x) = \sin(x)$?

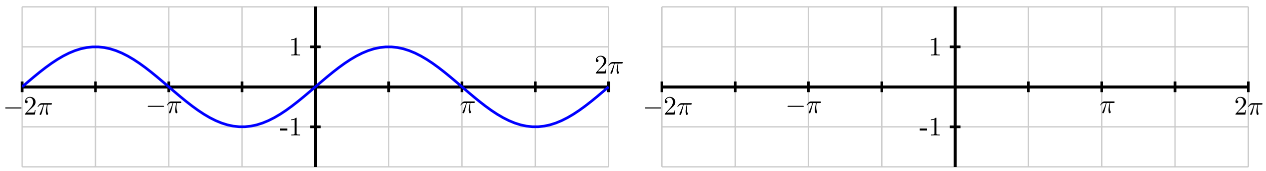


Figure 2.27: At left, the graph of $y = f(x) = \sin(x)$.

Activity 2.4–8

Consider the function $g(x) = \cos(x)$, which is graphed in Figure 2.28. Note carefully that the grid in the diagram does not have boxes that are 1×1, but rather approximately 1.57×1, as the horizontal scale of the grid is $\pi/2$ units per box.

(a) At each of $x = -2\pi, -\frac{3\pi}{2}, -\pi, -\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}, 2\pi$, use a straightedge to sketch an accurate tangent line to $y = g(x)$.

(b) Use the provided grid to estimate the slope of the tangent line you drew at each point. Again, note the scale of the axes and grid.

(c) Use the limit definition of the derivative to estimate $g'(\frac{\pi}{2})$ by using small values of h, and compare the result to your visual estimate for the slope of the tangent line to $y = g(x)$ at $x = \frac{\pi}{2}$ in (b). Using periodicity, what does this result suggest about $g'(-\frac{3\pi}{2})$? can symmetry on the graph help you estimate other slopes easily?

(d) Based on your work in (a), (b), and (c), sketch an accurate graph of $y = g'(x)$ on the axes adjacent to the graph of $y = g(x)$.

(e) What familiar function do you think is the derivative of $g(x) = \cos(x)$?

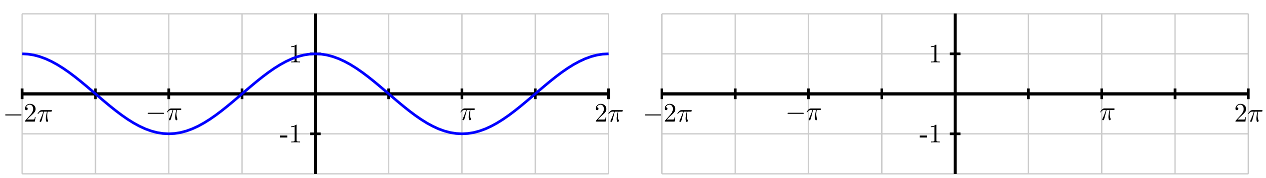


Figure 2.28: At left, the graph of $y = g(x) = \cos(x)$.

The results of the two preceding activities suggest that the sine and cosine functions not only have the beautiful interrelationships that are learned in a course in trigonometry – connections such as the identities $\sin^2(x) + \cos^2(x) = 1$ and $\cos(x - \frac{\pi}{2}) = \sin(x)$ – but that they are even further linked through calculus, as the derivative of each involves the other. The following rules summarize the results of the activities.

Sine and Cosine Functions:

For all real numbers x,

$$\frac{d}{dx}[\sin(x)] = \cos(x) \text{ and } \frac{d}{dx}[\cos(x)] = -\sin(x)$$

These two rules may be formally proved using the limit definition of the derivative and the expansion identities for $\sin(x+h)$ and $\cos(x+h)$, see Activity-2.4–9.

Activity 2.4–9

In this problem, we explore how the limit definition of the derivative more formally shows that $\frac{d}{dx}[\sin(x)] = \cos(x)$. Letting $f(x) = \sin(x)$, note that the limit definition of the derivative tells us that

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

(a) Recall the trigonometric identity for the sine of a sum of angles α and β:
$\sin(\alpha + \beta) = \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta)$. Use this identity and some algebra to show that

$$f'(x) = \lim_{h \to 0} \frac{\sin(x)(\cos(h) - 1) + \cos(x)\sin(h)}{h}.$$

(b) Next, note that as h changes, x remains constant. Explain why it therefore makes sense to say that

$$f'(x) = \sin(x) \cdot \lim_{h \to 0} \frac{\cos(h) - 1}{h} + \cos(x) \cdot \lim_{h \to 0} \frac{\sin(h)}{h}.$$

(c) Finally, use small values of h to estimate the values of the two limits in (c):

$$\lim_{h \to 0} \frac{\cos(h) - 1}{h} \text{ and } \lim_{h \to 0} \frac{\sin(h)}{h}.$$

(d) What do your results in (c) thus tell you about $f'(x)$?

(e) By emulating the steps taken above, use the limit definition of the derivative to argue convincingly that $\frac{d}{dx}[\cos(x)] = -\sin(x)$.

We have now added two additional functions to our library of basic functions whose derivatives we know: power functions, exponential functions, and the sine and cosine functions. The constant multiple and sum rules still hold, of course, and all of the inherent meaning of the derivative persists, regardless of the functions that are used to constitute a given choice of $f(x)$. The following activity puts our new knowledge of the derivatives of $\sin(x)$ and $\cos(x)$ to work.

Activity 2.4–10

Answer each of the following questions. Where a derivative is requested, be sure to label the derivative function with its name using proper notation.

(a) Determine the derivative of $h(t) = 3\cos(t) - 4\sin(t)$.

(b) Find the exact slope of the tangent line to $y = f(x) = 2x + \frac{\sin(x)}{2}$ at the point where $x = \frac{\pi}{6}$.

(c) Find the equation of the tangent line to $y = g(x) = x^2 + 2\cos(x)$ at the point where $x = \frac{\pi}{2}$.

(d) Determine the derivative of $p(z) = z^4 + 4^z + 4\cos(z) - \sin(\frac{\pi}{2})$.

(e) The function $P(t) = 24 + 8\sin(t)$ represents a population of a particular kind of animal that lives on a small island, where P is measured in hundreds and t is measured in decades since January 1, 2010. What is the instantaneous rate of change of P on January 1, 2030? What are the units of this quantity? Write a sentence in everyday language that explains how the population is behaving at this point in time.

Summary

In this section, we encountered the following important ideas:

- Given a differentiable function $y = f(x)$, we can express the derivative of f in several different notations: $f'(x)$, $\frac{df}{dx}$, $\frac{dy}{dx}$, and $\frac{d}{dx}[f(x)]$.
- The limit definition of the derivative leads to patterns among certain families of functions that enable us to compute derivative formulas without resorting directly to the limit definition. For example, if f is a power function of the form $f(x) = x^n$, then $f'(x) = nx^{n-1}$ for any real number n other than 0. This is called the Rule for Power Functions.
- We have stated a rule for derivatives of exponential functions in the same spirit as the rule for power functions: for any positive real number a, if $f(x) = a^x$, then $f'(x) = a^x \ln(a)$.
- If we are given a constant multiple of a function whose derivative we know, or a sum of functions whose derivatives we know, the Constant Multiple and Sum Rules make it straightforward to compute the derivative of the overall function. More formally, if $f(x)$ and $g(x)$ are differentiable with derivatives $f'(x)$ and $g'(x)$ and a and b are constants, then

$$\frac{d}{dx}\left[af(x) + bg(x)\right] = af'(x) + bg'(x).$$

- If we consider the graph of an exponential function $f(x) = a^x$ (where $a > 1$), the graph of $f'(x)$ behaves similarly, appearing exponential and as a possibly scaled version of the original function a^x. For $f(x) = 2^x$, careful analysis of the graph and its slopes suggests that $\frac{d}{dx}[2^x] = 2^x \ln(2)$.
- By carefully analyzing the graphs of $y = \sin(x)$ and $y = \cos(x)$, plus using the limit definition of the derivative at select points, we found that $\frac{d}{dx}[\sin(x)] = \cos(x)$ and $\frac{d}{dx}[\cos(x)] = -\sin(x)$.

Exercises

Terms and Concepts

1) What is the name of the rule which states that $\frac{d}{dx}(x^n) = nx^{n-1}$, where $n > 0$ is an integer?

2) Give an example of a function $f(x)$ where $f'(x) = f(x)$.

3) Give an example of a function $f(x)$ where $f'(x) = 0$.

4) The derivative rules introduced in this section explain how to compute the derivative of which of the following functions?

(a) $g(x) = 3x^2 - x + 17$

(b) $h(x) = 5\ln(x)$

(c) $j(x) = \sin(x)\cos(x)$

(d) $k(x) = e^{x^2}$

(e) $m(x) = \sqrt{x}$

(f) $f(x) = \dfrac{3}{x^2}$

Problems

In exercises 5–13, compute the derivative of the given function.

5) $f(x) = 7x^2 - 5x + 7$

6) $g(x) = 14x^3 + 7x^2 + 11x - 29$

7) $m(t) = 9t^5 - \frac{1}{8}t^3 + 3t - 8$

8) $f(\theta) = 9\sin(\theta) + 10\cos(\theta)$

9) $f(r) = 6e^r$

10) $g(t) = 10t^4 - \cos(t) + 7\sin(t)$

11) $p(s) = \frac{1}{4}s^4 + \frac{1}{3}s^3 + \frac{1}{2}s^2 + s + 1$

12) $h(t) = e^t - \sin(t) - \cos(t)$

13) $g(t) = (1+3t)^2$

In exercises 14–19, compute the first four derivatives of the given function.

14) $f(x) = x^6$

15) $g(x) = 2\cos(x)$

16) $h(t) = t^2 - e^t$

17) $p(\theta) = \theta^4 - \theta^3$

18) $f(\theta) = \sin(\theta) - \cos(\theta)$

19) $f(x) = 1,100$

In exercises 20–24, find the equation of the tangent line of the function at the given point.

20) $f(x) = x^3 - x$ at $x = 1$

21) $f(t) = e^t + 3$ at $t = 0$

22) $f(x) = 4\sin(x)$ at $x = \pi/2$

23) $f(x) = -2\cos(x)$ at $x = \pi/4$

24) $f(x) = 2x + 3$ at $x = 5$

25) Let f and g be differentiable functions for which the following information is known: $f(2) = 5$, $g(2) = -3$, $f'(2) = -1/2$, $g'(2) = 2$.

(a) Let h be the new function defined by the rule $h(x) = 3f(x) - 4g(x)$. Determine $h(2)$ and $h'(2)$.

(b) Find an equation for the tangent line to $y = h(x)$ at the point $(2, h(2))$.

(c) Let p be the function defined by the rule $p(x) = -2f(x) + \frac{1}{2}g(x)$. Is p increasing, decreasing, or neither at $a = 2$? Why?

(d) Estimate the value of $p(2.03)$ by using the local linearization of p at the point $(2, p(2))$.

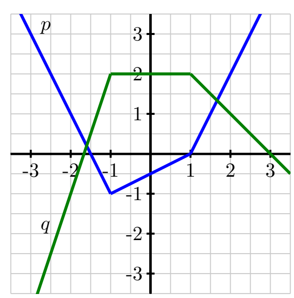
26) Consider the functions $r(t) = t^t$ and $s(t) = \arccos(t)$, for which you are given the facts that $r'(t) = t^t(\ln(t) + 1)$ and $s'(t) = -\frac{1}{\sqrt{1-t^2}}$. Do not be concerned with where these derivative formulas come from. We restrict our interest in both functions to the domain $0 < t < 1$.

(a) Let $w(t) = 3t^t - 2\arccos(t)$. Determine $w'(t)$.

(b) Find an equation for the tangent line to $y = w(t)$ at the point $(\frac{1}{2}, w(\frac{1}{2}))$.

(c) Let $v(t) = t^t + \arccos(t)$. Is v increasing or decreasing at the instant $t = \frac{1}{2}$? Why?

27) Let functions p and q be the piecewise linear functions given by their respective graphs given below. Use the graphs to answer the following questions.



(a) At what values of x is p not differentiable? At what values of x is q not differentiable? Why?

(b) Let $r(x) = p(x) + 2q(x)$. At what values of x is r not differentiable? Why?

(c) Determine $r'(-2)$ and $r'(0)$.

(d) Find an equation for the tangent line to $y = r(x)$ at the point $(2, r(2))$.

28) Use the definition of the derivative to prove the Constant Rule for differentiation.

29) Use the definition of the derivative to prove the Constant Multiple Rule for differentiation.

30) Use the definition of the derivative to prove the Sum Rule for differentiation.

2.5 *The product and quotient rules*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- How does the algebraic structure of a function direct us in computing its derivative using shortcut rules?
- How do we compute the derivative of a product of two basic functions in terms of the derivatives of the basic functions?
- How do we compute the derivative of a quotient of two basic functions in terms of the derivatives of the basic functions?
- How do the product and quotient rules combine with the sum and constant multiple rules to expand the library of functions we can quickly differentiate?
- What are the derivatives of the tangent, cotangent, secant, and cosecant functions?
- How do the derivatives of $\tan(x)$, $\cot(x)$, $\sec(x)$, and $\csc(x)$ combine with other derivative rules we have developed to expand the library of functions we can quickly differentiate?

Introduction

So far, the basic functions we know how to differentiate include power functions (x^n), exponential functions (a^x), and the two fundamental trigonometric functions ($\sin(x)$ and $\cos(x)$). With the sum rule and constant multiple rules, we can also compute the derivative of combined functions such as

$$f(x) = 7x^{11} - 4 \cdot 9^x + \pi \sin(x) - \sqrt{3}\cos(x),$$

because the function f is fundamentally a sum of basic functions. Indeed, we can now quickly say that $f'(x) = 77x^{10} - 4 \cdot 9^x \ln(9) + \pi \cos(x) + \sqrt{3}\sin(x)$.

But we can of course combine basic functions in ways other than multiplying them by constants and taking sums and differences. For example, we could consider the function that results from a product of two basic functions, such as

$$p(z) = z^3 \cos(z),$$

or another that is generated by the quotient of two basic functions, one like

$$q(t) = \frac{\sin(t)}{2^t}.$$

While the derivative of a sum is the sum of the derivatives, it turns out that the rules for computing derivatives of products and quotients are more complicated. In what follows we explore why this is the case, what the product and quotient rules actually say, and work to expand our repertoire of functions we can easily differentiate. To start, Preview Activity 2.5 asks you to investigate the derivative of a product and quotient of two polynomials.

Preview Activity 2.5

Let f and g be the functions defined by $f(t) = 2t^2$ and $g(t) = t^3 + 4t$.

(a) Determine $f'(t)$ and $g'(t)$.

(b) Let $p(t) = 2t^2(t^3 + 4t)$ and observe that $p(t) = f(t) \cdot g(t)$. Rewrite the formula for p by distributing the $2t^2$ term. Then, compute $p'(t)$ using the sum and constant multiple rules.

(c) True or false: $p'(t) = f'(t) \cdot g'(t)$.

(d) Let $q(t) = \dfrac{t^3 + 4t}{2t^2}$ and observe that $q(t) = \dfrac{g(t)}{f(t)}$. Rewrite the formula for q by dividing each term in the numerator by the denominator and simplify to write q as a sum of constant multiples of powers of t. Then, compute $q'(t)$ using the sum and constant multiple rules.

(e) True or false: $q'(t) = \dfrac{g'(t)}{f'(t)}$.

The product rule

As parts (b) and (d) of Preview Activity 2.5 show, it is not true in general that the derivative of a product of two functions is the product of the derivatives of those functions. Indeed, the rule for differentiating a function of the form $p(x) = f(x) \cdot g(x)$ in terms of the derivatives of f and g is more complicated than simply taking the product of the derivatives of f and g. To see further why this is the case, as well as to begin to understand how the product rule actually works, we consider an example involving meaningful functions.

Say that an investor is regularly purchasing stock in a particular company. Let $N(t)$ be a function that represents the number of shares owned on day t, where $t = 0$ represents the first day on which shares were purchased. Further, let $S(t)$ be a function that gives the value of one share of the stock on day t; note that the units on $S(t)$ are dollars per share. Moreover, to compute the total value on day t of the stock held by the investor, we use the function $V(t) = N(t) \cdot S(t)$. By taking the product

$$V(t) = N(t)\,\text{shares} \cdot S(t)\,\text{dollars per share},$$

we have the total value in dollars of the shares held. Observe that over time, both the number of shares and the value of a given share will vary. The derivative $N'(t)$ measures the rate at which the number of shares held is changing, while $S'(t)$ measures the rate at which the value per share is changing. The big question we'd like to answer is: how do these respective rates of change affect the rate of change of the total value function?

To help better understand the relationship among changes in N, S, and V, let's consider some specific data. Suppose that on day 100, the investor owns 520 shares of stock and the stock's current value is \$27.50 per share. This tells us that $N(100) = 520$ and $S(100) = 27.50$. In addition, say that on day 100, the investor purchases an additional 12 shares (so the number of shares held is rising at a rate of 12 shares per day), and that on that same day the price of the stock is rising at a rate of 0.75 dollars per share per day. Viewed in calculus notation, this tells us that $N'(100) = 12$ (shares per day) and $S'(100) = 0.75$ (dollars per share per day). At what rate is the value of the investor's total holdings changing on day 100?

Observe that the increase in total value comes from two sources: the growing number of shares, and the rising value of each share. If only the number of shares is rising (and the value of each share is constant), the rate at which which total value would rise is found by computing the product of the current value of the shares with the rate at which the number of shares is changing. That is, the rate at which total value would change is given by

$$S(100) \cdot N'(100) = 27.50 \, \frac{\text{dollars}}{\text{share}} \cdot 12 \, \frac{\text{shares}}{\text{day}} = 330 \, \frac{\text{dollars}}{\text{day}}.$$

Note particularly how the units make sense and explain that we are finding the rate at which the total value V is changing, measured in dollars per day. If instead the number of shares is constant, but the value of each share is rising, then the rate at which the total value would rise is found similarly by taking the product of the number of shares with the rate of change of share value. In particular, the rate total value is rising is

$$N(100) \cdot S'(100) = 520 \, \text{shares} \cdot 0.75 \, \frac{\text{dollars per share}}{\text{day}} = 390 \, \frac{\text{dollars}}{\text{day}}.$$

Of course, when both the number of shares is changing and the value of each share is changing, we have to include both of these sources, and hence the rate at which the total value is rising is

$$\begin{aligned} V'(100) &= S(100) \cdot N'(100) + N(100) \cdot S'(100) \\ &= 330 + 390 = 720 \, \frac{\text{dollars}}{\text{day}}. \end{aligned}$$

This tells us that we expect the total value of the investor's holdings to rise by about \$720 on the 100th day.[4]

Next, we expand our perspective from the specific example above to the more general and abstract setting of a product p of two differentiable functions, f and g. If we have $P(x) = f(x) \cdot g(x)$, our work above suggests that $P'(x) = f(x)g'(x) + g(x)f'(x)$. Indeed, a formal proof using the limit definition of the derivative can be given to show that the following rule, called the *product rule*, holds in general.

[4] While this example highlights why the product rule is true, there are some subtle issues to recognize. For one, if the stock's value really does rise exactly \$0.75 on day 100, and the number of shares really rises by 12 on day 100, then we'd expect that $V(101) = N(101) \cdot S(101) = 532 \cdot 28.25 = 15029$. If, as noted above, we expect the total value to rise by \$720, then with $V(100) = N(100) \cdot S(100) = 520 \cdot 27.50 = 14300$, then it seems like we should find that $V(101) = V(100) + 720 = 15020$. Why do the two results differ by 9? One way to understand why this difference occurs is to recognize that $N'(100) = 12$ represents an *instantaneous* rate of change, while our (informal) discussion has also thought of this number as the total change in the number of shares over the course of a single day. The formal proof of the product rule reconciles this issue by taking the limit as the change in the input tends to zero.

Product Rule:

If f and g are differentiable functions, then their product $P(x) = f(x) \cdot g(x)$ is also a differentiable function, and

$$P'(x) = f(x)g'(x) + g(x)f'(x).$$

Example 1

Use the definition of the derivative to prove the Product Rule for differentiation.

Solution. By the limit definition, we have

$$\frac{d}{dx}\Big(f(x)g(x)\Big) = \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h}.$$

We now do something a bit unexpected; add 0 to the numerator (so that nothing is changed) in the form of $-(x+h)g(x) + f(x+h)g(x)$, then do some regrouping as shown.

$$\begin{aligned}
&\frac{d}{dx}\Big(f(x)g(x)\Big) \\
&= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \quad \text{(now add 0 to the numerator)} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} \frac{\Big(f(x+h)g(x+h) - f(x+h)g(x)\Big) + \Big(f(x+h)g(x) - f(x)g(x)\Big)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x+h)g(x)}{h} + \lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} f(x+h)\frac{g(x+h) - g(x)}{h} + \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} g(x) \\
&= f(x)g'(x) + f'(x)g(x)
\end{aligned}$$

In light of the earlier example involving shares of stock, the product rule also makes sense intuitively: the rate of change of P should take into account both how fast f and g are changing, as well as how large f and g are at the point of interest. Furthermore, we note in words what the product rule says: if P is the product of two functions f (the first function) and g (the second), then "the derivative of P is the first times the derivative of the second, plus the second times the derivative of the first." It is often a helpful mental exercise to say this phrasing aloud when executing the product rule.

Example 2

Use the Product Rule to compute the derivative of $P(z) = z^3 \cdot \cos(z)$.

Solution. The first function is z^3 and the second function is $\cos(z)$. By the product rule, P' will be given by the first, z^3, times the derivative of the second, $-\sin(z)$, plus the second, $\cos(z)$, times the derivative of the first, $3z^2$. That is,

$$P'(z) = z^3(-\sin(z)) + \cos(z)3z^2 = -z^3\sin(z) + 3z^2\cos(z).$$

The following activity further explores the use of the product rule.

Activity 2.5–1

Use the product rule to answer each of the questions below. Throughout, be sure to carefully label any derivative you find by name. That is, if you're given a formula for $f(x)$, clearly label the formula you find for $f'(x)$. It is not necessary to algebraically simplify any of the derivatives you compute.

(a) Let $m(w) = 3w^{17}4^w$. Find $m'(w)$.

(b) Let $h(t) = (\sin(t) + \cos(t))t^4$. Find $h'(t)$.

(c) Determine the slope of the tangent line to the curve $y = f(x)$ at the point where $a = 1$ if f is given by the rule $f(x) = e^x \sin(x)$.

(d) Find the tangent line approximation $L(x)$ to the function $y = g(x)$ at the point where $a = -1$ if g is given by the rule $g(x) = (x^2 + x)2^x$.

The quotient rule

Because quotients and products are closely linked, we can use the product rule to understand how to take the derivative of a quotient. In particular, let $Q(x)$ be defined by $Q(x) = f(x)/g(x)$, where f and g are both differentiable functions. We desire a formula for Q' in terms of f, g, f', and g'. It turns out that Q is differentiable everywhere that $g(x) \neq 0$. Moreover, taking the formula $Q = f/g$ and multiplying both sides by g, we can observe that

$$f(x) = Q(x) \cdot g(x).$$

Thus, we can use the product rule to differentiate f. Doing so,

$$f'(x) = Q(x)g'(x) + g(x)Q'(x).$$

Since we want to know a formula for Q', we work to solve this most recent equation for $Q'(x)$, finding first that

$$Q'(x)g(x) = f'(x) - Q(x)g'(x).$$

Dividing both sides by $g(x)$, we have

$$Q'(x) = \frac{f'(x) - Q(x)g'(x)}{g(x)}.$$

Finally, we also recall that $Q(x) = \frac{f(x)}{g(x)}$. Using this expression in the preceding equation and simplifying, we have

$$\begin{aligned} Q'(x) &= \frac{f'(x) - \frac{f(x)}{g(x)}g'(x)}{g(x)} \\ &= \frac{f'(x) - \frac{f(x)}{g(x)}g'(x)}{g(x)} \cdot \frac{g(x)}{g(x)} \\ &= \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}. \end{aligned}$$

This shows the fundamental argument for why the *quotient rule* holds.

Quotient Rule:

If f and g are differentiable functions, then their quotient $Q(x) = \frac{f(x)}{g(x)}$ is also a differentiable function for all x where $g(x) \neq 0$, and

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

Like the product rule, it can be helpful to think of the quotient rule verbally. If a function Q is the quotient of a top function f and a bottom function g, then Q' is given by "the bottom times the derivative of the top, minus the top times the derivative of the bottom, all over the bottom squared."

Example 3

Use the Quotient Rule to compute the derivative of $Q(t) = \sin(t)/2^t$.

Solution. We can identify the top function as $\sin(t)$ and the bottom function as 2^t. By the quotient rule, we then have that Q' will be given by the bottom, 2^t, times the derivative of the top, $\cos(t)$, minus the top, $\sin(t)$, times the derivative of the bottom, $2^t \ln(2)$, all over the bottom squared, $(2^t)^2$. That is,

$$Q'(t) = \frac{2^t \cos(t) - \sin(t) 2^t \ln(2)}{(2^t)^2}.$$

In this particular example, it is possible to simplify $Q'(t)$ by removing a factor of 2^t from both the numerator and denominator, hence finding that

$$Q'(t) = \frac{\cos(t) - \sin(t)\ln(2)}{2^t}.$$

In general, we must be careful in doing any such simplification, as we

> don't want to correctly execute the quotient rule but then find an incorrect overall derivative due to an algebra error. As such, we will often place more emphasis on correctly using derivative rules than we will on simplifying the result that follows.

The following activity further explores the use of the quotient rule.

Activity 2.5–2

Use the quotient rule to answer each of the questions below. Throughout, be sure to carefully label any derivative you find by name. That is, if you're given a formula for $f(x)$, clearly label the formula you find for $f'(x)$. It is not necessary to algebraically simplify any of the derivatives you compute.

(a) Let $r(z) = \dfrac{3^z}{z^4+1}$. Find $r'(z)$.

(b) Let $v(t) = \dfrac{\sin(t)}{\cos(t)+t^2}$. Find $v'(t)$.

(c) Determine the slope of the tangent line to the curve $R(x) = \dfrac{x^2-2x-8}{x^2-9}$ at the point where $x = 0$.

(d) When a camera flashes, the intensity I of light seen by the eye is given by the function
$$I(t) = \frac{100t}{e^t},$$
where I is measured in candles and t is measured in milliseconds. Compute $I'(0.5)$, $I'(2)$, and $I'(5)$; include appropriate units on each value; and discuss the meaning of each.

Combining rules

One of the challenges to learning to apply various derivative shortcut rules correctly and effectively is recognizing the fundamental structure of a function. For instance, consider the function given by

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

How do we decide which rules to apply? Our first task is to recognize the overall structure of the given function. Observe that the function f is fundamentally a sum of two slightly less complicated functions, so we can apply the sum rule[5] and get

$$\begin{aligned} f'(x) &= \frac{d}{dx}\left[x\sin(x) + \frac{x^2}{\cos(x)+2}\right] \\ &= \frac{d}{dx}\left[x\sin(x)\right] + \frac{d}{dx}\left[\frac{x^2}{\cos(x)+2}\right] \end{aligned}$$

[5] When taking a derivative that involves the use of multiple derivative rules, it is often helpful to use the notation $\frac{d}{dx}[\ ]$ to wait to apply subsequent rules. This is demonstrated in each of the two examples presented here.

Now, the left-hand term above is a product, so the product rule is needed there, while the right-hand term is a quotient, so the quotient rule is required. Applying these rules respectively, we find that

$$\begin{aligned} f'(x) &= (x\cos(x)+\sin(x)) + \frac{(\cos(x)+2)2x - x^2(-\sin(x))}{(\cos(x)+2)^2} \\ &= x\cos(x)+\sin(x) + \frac{2x\cos(x)+4x^2+x^2\sin(x)}{(\cos(x)+2)^2}. \end{aligned}$$

We next consider how the situation changes with the function defined by

$$s(y) = \frac{y \cdot 7^y}{y^2+1}.$$

Overall, s is a quotient of two simpler function, so the quotient rule will be needed. Here, we execute the quotient rule and use the notation $\frac{d}{dy}$ to defer the computation of the derivative of the numerator and derivative of the denominator. Thus,

$$s'(y) = \frac{(y^2+1)\cdot \frac{d}{dy}\left[y\cdot 7^y\right] - y\cdot 7^y \cdot \frac{d}{dy}\left[y^2+1\right]}{(y^2+1)^2}.$$

Now, there remain two derivatives to calculate. The first one, $\frac{d}{dy}\left[y\cdot 7^y\right]$ calls for use of the product rule, while the second, $\frac{d}{dy}\left[y^2+1\right]$ takes only an elementary application of the sum rule. Applying these rules, we now have

$$s'(y) = \frac{(y^2+1)[y\cdot 7^y\ln(7) + 7^y\cdot 1] - y\cdot 7^y[2y]}{(y^2+1)^2}.$$

While some minor simplification is possible, we are content to leave $s'(y)$ in its current form, having found the desired derivative of s. In summary, to compute the derivative of s, we applied the quotient rule. In so doing, when it was time to compute the derivative of the top function, we used the product rule; at the point where we found the derivative of the bottom function, we used the sum rule.

In general, one of the main keys to success in applying derivative rules is to recognize the structure of the function, followed by the careful and diligent application of relevant derivative rules. The best way to get good at this process is by doing a large number of exercises, and the next activity provides some practice and exploration to that end.

Activity 2.5–3

Use relevant derivative rules to answer each of the questions below. Throughout, be sure to use proper notation and carefully label any derivative you find by name.

(a) Let $f(r) = (5r^3 + \sin(r))(4^r - 2\cos(r))$. Find $f'(r)$.

(b) Let $p(t) = \dfrac{\cos(t)}{t^6 \cdot 6^t}$. Find $p'(t)$.

(c) Let $g(z) = 3z^7 e^z - 2z^2 \sin(z) + \dfrac{z}{z^2+1}$. Find $g'(z)$.

(d) A moving particle has its position in feet at time t in seconds given by the function $s(t) = \dfrac{3\cos(t) - \sin(t)}{e^t}$. Find the particle's instantaneous velocity at the moment $t = 1$.

(e) Suppose that $f(x)$ and $g(x)$ are differentiable functions and it is known that $f(3) = -2$, $f'(3) = 7$, $g(3) = 4$, and $g'(3) = -1$. If $p(x) = f(x) \cdot g(x)$ and $q(x) = \dfrac{f(x)}{g(x)}$, calculate $p'(3)$ and $q'(3)$.

As the algebraic complexity of the functions we are able to differentiate continues to increase, it is important to remember that all of the derivative's meaning continues to hold. Regardless of the structure of the function f, the value of $f'(a)$ tells us the instantaneous rate of change of f with respect to x at the moment $x = a$, as well as the slope of the tangent line to $y = f(x)$ at the point $(a, f(a))$.

Derivatives of the tangent, cotangent, secant, and cosecant functions

Activity 2.5–4

Consider the function $f(x) = \tan(x)$, and remember that $\tan(x) = \dfrac{\sin(x)}{\cos(x)}$.

(a) What is the domain of f?

(b) Use the quotient rule to show that one expression for $f'(x)$ is

$$f'(x) = \frac{\cos(x)\cos(x) + \sin(x)\sin(x)}{\cos^2(x)}.$$

(c) What is the Fundamental Trigonometric Identity? How can this identity be used to find a simpler form for $f'(x)$?

(d) Recall that $\sec(x) = \dfrac{1}{\cos(x)}$. How can we express $f'(x)$ in terms of the secant function?

(e) For what values of x is $f'(x)$ defined? How does this set compare to the domain of f?

In Activity 2.5–4, we found that the derivative of the tangent function can be expressed in several ways, but most simply in terms of the secant function. Next, we develop the derivative of the cotangent function.

Let $g(x) = \cot(x)$. To find $g'(x)$, we observe that $g(x) = \frac{\cos(x)}{\sin(x)}$ and apply the quotient rule. Hence

$$\begin{aligned} g'(x) &= \frac{\sin(x)(-\sin(x)) - \cos(x)\cos(x)}{\sin^2(x)} \\ &= -\frac{\sin^2(x) + \cos^2(x)}{\sin^2(x)} \end{aligned}$$

By the Fundamental Trigonometric Identity, we see that $g'(x) = -\frac{1}{\sin^2(x)}$; recalling that $\csc(x) = \frac{1}{\sin(x)}$, it follows that we can most simply express g' by the rule

$$g'(x) = -\csc^2(x).$$

Note that neither g nor g' is defined when $\sin(x) = 0$, which occurs at every integer multiple of π. Hence we have the following rule.

Cotangent Function:

For all real numbers x such that $x \neq k\pi$, where $k = 0, \pm 1, \pm 2, \ldots$,

$$\frac{d}{dx}[\cot(x)] = -\csc^2(x).$$

Observe that the shortcut rule for the cotangent function is very similar to the rule we discovered in Activity 2.5–4 for the tangent function.

Tangent Function:

For all real numbers x such that $x \neq \frac{k\pi}{2}$, where $k = \pm 1, \pm 2, \ldots$,

$$\frac{d}{dx}[\tan(x)] = \sec^2(x).$$

In the next two activities, we develop the rules for differentiating the secant and cosecant functions.

Activity 2.5–5

Let $h(x) = \sec(x)$ and recall that $\sec(x) = \frac{1}{\cos(x)}$.

(a) What is the domain of h?

(b) Use the quotient rule to develop a formula for $h'(x)$ that is expressed completely in terms of $\sin(x)$ and $\cos(x)$.

(c) How can you use other relationships among trigonometric functions to write $h'(x)$ only in terms of $\tan(x)$ and $\sec(x)$?

(d) What is the domain of h'? How does this compare to the domain of h?

Activity 2.5–6

Let $p(x) = \csc(x)$ and recall that $\csc(x) = \dfrac{1}{\sin(x)}$.

(a) What is the domain of p?

(b) Use the quotient rule to develop a formula for $p'(x)$ that is expressed completely in terms of $\sin(x)$ and $\cos(x)$.

(c) How can you use other relationships among trigonometric functions to write $p'(x)$ only in terms of $\cot(x)$ and $\csc(x)$?

(d) What is the domain of p'? How does this compare to the domain of p?

The quotient rule has thus enabled us to determine the derivatives of the tangent, cotangent, secant, and cosecant functions, expanding our overall library of basic functions we can differentiate. Moreover, we observe that just as the derivative of any polynomial function is a polynomial, and the derivative of any exponential function is another exponential function, so it is that the derivative of any basic trigonometric function is another function that consists of basic trigonometric functions. This makes sense because all trigonometric functions are periodic, and hence their derivatives will be periodic, too.

As has been and will continue to be the case throughout our work in Chapter 2, the derivative retains all of its fundamental meaning as an instantaneous rate of change and as the slope of the tangent line to the function under consideration. Our present work primarily expands the list of functions for which we can quickly determine a formula for the derivative. Moreover, with the addition of $\tan(x)$, $\cot(x)$, $\sec(x)$, and $\csc(x)$ to our library of basic functions, there are many more functions we can differentiate through the sum, constant multiple, product, and quotient rules.

Example 4

Find the derivative of $f(x) = \dfrac{5x^2}{\sin(x)}$ was found using the Quotient Rule. Rewrite f as $f(x) = 5x^2\csc(x)$, find f' using the Product rule and verify the two answers are the same.

Solution. Directly applying the Quotient Rule gives:

$$\begin{aligned}\frac{d}{dx}\left(\frac{5x^2}{\sin(x)}\right) &= \frac{\sin(x)\cdot 10x - 5x^2\cdot\cos(x)}{\sin^2(x)}\\ &= \frac{10x\sin(x) - 5x^2\cos(x)}{\sin^2(x)}.\end{aligned}$$

We now find f' using the Product Rule, considering f as $f(x) = 5x^2\csc(x)$.

$$\begin{aligned}f'(x) &= \frac{d}{dx}\Big(5x^2\csc(x)\Big)\\ &= 5x^2(-\csc(x)\cot(x)) + 10x\csc(x) \quad \text{(now rewrite trig functions)}\\ &= 5x^2\cdot\frac{-1}{\sin(x)}\cdot\frac{\cos(x)}{\sin(x)} + \frac{10x}{\sin(x)}\\ &= \frac{-5x^2\cos(x)}{\sin^2(x)} + \frac{10x}{\sin(x)} \quad \text{(get common denominator)}\\ &= \frac{10x\sin(x) - 5x^2\cos(x)}{\sin^2(x)}\end{aligned}$$

Finding f' using either method returned the same result. At first, the answers looked different, but some algebra verified they are the same. In general, there is not one final form that we seek; the immediate result from the Product Rule is fine. Work to "simplify" your results into a form that is most readable and useful to you.

Activity 2.5–7

Answer each of the following questions. Where a derivative is requested, be sure to label the derivative function with its name using proper notation.

(a) Let $f(x) = 5\sec(x) - 2\csc(x)$. Find the slope of the tangent line to f at the point where $x = \frac{\pi}{3}$.

(b) Let $p(z) = z^2\sec(z) - z\cot(z)$. Find the instantaneous rate of change of p at the point where $z = \frac{\pi}{4}$.

(c) Let $h(t) = \dfrac{\tan(t)}{t^2+1} - 2e^t\cos(t)$. Find $h'(t)$.

(d) Let $g(r) = \dfrac{r\sec(r)}{5^r}$. Find $g'(r)$.

(e) When a mass hangs from a spring and is set in motion, the object's position oscillates in a way that the size of the oscillations decrease. This is usually called a *damped oscillation*. Suppose that for a particular object, its displacement from equilibrium (where the object sits at rest) is modeled by the function

$$s(t) = \frac{15\sin(t)}{e^t}.$$

Assume that s is measured in inches and t in seconds. Sketch a graph of this function for $t \geq 0$ to see how it represents the situation described. Then compute ds/dt, state the units on this function, and explain what it tells you about the object's motion. Finally, compute and interpret $s'(2)$.

Summary

In this section, we encountered the following important ideas:

- If a function is a sum, product, or quotient of simpler functions, then we can use the sum, product, or quotient rules to differentiate the overall function in terms of the simpler functions and their derivatives.
- The product rule tells us that if P is a product of differentiable functions f and g according to the rule $P(x) = f(x)g(x)$, then
$$P'(x) = f(x)g'(x) + g(x)f'(x).$$
- The quotient rule tells us that if Q is a quotient of differentiable functions f and g according to the rule $Q(x) = \frac{f(x)}{g(x)}$, then
$$Q'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}.$$
- The product and quotient rules now complement the constant multiple and sum rules and enable us to compute the derivative of any function that consists of sums, constant multiples, products, and quotients of basic functions we already know how to differentiate. For instance, if F has the form
$$F(x) = \frac{2a(x) - 5b(x)}{c(x) \cdot d(x)},$$
then F is fundamentally a quotient, and the numerator is a sum of constant multiples and the denominator is a product. Hence the derivative of F can be found by applying the quotient rule and then using the sum and constant multiple rules to differentiate the numerator and the product rule to differentiate the denominator.
- The derivatives of the other four trigonometric functions are
$$\frac{d}{dx}[\tan(x)] = \sec^2(x), \quad \frac{d}{dx}[\cot(x)] = -\csc^2(x),$$
$$\frac{d}{dx}[\sec(x)] = \sec(x)\tan(x), \text{ and } \frac{d}{dx}[\csc(x)] = -\csc(x)\cot(x).$$
Each derivative exists and is defined on the same domain as the original function. For example, both the tangent function and its derivative are defined for all real numbers x such that $x \neq \frac{k\pi}{2}$, where $k = \pm 1, \pm 2, \ldots$.
- The above four rules for the derivatives of the tangent, cotangent, secant, and cosecant can be used along with the rules for power functions, exponential functions, and the sine and cosine, as well as the sum, constant multiple, product, and quotient rules, to quickly differentiate a wide range of different functions.

Exercises

Terms and Concepts

1) T/F: The Product Rule states that $\frac{d}{dx}(x^2\sin(x)) = 2x\cos(x)$.

2) T/F: The Quotient Rule states that $\frac{d}{dx}\left(\frac{x^2}{\sin(x)}\right) = \frac{\cos(x)}{2x}$.

3) T/F: Regardless of the function, there is always exactly one right way of computing its derivative.

Problems

In exercises 4–7,

(a) Use the Product Rule to differentiate the function.

(b) Manipulate the function algebraically and differentiate without the Product Rule.

(c) Show that the answers from (a) and (b) are equivalent.

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

5) $g(x) = 2x^2(5x^3)$

6) $m(t) = (2t-1)(t+4)$

7) $f(\theta) = (\theta^2+5)(3-\theta^3)$

In exercises 8–12,

(a) Use the Quotient Rule to differentiate the function.

(b) Manipulate the function algebraically and differentiate without the Quotient Rule.

(c) Show that the answers from (a) and (b) are equivalent.

8) $f(x) = \dfrac{x^2+3}{x}$

9) $g(x) = \dfrac{x^3-2x^2}{2x^2}$

10) $h(s) = \dfrac{3}{4s^3}$

11) $f(t) = \dfrac{t^2-1}{t+1}$

12) $f(x) = \dfrac{x^4+2x^3}{x+2}$

In exercises 13–26, differentiate the given function.

13) $f(x) = x\sin(x)$

14) $f(t) = \dfrac{1}{t^2}(\csc(t)-4)$

15) $f(x) = \dfrac{x+7}{\sqrt{x}}$

16) $g(x) = \dfrac{x^5}{\cos(x)-2x^2}$

17) $h(t) = \cot(t) - e^t$

18) $h(x) = 7x^2+6x-2$

19) $f(x) = (16x^3+24x^2+3x)\dfrac{7x-1}{16x^3+24x^2+3x}$

20) $f(t) = \sqrt[5]{t}(\sec(t)+e^t)$

21) $f(x) = \dfrac{\sin(x)}{\cos(x)+3}$

22) $g(x) = e^2(\sin(\pi/4)-1)$

23) $g(t) = 4t^3e^t - \sin(t)\cos(t)$

24) $h(t) = \dfrac{2^t+3}{3^t+2}$

25) $f(x) = x^2e^x\tan(x)$

26) $g(x) = 2x\sin(x)\sec(x)$

In exercises 27–30, find the equation of the tangent line of the function at the given point.

27) $g(s) = e^s(s^2+2)$ at $(0,2)$

28) $g(t) = t\sin(t)$ at $(3\pi/2, -3\pi/2)$

29) $f(x) = \dfrac{x^2}{x-1}$ at $(2,4)$

30) $f(t) = \dfrac{\cos(t)-8t}{t+1}$ at $(0,-5)$

In exercises 31–34, find the x-values where the function has a horizontal tangent line.

31) $f(x) = 6x^2-18x-24$

32) $f(x) = x\sin(x)$ on $[-1,1]$

33) $g(x) = \dfrac{x}{x+1}$

34) $h(x) = \dfrac{x^2}{x+1}$

In exercises 35–38, find the indicated derivative.

35) $f(x) = x\sin(x)$; find $f''(x)$

36) $f(x) = x\sin(x)$; find $f^{(4)}(x)$

37) $f(x) = \csc(x)$; find $f''(x)$

38) $f(x) = (x^3-5x+2)(x^2+x-7)$; find $f^{(8)}(x)$

39) Let f and g be differentiable functions for which the following information is known: $f(2)=5$, $g(2)=-3$, $f'(2)=-1/2$, $g'(2)=2$.

(a) Let h be the new function defined by the rule $h(x) = g(x)\cdot f(x)$. Determine $h(2)$ and $h'(2)$.

(b) Find an equation for the tangent line to $y = h(x)$ at the point $(2, h(2))$.

(c) Let r be the function defined by the rule $r(x) = \frac{g(x)}{f(x)}$. Is r increasing, decreasing, or neither at $a=2$? Why?

(d) Estimate the value of $r(2.06)$ by using the local linearization of p at the point $(2, p(2))$.

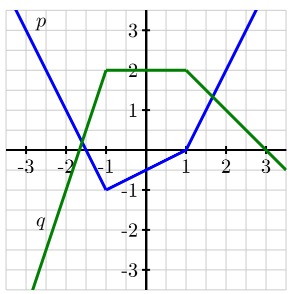
40) Consider the functions $r(t) = t^t$ and $s(t) = \arccos(t)$, for which you are given the facts that $r'(t) = t^t(\ln(t) + 1)$ and $s'(t) = -\frac{1}{\sqrt{1-t^2}}$. Do not be concerned with where these derivative formulas come from. We restrict our interest in both functions to the domain $0 < t < 1$.

   (a) Let $w(t) = t^t \arccos(t)$. Determine $w'(t)$.

   (b) Find an equation for the tangent line to $y = w(t)$ at the point $(\frac{1}{2}, w(\frac{1}{2}))$.

   (c) Let $v(t) = \frac{t^t}{\arccos(t)}$. Is v increasing or decreasing at the instant $t = \frac{1}{2}$? Why?

41) Let functions p and q be the piecewise linear functions given by their respective graphs below. Use the graphs to answer the following questions.



   (a) Let $r(x) = p(x) \cdot q(x)$. Determine $r'(-2)$ and $r'(0)$.

   (b) Are there values of x for which $r'(x)$ does not exist? If so, which values, and why?

   (c) Find an equation for the tangent line to $y = r(x)$ at the point $(2, r(2))$.

   (d) Let $z(x) = \frac{q(x)}{p(x)}$. Determine $z'(0)$ and $z'(2)$.

   (e) Are there values of x for which $z'(x)$ does not exist? If so, which values, and why?

42) A farmer with large land holdings has historically grown a wide variety of crops. With the price of ethanol fuel rising, he decides that it would be prudent to devote more and more of his acreage to producing corn. As he grows more and more corn, he learns efficiencies that increase his yield per acre. In the present year, he used 7000 acres of his land to grow corn, and that land had an average yield of 170 bushels per acre. At the current time, he plans to increase his number of acres devoted to growing corn at a rate of 600 acres/year, and he expects that right now his average yield is increasing at a rate of 8 bushels per acre per year. Use this information to answer the following questions.

   (a) Say that the present year is $t = 0$, that $A(t)$ denotes the number of acres the farmer devotes to growing corn in year t, $Y(t)$ represents the average yield in year t (measured in bushels per acre), and $C(t)$ is the total number of bushels of corn the farmer produces. What is the formula for $C(t)$ in terms of $A(t)$ and $Y(t)$? Why?

   (b) What is the value of $C(0)$? What does it measure?

   (c) Write an expression for $C'(t)$ in terms of $A(t)$, $A'(t)$, $Y(t)$, and $Y'(t)$. Explain your thinking.

   (d) What is the value of $C'(0)$? What does it measure?

   (e) Based on the given information and your work above, estimate the value of $C(1)$.

43) Let $f(v)$ be the gas consumption (in liters/km) of a car going at velocity v (in km/hour). In other words, $f(v)$ tells you how many liters of gas the car uses to go one kilometer if it is traveling at v kilometers per hour. In addition, suppose that $f(80) = 0.05$ and $f'(80) = 0.0004$.

   (a) Let $g(v)$ be the distance the same car goes on one liter of gas at velocity v. What is the relationship between $f(v)$ and $g(v)$? Hence find $g(80)$ and $g'(80)$.

   (b) Let $h(v)$ be the gas consumption in liters per hour of a car going at velocity v. In other words, $h(v)$ tells you how many liters of gas the car uses in one hour if it is going at velocity v. What is the algebraic relationship between $h(v)$ and $f(v)$? Hence find $h(80)$ and $h'(80)$.

   (c) How would you explain the practical meaning of these function and derivative values to a driver who knows no calculus? Include units on each of the function and derivative values you discuss in your response.

44) An object moving vertically has its height at time t (measured in feet, with time in seconds) given by the function $h(t) = 3 + \frac{2\cos(t)}{1.2^t}$.

   (a) What is the object's instantaneous velocity when $t = 2$?

   (b) What is the object's acceleration at the instant $t = 2$?

   (c) Describe in everyday language the behavior of the object at the instant $t = 2$.

45) Let $f(x) = \sin(x)\cot(x)$.

   (a) Use the product rule to find $f'(x)$.

   (b) True or false: for all real numbers x, $f(x) = \cos(x)$.

   (c) Explain why the function that you found in (a) is almost the opposite of the sine function, but not quite. (Hint: convert all of the trigonometric functions in (a) to sines and cosines, and work to simplify. Think carefully about the domain of f and the domain of f'.)

2.6 *The chain rule*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- What is a composite function and how do we recognize its structure algebraically?
- Given a composite function $C(x) = f(g(x))$ that is built from differentiable functions f and g, how do we compute $C'(x)$? What is the statement of the Chain Rule?

Introduction

In addition to learning how to differentiate a variety of basic functions, we have also been developing our ability to understand how to use rules to differentiate certain algebraic combinations of them. For example, we not only know how to take the derivative of $f(x) = \sin(x)$ and $g(x) = x^2$, but now we can quickly find the derivative of each of the following combinations of f and g:

$$s(x) = 3x^2 - 5\sin(x),$$

$$p(x) = x^2 \sin(x), \text{and}$$

$$q(x) = \frac{\sin(x)}{x^2}.$$

Finding s' uses the sum and constant multiple rules, determining p' requires the product rule, and q' can be attained with the quotient rule. Again, we note the importance of recognizing the algebraic structure of a given function in order to find its derivative: $s(x) = 3g(x) - 5f(x)$, $p(x) = g(x) \cdot f(x)$, and $q(x) = \frac{f(x)}{g(x)}$.

There is one more natural way to algebraically combine basic functions, and that is by *composing* them. For instance, let's consider the function

$$C(x) = \sin(x^2),$$

and observe that any input x passes through a *chain* of functions. In particular, in the process that defines the function $C(x)$, x is first squared, and then the sine of the result is taken. Using an arrow diagram,

$$x \longrightarrow x^2 \longrightarrow \sin(x^2).$$

In terms of the elementary functions f and g, we observe that x is first input in the function g, and then the result is used as the input in f. Said differently, we can write

$$C(x) = f(g(x)) = \sin(x^2)$$

and say that C is the *composition* of f and g.

The main question that we answer in the present section is: given a composite function $C(x) = f(g(x))$ that is built from differentiable functions f and g, how do we compute $C'(x)$ in terms of f, g, f', and g'? In the same way that the rate of change of a product of two functions, $p(x) = f(x) \cdot g(x)$, depends on the behavior of both f and g, it makes sense intuitively that the rate of change of a composite function $C(x) = f(g(x))$ will also depend on some combination of f and g and their derivatives. The rule that describes how to compute C' in terms of f and g and their derivatives will be called the *chain rule*.

But before we can learn what the chain rule says and why it works, we first need to be comfortable decomposing composite functions so that we can correctly identify the functions that produce the composition, as we did in the example above with $C(x) = \sin(x^2)$.

Preview Activity 2.6

For each function given below, identify its fundamental algebraic structure. In particular, is the given function a sum, product, quotient, or composition of basic functions? If the function is a composition of basic functions, state a formula for the functions that make up the composition so that the overall composite function can be written in the form $f(g(x))$. If the function is a sum, product, or quotient of basic functions, use the appropriate rule to determine its derivative.

(a) $h(x) = \tan(2^x)$

(b) $p(x) = 2^x \tan(x)$

(c) $r(x) = (\tan(x))^2$

(d) $m(x) = e^{\tan(x)}$

(e) $w(x) = \sqrt{x} + \tan(x)$

(f) $z(x) = \sqrt{\tan(x)}$

Chaining together rates of change

Before we continue our discussion of the chain rule, let's focus on how we can "chain" the rates of change of some quantity together to produce yet another rate of change.

Suppose three friends x, y, and u are chopping wood, and they each chop at different rates. If y chops twice as fast as u, then we can notate the rate at which y chops wood with respect to u, using one of the alternate notations for the derivative, as

$$\frac{dy}{du} = 2.$$

And the notation is read as "The rate of change of y with respect to u is 2."

If u chops six times as fast as x, then we can notate the rate at which u chops wood with respect to x as

$$\frac{du}{dx} = 6.$$

This notation is read as "The rate of change of u with respect to x is 6."

A natural question would then be "How much faster is y chopping wood than x?" Or "What is the rate of change of y with respect to x?" Or "What is $\frac{dy}{dx}$?"

We can find it by multiplying the rate at which y chops wood with respect to u times the rate at which u chops wood with respect to x, yielding:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

So we have that y chops $12 = 2 \cdot 6$ times faster than x.

Now suppose there are 26 friends a, b, c, d, ..., x, y, and z. Again, they are all chopping wood and we know the following rates of change:

$$\frac{da}{db}, \frac{db}{dc}, \frac{dc}{dd}, \ldots, \frac{dx}{dy}, \text{ and } \frac{dy}{dz}.$$

How can we compute $\frac{da}{dz}$? Once again we'll multiply the rates of change together, yielding

$$\frac{da}{dz} = \frac{da}{db} \cdot \frac{db}{dc} \cdot \frac{dc}{dd} \cdot \ \cdots \ \cdot \frac{dx}{dy} \cdot \frac{dy}{dz}.$$

It is important to realize that we *are not* canceling these terms; the derivative notation of $\frac{dy}{dx}$ or $\frac{db}{dc}$ is one symbol. It is equally important to realize that this notation was chosen precisely because of this behavior. It makes applying the Chain Rule easy with multiple variables. For instance,

$$\frac{dy}{dt} = \frac{dy}{d\bigcirc} \cdot \frac{d\bigcirc}{d\triangle} \cdot \frac{d\triangle}{dt}.$$

where $\bigcirc$ and $\triangle$ are any variables you'd like to use.

One of the most common ways of "visualizing" the Chain Rule is to consider a set of gears, as shown in Figure 2.29. The gears have 36, 18, and 6 teeth, respectively. That means for every revolution of the x gear, the u gear revolves twice. That is: $\frac{du}{dx} = 2$. Likewise, every revolution of u causes 3 revolutions of y: $\frac{dy}{du} = 3$. How does y change with respect to x? For each revolution of x, y revolves 6 times; that is,

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = 2 \cdot 3 = 6.$$

We can then extend the Chain Rule with more variables by adding more gears to the picture.

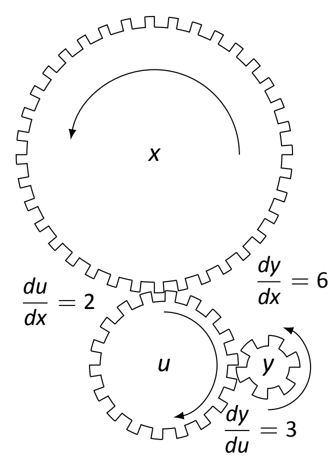


Figure 2.29: A series of gears to demonstrate the Chain Rule. Note how $\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$

The chain rule

Let's now return to the question of differentiating composite functions and the function $f(x) = \sin(x^2)$. From Preview Activity 2.6, we know that we can "decompose" this composition by identifying the two functions that make up the composition. For this function, that would be $\sin(x)$ and x^2, and we can rewrite our original function as

$$y = \sin(u) \quad \text{where} \quad u = x^2$$

by introducing a simple substitution into the composition.

We'd like to know $f'(x)$, or alternately $\frac{dy}{dx}$, and as we saw in the previous subsection, if we know $\frac{dy}{du}$ and $\frac{du}{dx}$, then we have

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

Since we've rewritten $y = \sin(u)$ and $u = x^2$, then their individual respective derivatives are

$$\frac{dy}{du} = \cos(u) \quad \text{and} \quad \frac{du}{dx} = 2x,$$

and we can compute the derivative of the original function as

$$\begin{aligned} \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= \cos(u) \cdot 2x \\ &= \cos(x^2) \cdot 2x \end{aligned}$$

where we write our computed derivative in terms of the variables with which we started.

We now consider a second example to gain further understanding of how differentiating a composite function involves the basic functions that combine to form it.

Example 1

Let $C(x) = \sin(2x)$. Use the double angle identity to rewrite C as a product of basic functions, and use the product rule to find C'. Rewrite C' in the simplest form possible.

Solution. By the double angle identity for the sine function,

$$C(x) = \sin(2x) = 2\sin(x)\cos(x).$$

Applying the product rule and simplifying,

$$C'(x) = 2\sin(x)(-\sin(x)) + \cos(x)(2\cos(x)) = 2(\cos^2(x) - \sin^2(x)).$$

Next, we recall that one of the double angle identities for the cosine

> function tells us that
>
> $$\cos(2x) = \cos^2(x) - \sin^2(x).$$
>
> Substituting this result in our expression for $C'(x)$, we now have that $C'(x) = 2\cos(2x)$.

So from Example 1, we see that if $C(x) = \sin(2x)$, then $C'(x) = 2\cos(2x)$. Letting $u(x) = 2x$ and $f(u) = \sin(u)$, we observe that $C(x) = f(u(x))$. Moreover, with $\frac{du}{dx} = 2$ and $\frac{df}{du} = \cos(u)$, it follows that we can view the structure of $C'(x)$ as

$$C'(x) = \frac{dC}{dx} = 2\cos(2x) = \frac{du}{dx} \cdot \frac{df}{du}.$$

In this example, we see that for the composite function $C(x) = f(u(x))$, the derivative C' is constituted by multiplying the derivatives of f, $\frac{df}{du}$, and u, $\frac{du}{dx}$ but with the special condition that f' is evaluated at $u(x)$.

It makes sense intuitively that these two quantities are involved in understanding the rate of change of a composite function: if we are considering $C(x) = f(u(x))$ and asking how fast C is changing at a given x value as x changes, it clearly matters (a) how fast u is changing with respect to x, as well as how fast f is changing with respect to $u(x)$. It turns out that this structure holds not only for the functions in Example 1, but indeed for all differentiable functions[6] as is stated in the Chain Rule.

[6] Like other differentiation rules, the Chain Rule can be proved formally using the limit definition of the derivative.

Chain Rule

If g is differentiable at x and f is differentiable at $g(x)$, then the composite function C defined by $C(x) = f(g(x))$ is differentiable at x and

$$C'(x) = f'(g(x))g'(x).$$

Alternately, if y can be written as a differentiable function of u where u is a differentiable function of x, then

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

This version is know as the Leibniz notation of the Chain Rule.

As with the product and quotient rules, it is often helpful to think verbally about what the chain rule says: "If C is a composite function defined by an *outer* function f and an *inner* function

g, then C' is given by the derivative of the outer function, evaluated at the inner function, times the derivative of the inner function."

At least initially in working particular examples requiring the chain rule, it can be also be helpful to clearly identify the inner function g and outer function f, compute their derivatives individually, and then put all of the pieces together to generate the derivative of the overall composite function. To see what we mean by this, consider the function

$$r(x) = (\tan(x))^2.$$

The function r is composite, with inner function $g(x) = \tan(x)$ and outer function $f(x) = x^2$. Organizing the key information involving f, g, and their derivatives, we have

$$\begin{array}{ll} f(x) = x^2 & g(x) = \tan(x) \\ f'(x) = 2x & g'(x) = \sec^2(x) \\ f'(g(x)) = 2\tan(x) & \end{array}$$

Applying the chain rule, which tells us that $r'(x) = f'(g(x))g'(x)$, we find that for $r(x) = (\tan(x))^2$, its derivative is

$$r'(x) = 2\tan(x)\sec^2(x).$$

Using the Leibniz notation to differentiate this function, we first rewrite

$$r = u^2 \quad \text{where} \quad u = \tan(x).$$

Their individual respective derivatives are

$$\frac{dr}{du} = 2u \quad \text{and} \quad \frac{du}{dx} = \sec^2(x);$$

therefore,

$$\begin{aligned} \frac{dr}{dx} &= \frac{dr}{du} \cdot \frac{du}{dx} \\ &= 2u \cdot \sec^2(x) \\ &= 2\tan(x)\sec^2(x). \end{aligned}$$

As a side note, we remark that another way to write $r(x)$ is $r(x) = \tan^2(x)$. Observe that in this format, the composite nature of the function is more implicit, but this is common notation for powers of trigonometric functions: $\cos^4(x)$, $\sin^5(x)$, and $\sec^2(x)$ are all composite functions, with the outer function a power function and the inner function a trigonometric one.

Example 2

Find the derivatives of the following functions:

1) $y = \sec(3x)$
2) $y = \sqrt{4x^3 - 2x^2}$
3) $y = e^{-x^2}$

Solution.

1) Consider $y = \sec(3x)$. Recognize that this is a composition of functions, where $f(x) = \sec x$ and $g(x) = 3x$. Thus

$$y' = f'(g(x)) \cdot g'(x) = \sec(3x)\tan(3x) \cdot 3 = 3\sec(3x)\tan(3x).$$

Using the Leibniz notation to differentiate this function, we first rewrite

$$y = \sec(u) \quad \text{where} \quad u = 3x.$$

Their individual respective derivatives are

$$\frac{dy}{du} = \sec(u)\tan(u) \quad \text{and} \quad \frac{du}{dx} = 3;$$

therefore,

$$\begin{aligned} \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= \sec(u)\tan(u) \cdot 3 \\ &= 3\sec(3x)\tan(3x). \end{aligned}$$

2) Recognize that $y = \sqrt{4x^3 - 2x^2}$ is the composition of $f(x) = \sqrt{x}$ and $g(x) = 4x^3 - 2x^2$. Also, recall that

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

This leads us to:

$$\begin{aligned} y' &= \frac{1}{2\sqrt{4x^3 - 2x^2}} \cdot (12x^2 - 4x) \\ &= \frac{12x^2 - 4x}{2\sqrt{4x^3 - 2x^2}} = \frac{4x(3x-1)}{2\sqrt{4x^3 - 2x^2}} \\ &= \frac{2x(3x-1)}{\sqrt{4x^3 - 2x^2}}. \end{aligned}$$

Using the Leibniz notation to differentiate this function, we first rewrite

$$y = \sqrt{u} \quad \text{where} \quad u = 4x^3 - 2x^2.$$

Their individual respective derivatives are

$$\frac{dy}{du} = \frac{1}{2}u^{-1/2} \quad \text{and} \quad \frac{du}{dx} = 12x^2 - 4x;$$

therefore,

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{du}\cdot\frac{du}{dx}\\ &= \frac{1}{2\sqrt{u}}\cdot 12x^2-4x\\ &= \frac{2x(3x-1)}{\sqrt{4x^3-2x^2}}.\end{aligned}$$

3) Recognize that $y = e^{-x^2}$ is the composition of $f(x) = e^x$ and $g(x) = -x^2$. Remembering that $f'(x) = e^x$, we have

$$y' = e^{-x^2}\cdot(-2x) = (-2x)e^{-x^2}.$$

Using the Leibniz notation to differentiate this function, we first rewrite

$$y = e^u \quad \text{where} \quad u = -x^2.$$

Their individual respective derivatives are

$$\frac{dy}{du} = e^u \quad \text{and} \quad \frac{du}{dx} = -2x;$$

therefore,

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{du}\cdot\frac{du}{dx}\\ &= e^u\cdot(-2x)\\ &= (-2x)e^{-x^2}.\end{aligned}$$

The chain rule now substantially expands the library of functions we can differentiate, as the following activity demonstrates.

Activity 2.6–1

For each function given below, identify an inner function g and outer function f to write the function in the form $f(g(x))$. Then, determine $f'(x)$, $g'(x)$, and $f'(g(x))$, and finally apply the chain rule to determine the derivative of the given function. Also, use the Leibniz notation of the chain rule for each function.

(a) $h(x) = \cos(x^4)$

(b) $p(x) = \sqrt{\tan(x)}$

(c) $s(x) = 2^{\sin(x)}$

(d) $z(x) = \cot^5(x)$

(e) $m(x) = (\sec(x) + e^x)^9$

Using multiple rules simultaneously

The chain rule now joins the sum, constant multiple, product, and quotient rules in our collection of the different techniques for finding the derivative of a function through understanding its algebraic structure and the basic functions that constitute it. It takes substantial practice to get comfortable with navigating multiple rules in a single problem; using proper notation and taking a few extra steps can be particularly helpful as well. We demonstrate with an example and then provide further opportunity for practice in the following activity.

Example 3

Find a formula for the derivative of $h(t) = 3^{t^2+2t}\sec^4(t)$.

Solution. We first observe that the most basic structure of h is that it is the product of two functions: $h(t) = a(t) \cdot b(t)$ where $a(t) = 3^{t^2+2t}$ and $b(t) = \sec^4(t)$. Therefore, we see that we will need to use the product rule to differentiate h. When it comes time to differentiate a and b in their roles in the product rule, we observe that since each is a composite function, the chain rule will be needed. We therefore begin by working separately to compute $a'(t)$ and $b'(t)$.

Writing $a(t) = f(g(t)) = 3^{t^2+2t}$, and finding the derivatives of f and g, we have

$$\begin{aligned} f(t) &= 3^t & g(t) &= t^2 + 2t \\ f'(t) &= 3^t \ln(3) & g'(t) &= 2t + 1 \\ f'(g(t)) &= 3^{t^2+2t}\ln(3) \end{aligned}$$

Thus, by the chain rule, it follows that $a'(t) = f'(g(t))g'(t) = 3^{t^2+2t}\ln(3)(2t + 1)$.

Turning next to b, we write $b(t) = r(s(t)) = \sec^4(t)$ and find the derivatives of r and g. Doing so,

$$\begin{aligned} r(t) &= t^4 & s(t) &= \sec(t) \\ r'(t) &= 4t^3 & s'(t) &= \sec(t)\tan(t) \\ r'(s(t)) &= 4\sec^3(t) \end{aligned}$$

By the chain rule, we now know that

$$b'(t) = r'(s(t))s'(t) = 4\sec^3(t)\sec(t)\tan(t) = 4\sec^4(t)\tan(t).$$

Now we are finally ready to compute the derivative of the overall function h. Recalling that $h(t) = 3^{t^2+2t}\sec^4(t)$, by the product rule we have

$$h'(t) = 3^{t^2+2t}\frac{d}{dt}[\sec^4(t)] + \sec^4(t)\frac{d}{dt}[3^{t^2+2t}].$$

From our work above with a and b, we know the derivatives of 3^{t^2+2t} and $\sec^4(t)$, and therefore

$$h'(t) = 3^{t^2+2t}4\sec^4(t)\tan(t) + \sec^4(t)3^{t^2+2t}\ln(3)(2t+2).$$

Example 4

Find the derivatives of $f(x) = \dfrac{5x^3}{e^{-x^2}}$.

Solution. We must employ the Quotient Rule along with the Chain

Rule. Again, proceed step–by–step.

$$\begin{aligned} f'(x) &= \frac{e^{-x^2}(15x^2) - 5x^3((-2x)e^{-x^2})}{(e^{-x^2})^2} \\ &= \frac{e^{-x^2}(30x^4 + 15x^2)}{e^{-2x^2}} \\ &= e^{x^2}(30x^4 + 15x^2). \end{aligned}$$

Activity 2.6–2

For each of the following functions, find the function's derivative. State the rule(s) you use, label relevant derivatives appropriately, and be sure to clearly identify your overall answer.

(a) $p(r) = 4\sqrt{r^6 + 2e^r}$

(b) $m(v) = \sin(v^2)\cos(v^3)$

(c) $h(y) = \dfrac{\cos(10y)}{e^{4y} + 1}$

(d) $s(z) = 2^{z^2 \sec(z)}$

(e) $c(x) = \sin(e^{x^2})$

The chain rule now adds substantially to our ability to do different familiar problems that involve derivatives. Whether finding the equation of the tangent line to a curve, the instantaneous velocity of a moving particle, or the instantaneous rate of change of a certain quantity, if the function under consideration involves a composition of other functions, the chain rule is indispensable.

Activity 2.6–3

Use known derivative rules, including the chain rule, as needed to answer each of the following questions.

(a) Find an equation for the tangent line to the curve $y = \sqrt{e^x + 3}$ at the point where $x = 0$.

(b) If $s(t) = \dfrac{1}{(t^2 + 1)^3}$ represents the position function of a particle moving horizontally along an axis at time t (where s is measured in inches and t in seconds), find the particle's instantaneous velocity at $t = 1$. Is the particle moving to the left or right at that instant?

(c) At sea level, air pressure is 30 inches of mercury. At an altitude of h feet above sea level, the air pressure, P, in inches of mercury, is given by the function

$$P = 30e^{-0.0000323h}.$$

Compute dP/dh and explain what this derivative function tells you about air pressure, including a discussion of the units on dP/dh. In addition, determine how fast the air pressure is changing for a pilot of a small plane passing through an altitude of 1000 feet.

(d) Suppose that $f(x)$ and $g(x)$ are differentiable functions and that the following information about them is known:

x	$f(x)$	$f'(x)$	$g(x)$	$g'(x)$
-1	2	-5	-3	4
2	-3	4	-1	2

If $C(x)$ is a function given by the formula $f(g(x))$, determine $C'(2)$. In addition, if $D(x)$ is the function $f(f(x))$, find $D'(-1)$.

The composite version of basic function rules

As we gain more experience with differentiating complicated functions, we will become more comfortable in the process of simply writing down the derivative without taking multiple steps. We demonstrate part of this perspective here by showing how we can find a composite rule that corresponds to two of our basic functions. For instance, we know that $\frac{d}{dx}[\sin(x)] = \cos(x)$. If we instead want to know

$$\frac{d}{dx}[\sin(u(x))],$$

where u is a differentiable function of x, then this requires the chain rule with the sine function as the outer function. Applying the chain rule,

$$\frac{d}{dx}[\sin(u(x))] = \cos(u(x)) \cdot u'(x).$$

Similarly, since $\frac{d}{dx}[a^x] = a^x \ln(a)$, it follows by the chain rule that

$$\frac{d}{dx}[a^{u(x)}] = a^{u(x)} \ln(a).$$

In the process of getting comfortable with derivative rules, an excellent exercise is to write down a list of all basic functions whose derivatives are known, list those derivatives, and then write the corresponding chain rule for the composite version with the inner function being an unknown function $u(x)$ and the outer function being the known basic function. These versions of the chain rule are particularly simple when the inner function is linear, since the derivative of a linear function is a constant. For instance,

$$\frac{d}{dx}\left[(5x+7)^{10}\right] = 10(5x+7)^9 \cdot 5,$$

$$\frac{d}{dx}[\tan(17x)] = 17\sec^2(17x), \text{ and}$$

$$\frac{d}{dx}\left[e^{-3x}\right] = -3e^{-3x}.$$

Summary

In this section, we encountered the following important ideas:

- A composite function is one where the input variable x first passes through one function, and then the resulting output passes through another. For example, the function $h(x) = 2^{\sin(x)}$ is composite since $x \longrightarrow \sin(x) \longrightarrow 2^{\sin(x)}$.
- Given a composite function $C(x) = f(g(x))$ that is built from differentiable functions f and g, the chain rule tells us that we compute $C'(x)$ in terms of f, g, f', and g' according to the formula

$$C'(x) = f'(g(x))g'(x).$$

- Given a composite function $y = f(g(x))$ that can be rewritten as $y = f(u)$ where $u = g(x)$, then the Leibniz version of the chain rule tells us that we compute $\frac{dy}{dx}$ according to the formula

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

Exercises

Terms and Concepts

1) T/F: The chain rule describes how to evaluate the derivative of a composition of functions.

2) T/F: The Generalized Power Rule states that $\frac{d}{dx}\Big(g(x)^n\Big) = n\big(g(x)\big)^{n-1}$.

3) T/F: $\frac{d}{dx}(3^x) \approx 1.1 \cdot 3^x$.

4) T/F: $\frac{dx}{dy} = \frac{dx}{dt} \cdot \frac{dt}{dy}$

5) Let $u(x)$ be a differentiable function. For each of the following functions, determine the derivative. Each response will involve u and/or u'.
   (a) $p(x) = e^{u(x)}$
   (b) $q(x) = u(e^x)$
   (c) $r(x) = \cot(u(x))$
   (d) $s(x) = u(\cot(x))$
   (e) $a(x) = u(x^4)$
   (f) $b(x) = u^4(x)$

Problems

In exercises 6–23, compute the derivative of the given function.

6) $f(x) = (4x^3 - x)^{10}$

7) $f(t) = (3t-2)^5$

8) $g(x) = (\sin(x) + \cos(x))^3$

9) $h(t) = e^{3t^2+t-1}$

10) $f(x) = \left(x + \frac{1}{x}\right)^4$

11) $f(x) = \cos(3x)$

12) $g(x) = \tan(5x)$

13) $h(t) = \sin^4(2t)$

14) $p(t) = \cos^3(t^2 + 3t + 1)$

15) $f(x) = 4^r$

16) $g(t) = 5^{\cos(t)}$

17) $g(t = 15^2$

18) $m(w) = \frac{3^w}{2^w}$

19) $m(w) = \frac{3^w + 1}{2^w}$

20) $f(x) = \frac{3^{x^2} + x}{2^{x^2}}$

21) $f(x) = x^2\sin(5x)$

22) $g(x) = \cos(t^2 + 3t)\sin(5t - 7)$

23) $g(t) = \cos\left(\frac{1}{x}\right)e^{5x^2}$

In exercises 24–27, find the equation of the tangent line to the graph of the function at the given point.

24) $f(x) = (4x^3 - x)^{10}$ at $x = 0$

25) $f(t) = (3t-2)^5$ at $t = 1$

26) $g(x) = (\sin(x) + \cos(x))^3$ at $x = \pi/2$

27) $h(t) = e^{3t^2+t-1}$ at $t = -1$

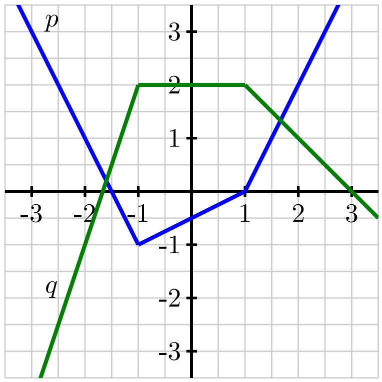
28) Consider the basic functions $f(x) = x^3$ and $g(x) = \sin(x)$.
   (a) Let $h(x) = f(g(x))$. Find the exact instantaneous rate of change of h at the point where $x = \frac{\pi}{4}$.
   (b) Which function is changing most rapidly at $x = 0.25$: $h(x) = f(g(x))$ or $r(x) = g(f(x))$? Why?
   (c) Let $h(x) = f(g(x))$ and $r(x) = g(f(x))$. Which of these functions has a derivative that is periodic? Why?

29) If a spherical tank of radius 4 feet has h feet of water present in the tank, then the volume of water in the tank is given by the formula

$$V = \frac{\pi}{3}h^2(12 - h).$$

   (a) At what instantaneous rate is the volume of water in the tank changing with respect to the *height* of the water at the instant $h = 1$? What are the units on this quantity?
   (b) Now suppose that the height of water in the tank is being regulated by an inflow and outflow (e.g., a faucet and a drain) so that the height of the water at time t is given by the rule $h(t) = \sin(\pi t) + 1$, where t is measured in hours (and h is still measured in feet). At what rate is the height of the water changing with respect to time at the instant $t = 2$?
   (c) Continuing under the assumptions in (b), at what instantaneous rate is the volume of water in the tank changing with respect to *time* at the instant $t = 2$?
   (d) What are the main differences between the rates found in (a) and (c)? Include a discussion of the relevant units.

30) Let functions p and q be the piecewise linear functions given by their respective graphs below. Use the graphs to answer the following questions.



(a) Let $C(x) = p(q(x))$. Determine $C'(0)$ and $C'(3)$.

(b) Find a value of x for which $C'(x)$ does not exist. Explain your thinking.

(c) Let $Y(x) = q(q(x))$ and $Z(x) = q(p(x))$. Determine $Y'(-2)$ and $Z'(0)$.

2.7 *Derivatives of functions given implicitly*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- What does it mean to say that a curve is an implicit function of x, rather than an explicit function of x?
- How does implicit differentiation enable us to find a formula for $\frac{dy}{dx}$ when y is an implicit function of x?
- In the context of an implicit curve, how can we use $\frac{dy}{dx}$ to answer important questions about the tangent line to the curve?

Introduction

In all of our studies with derivatives to date, we have worked in a setting where we can express a formula for the function of interest explicitly in terms of x. But there are many interesting curves that are determined by an equation involving x and y for which it is impossible to solve for y in terms of x.

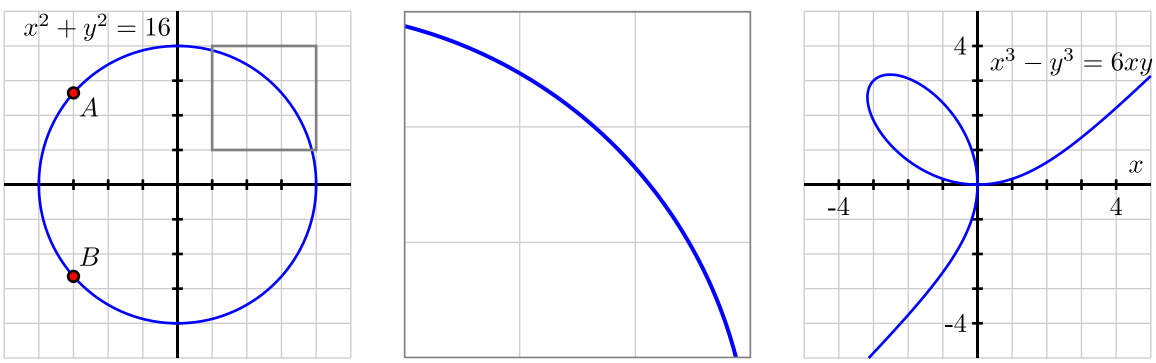


Figure 2.30: At left, the circle given by $x^2 + y^2 = 16$. In the middle, the portion of the circle $x^2 + y^2 = 16$ that has been highlighted in the box at left. And at right, the lemniscate given by $x^3 - y^3 = 6xy$.

Perhaps the simplest and most natural of all such curves are circles. Because of the circle's symmetry, for each x value strictly between the endpoints of the horizontal diameter, there are two corresponding y-values. For instance, in Figure 2.30, we have labeled $A = (-3, \sqrt{7})$ and $B = (-3, -\sqrt{7})$, and these points demonstrate that the circle fails the vertical line test. Hence, it is impossible to represent the circle through a single function of the form $y = f(x)$. At the same time, portions of the circle can be represented explicitly as a function of x, such as the highlighted arc that is magnified in the center of Figure 2.30. Moreover, it is evident that the circle is locally linear, so we ought to be able to find a tangent line to the curve at every point; thus, it makes sense to wonder if we can compute $\frac{dy}{dx}$ at any point on the circle, even though we cannot write y explicitly as a function of x. Finally, we note that the righthand curve in Figure 2.30 is called a *lemniscate* and is just one of many fascinating possibilities for implicitly given curves.

In working with implicit functions, we will often be interested in finding an equation for $\frac{dy}{dx}$ that tells us the slope of the tangent line to the curve at a point (x, y). To do so, it will be necessary for us to work with y while thinking of y as a function of x, but without being able to write an explicit formula for y in terms of x. The following preview activity reminds us of some ways we can compute derivatives of functions in settings where the function's formula is not known. For instance, recall the earlier example $\frac{d}{dx}[e^{u(x)}] = e^{u(x)}u'(x)$.

Preview Activity 2.7

Let f be a differentiable function of x (whose formula is not known) and recall that $\frac{d}{dx}[f(x)]$ and $f'(x)$ are interchangeable notations. Determine each of the following derivatives of combinations of explicit functions of x, the unknown function f, and an arbitrary constant c.

(a) $\frac{d}{dx}\left[x^2 + f(x)\right]$

(b) $\frac{d}{dx}\left[x^2 f(x)\right]$

(c) $\frac{d}{dx}\left[c + x + f(x)^2\right]$

(d) $\frac{d}{dx}\left[f(x^2)\right]$

(e) $\frac{d}{dx}\left[xf(x) + f(cx) + cf(x)\right]$

Implicit Differentiation

Because a circle is perhaps the simplest of all curves that cannot be represented explicitly as a single function of x, we begin our exploration of implicit differentiation with the example of the circle given by $x^2 + y^2 = 16$. It is visually apparent that this curve is locally linear, so it makes sense for us to want to find the slope of the tangent line to the curve at any point, and moreover to think that the curve is differentiable. The big question is: how do we find a formula for $\frac{dy}{dx}$, the slope of the tangent line to the circle at a given point on the circle? By viewing y as an *implicit*[7] function of x, we essentially think of y as some function whose formula $f(x)$ is unknown, but which we can differentiate. Just as y represents an unknown formula, so too its derivative with respect to x, $\frac{dy}{dx}$, will be (at least temporarily) unknown.

[7] Essentially the idea of an implicit function is that it can be broken into pieces where each piece can be viewed as an explicit function of x, and the combination of those pieces constitutes the full implicit function. For the circle, we could choose to take the top half as one explicit function of x, and the bottom half as another.

Consider the equation $x^2 + y^2 = 16$ and view y as an unknown differentiable function of x. Differentiating both sides of the equation with respect to x, we have

$$\frac{d}{dx}\left[x^2 + y^2\right] = \frac{d}{dx}[16].$$

On the right, the derivative of the constant 16 is 0, and on the left we can apply the sum rule, so it follows that

$$\frac{d}{dx}\left[x^2\right] + \frac{d}{dx}\left[y^2\right] = 0.$$

Next, it is essential that we recognize the different roles being played by x and y. Since x is the independent variable, it is the variable with respect to which we are differentiating, and thus $\frac{d}{dx}\left[x^2\right] = 2x$. But y is the dependent variable and y is an implicit function of x. Thus, when we want to compute $\frac{d}{dx}[y^2]$ it is identical to the situation in Preview Activity 2.7 where we computed $\frac{d}{dx}[f(x)^2]$. In both situations, we have an unknown function being squared, and we seek the derivative of the result. This requires the chain rule, by which we find that $\frac{d}{dx}[y^2] = 2y^1\frac{dy}{dx}$. Therefore, continuing our work in differentiating both sides of $x^2 + y^2 = 16$, we now have that

$$2x + 2y\frac{dy}{dx} = 0.$$

Since our goal is to find an expression for $\frac{dy}{dx}$, we solve this most recent equation for $\frac{dy}{dx}$. Subtracting $2x$ from both sides and dividing by $2y$,

$$\frac{dy}{dx} = -\frac{2x}{2y} = -\frac{x}{y}.$$

There are several important things to observe about the result that $\frac{dy}{dx} = -\frac{x}{y}$. First, this expression for the derivative involves both x and y. It makes sense that this should be the case, since for each value of x between -4 and 4, there are two corresponding points on the circle, and the slope of the tangent line is different at each of these points. Second, this formula is entirely consistent with our understanding of circles. If we consider the radius from the origin to the point (a, b), the slope of this line segment is $m_r = \frac{b}{a}$. The tangent line to the circle at (a, b) will be perpendicular to the radius, and thus have slope $m_t = -\frac{a}{b}$, as shown in Figure 2.31. Finally, the slope of the tangent line is zero at $(0, 4)$ and $(0, -4)$, and is undefined at $(-4, 0)$ and $(4, 0)$; all of these values are consistent with the formula $\frac{dy}{dx} = -\frac{x}{y}$.

We consider the following more complicated example to investigate and demonstrate some additional algebraic issues that arise in problems involving implicit differentiation.

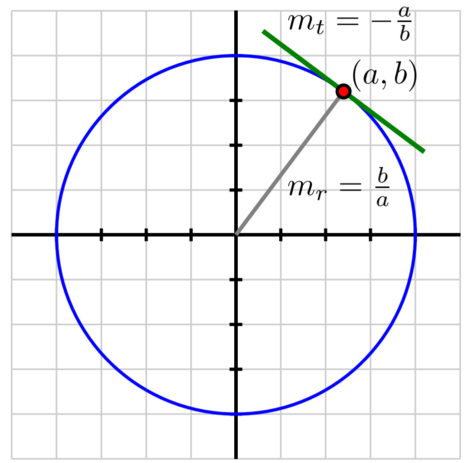


Figure 2.31: The circle given by $x^2 + y^2 = 16$ with point (a, b) on the circle and the tangent line at that point, with labeled slopes of the radial line, m_r, and tangent line, m_t.

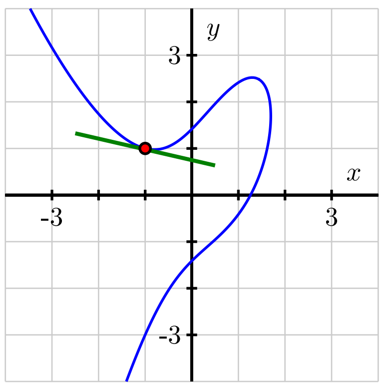


Figure 2.32: The curve $x^3 + y^2 - 2xy = 2$.

Example 1

For the curve given implicitly by $x^3 + y^2 - 2xy = 2$, shown in Figure 2.32, find the slope of the tangent line at $(-1, 1)$.

Solution. We begin by differentiating the curve's equation implicitly. Taking the derivative of each side with respect to x,

$$\frac{d}{dx}\left[x^3 + y^2 - 2xy\right] = \frac{d}{dx}[2],$$

by the sum rule and the fact that the derivative of a constant is zero, we have

$$\frac{d}{dx}[x^3] + \frac{d}{dx}[y^2] - \frac{d}{dx}[2xy] = 0.$$

For the three derivatives we now must execute, the first uses the simple power rule, the second requires the chain rule (since y is an implicit function of x), and the third necessitates the product rule (again since y is a function of x). Applying these rules, we now find that

$$3x^2 + 2y\frac{dy}{dx} - [2x\frac{dy}{dx} + 2y] = 0.$$

Remembering that our goal is to find an expression for $\frac{dy}{dx}$ so that we can determine the slope of a particular tangent line, we want to solve the preceding equation for $\frac{dy}{dx}$. To do so, we get all of the terms involving $\frac{dy}{dx}$ on one side of the equation and then factor. Expanding and then subtracting $3x^2 - 2y$ from both sides, it follows that

$$2y\frac{dy}{dx} - 2x\frac{dy}{dx} = 2y - 3x^2.$$

Factoring the left side to isolate $\frac{dy}{dx}$, we have

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

Finally, we divide both sides by $(2y - 2x)$ and conclude that

$$\frac{dy}{dx} = \frac{2y - 3x^2}{2y - 2x}.$$

Here again, the expression for $\frac{dy}{dx}$ depends on both x and y. To find the slope of the tangent line at $(-1, 1)$, we substitute this point in the formula for $\frac{dy}{dx}$, using the notation

$$\left.\frac{dy}{dx}\right|_{(-1,1)} = \frac{2(1) - 3(-1)^2}{2(1) - 2(-1)} = -\frac{1}{4}.$$

This value matches our visual estimate of the slope of the tangent line shown in Figure 2.32.

Example 1 shows that it is possible when differentiating implicitly to have multiple terms involving $\frac{dy}{dx}$. Regardless of the particular curve involved, our approach will be similar each time. After differentiating, we expand so that each side of the equation is a sum of terms, some of which involve $\frac{dy}{dx}$. Next, addition and subtraction are used to get all terms involving $\frac{dy}{dx}$ on one side of the equation, with all remaining terms on the other. Finally, we factor to get a single instance of $\frac{dy}{dx}$, and then divide to solve for $\frac{dy}{dx}$.

Note, too, that since $\frac{dy}{dx}$ is often a function of both x and y, we use the notation

$$\left.\frac{dy}{dx}\right|_{(a,b)}$$

to denote the evaluation of $\frac{dy}{dx}$ at the point (a, b). This is analogous to writing $f'(a)$ when f' depends on a single variable.

Finally, there is a big difference between writing $\frac{d}{dx}$ and $\frac{dy}{dx}$. For example,

$$\frac{d}{dx}[x^2 + y^2]$$

gives an instruction to take the derivative with respect to x of the quantity $x^2 + y^2$, presumably where y is a function of x. On the other hand,

$$\frac{dy}{dx}(x^2 + y^2)$$

means the product of the derivative of y with respect to x with the quantity $x^2 + y^2$. Understanding this notational subtlety is essential.

Example 2

Find the equation of the line tangent to the curve of the implicitly defined function $\sin(y) + y^3 = 6 - x^3$ at the point $(\sqrt[3]{6}, 0)$.

Solution. We start by taking the derivative of both sides (thus maintaining the equality.) We have :

$$\frac{d}{dx}\Big(\sin(y) + y^3\Big) = \frac{d}{dx}\Big(6 - x^3\Big).$$

The right hand side is easy; it returns $-3x^2$.

The left hand side requires more consideration. We take the derivative term–by–term, and we can see that

$$\frac{d}{dx}\Big(\sin(y)\Big) = \cos(y) \cdot y'.$$

We apply the same process to the y^3 term.

$$\frac{d}{dx}\Big(y^3\Big) = \frac{d}{dx}\Big((y)^3\Big) = 3(y)^2 \cdot y'.$$

Putting this together with the right hand side, we have

$$\cos(y)y' + 3y^2y' = -3x^2.$$

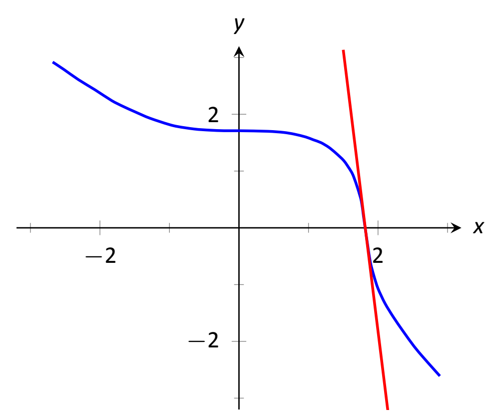


Figure 2.33: The function $\sin(y) + y^3 = 6 - x^3$ and its tangent line at the point $(\sqrt[3]{6}, 0)$.

Now solve for y'.

$$\begin{aligned}\cos(y)y' + 3y^2y' &= -3x^2. \\ \left(\cos(y) + 3y^2\right)y' &= -3x^2 \\ y' &= \frac{-3x^2}{\cos(y) + 3y^2}\end{aligned}$$

We find the slope of the tangent line at the point $(\sqrt[3]{6}, 0)$ by substituting $\sqrt[3]{6}$ for x and 0 for y. Thus at the point $(\sqrt[3]{6}, 0)$, we have the slope as

$$y' = \frac{-3(\sqrt[3]{6})^2}{\cos(0) + 3 \cdot 0^2} = \frac{-3\sqrt[3]{36}}{1} \approx -9.91.$$

Therefore the equation of the tangent line to the implicitly defined function $\sin(y) + y^3 = 6 - x^3$ at the point $(\sqrt[3]{6}, 0)$ is

$$y = -3\sqrt[3]{36}(x - \sqrt[3]{6}) + 0 \approx -9.91x + 18.$$

The curve and this tangent line are shown in Figure 2.33.

Activity 2.7–1

Consider the curve defined by the equation $x = y^5 - 5y^3 + 4y$, whose graph is pictured in Figure 2.34.

(a) Explain why it is not possible to express y as an explicit function of x.

(b) Use implicit differentiation to find a formula for dy/dx.

(c) Use your result from part (b) to find an equation of the line tangent to the graph of $x = y^5 - 5y^3 + 4y$ at the point $(0, 1)$.

(d) Use your result from part (b) to determine all of the points at which the graph of $x = y^5 - 5y^3 + 4y$ has a vertical tangent line.

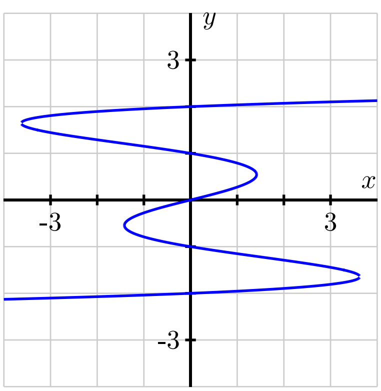


Figure 2.34: The curve $x = y^5 - 5y^3 + 4y$.

Two natural questions to ask about any curve involve where the tangent line can be vertical or horizontal. To be horizontal, the slope of the tangent line must be zero, while to be vertical, the slope must be undefined. It is typically the case when differentiating implicitly that the formula for $\frac{dy}{dx}$ is expressed as a quotient of functions of x and y, say

$$\frac{dy}{dx} = \frac{p(x, y)}{q(x, y)}.$$

Thus, we observe that the tangent line will be horizontal precisely when the numerator is zero and the denominator is nonzero, making the slope of the tangent line zero. Similarly, the tangent line will be vertical whenever $q(x, y) = 0$ and $p(x, y) \neq 0$, making the slope undefined. If both x and y are involved in an equation such as $p(x, y) = 0$, we try to solve for one of them in terms of the other, and then use the resulting condition in the original equation that defines the curve to find an equation in a single variable that we can solve to determine the point(s) that lie on the curve at which the condition holds. It is not always possible to execute the desired algebra due to the possibly complicated combinations of functions that often arise.

Example 3

Given the implicitly defined function $\sin(x^2y^2) + y^3 = x + y$, find y'.

Solution. Differentiating term by term, we find the most difficulty in the first term. It requires both the Chain and Product Rules.

$$\begin{aligned}\frac{d}{dx}\Big(\sin(x^2y^2)\Big) &= \cos(x^2y^2)\cdot\frac{d}{dx}\Big(x^2y^2\Big)\\ &= \cos(x^2y^2)\cdot\big(x^2(2yy') + 2xy^2\big)\\ &= 2(x^2yy' + xy^2)\cos(x^2y^2).\end{aligned}$$

We leave the derivatives of the other terms to the reader. After taking the derivatives of both sides, we have

$$2(x^2yy' + xy^2)\cos(x^2y^2) + 3y^2y' = 1 + y'.$$

We now have to be careful to properly solve for y', particularly because of the product on the left. It is best to multiply out the product. Doing this, we get

$$2x^2y\cos(x^2y^2)y' + 2xy^2\cos(x^2y^2) + 3y^2y' = 1 + y'.$$

From here we can safely move around terms to get the following:

$$2x^2y\cos(x^2y^2)y' + 3y^2y' - y' = 1 - 2xy^2\cos(x^2y^2).$$

Then we can solve for y' to get

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

A graph of this implicit function is given in Figure 2.35. It is easy to verify that the points $(0,0)$, $(0,1)$ and $(0,-1)$ all lie on the graph. We can find the slopes of the tangent lines at each of these points using our formula for y'.

At $(0,0)$, the slope is -1.
At $(0,1)$, the slope is $1/2$.
At $(0,-1)$, the slope is also $1/2$.

The tangent lines have been added to the graph of the function in Figure 2.36.

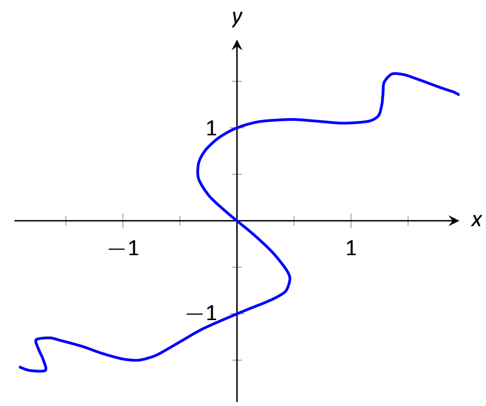


Figure 2.35: A graph of the implicitly defined function $\sin(x^2y^2) + y^3 = x + y$.

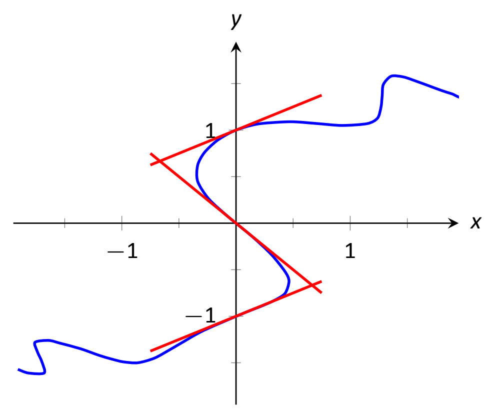


Figure 2.36: A graph of the implicitly defined function $\sin(x^2y^2) + y^3 = x + y$ and certain tangent lines.

Activity 2.7–2

Consider the curve defined by the equation $y(y^2-1)(y-2) = x(x-1)(x-2)$, whose graph is pictured in Figure 2.37.

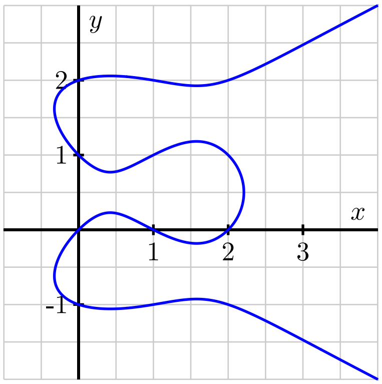


Figure 2.37: The curve $y(y^2-1)(y-2) = x(x-1)(x-2)$.

Through implicit differentiation, it can be shown that

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

Use this fact to answer each of the following questions.

(a) Determine all points (x, y) at which the tangent line to the curve is horizontal. (Use technology appropriately to find the needed zeros of the relevant polynomial function.)

(b) Determine all points (x, y) at which the tangent line is vertical. (Use technology appropriately to find the needed zeros of the relevant polynomial function.)

(c) Find the equation of the tangent line to the curve at one of the points where $x = 1$.

Implicit Differentiation and the Second Derivative

We can use implicit differentiation to find higher order derivatives. In theory, this is simple: first find $\frac{dy}{dx}$, then take its derivative with respect to x. In practice, it is not hard, but it often requires a bit of algebra. We demonstrate this in an example.

Example 4

Given $x^2 + y^2 = 1$, find $\dfrac{d^2y}{dx^2} = y''$.

Solution. Taking derivatives, we get $2x + 2yy' = 0$. Solving for y' gives:

$$y' = \frac{-x}{y}.$$

To find y'', we apply implicit differentiation to y'.

$$\begin{aligned} y'' &= \frac{d}{dx}(y') \\ &= \frac{d}{dx}\left(-\frac{x}{y}\right) \\ &= -\frac{y(1) - x(y')}{y^2} \end{aligned}$$

replace y' with $-x/y$:

$$\begin{aligned} &= -\frac{y - x(-x/y)}{y^2} \\ &= -\frac{y + x^2/y}{y^2} \end{aligned}$$

While this is not a particularly simple expression, it is usable. We can see that $y'' > 0$ when $y < 0$ and $y'' < 0$ when $y > 0$. In Section **??**, we will see how this relates to the shape of the graph.

The closing activity in this section offers more opportunities to practice implicit differentiation.

Activity 2.7–3

For each of the following curves, use implicit differentiation to find dy/dx, d^2y/dx^2, and determine the equation of the tangent line at the given point.

(a) $x^3 - y^3 = 6xy, \quad (-3, 3)$

(b) $\sin(y) + y = x^3 + x, \quad (0, 0)$

(c) $xe^{-xy} = y^2, \quad (0.571433, 1)$

Summary

In this section, we encountered the following important ideas:

- When we have an equation involving x and y where y cannot be solved for explicitly in terms of x, but where portions of the curve can be thought of as being generated by explicit functions of x, we say that y is an implicit function of x. A good example of such a curve is the unit circle.
- In the process of implicit differentiation, we take the equation that generates an implicitly given curve and differentiate both sides with respect to x while treating y as a function of x. In so doing, the chain rule leads $\frac{dy}{dx}$ to arise, and then we may subsequently solve for $\frac{dy}{dx}$ using algebra.
- While $\frac{dy}{dx}$ may now involve both the variables x and y, $\frac{dy}{dx}$ still measures the slope of the tangent line to the curve, and thus this derivative may be used to decide when the tangent line is horizontal ($\frac{dy}{dx} = 0$) or vertical ($\frac{dy}{dx}$ is undefined), or to find the equation of the tangent line at a particular point on the curve.

Exercises

Terms and Concepts

1) In your own words, explain the difference between implicit functions and explicit functions.

2) Implicit differentiation is based on what other differentiation rule?

Problems

In exercises 3–6, compute the derivative of the given function.

3) $f(x) = \sqrt[3]{x}$

4) $f(t) = \sqrt{1 - t^2}$

5) $g(t) = \sqrt{t} \sin t$

6) $h(x) = x^{1.5}$

In exercises 7–17, compute the derivative using implicit differentiation.

7) $x^4 + y^2 + y = 7$

8) $x^{2/5} + y^{2/5} = 1$

9) $\cos(x) + \sin(y) = 1$

10) $\dfrac{x}{y} = 10$

11) $\dfrac{y}{x} = 10$

12) $x^2 e^2 + 2^y = 5$

13) $x^2 \tan y = 50$

14) $(3x^2 + 2y^3)^4 = 2$

15) $(y^2 + 2y - x)^2 = 200$

16) $\dfrac{x^2 + y}{x + y^2} = 17$

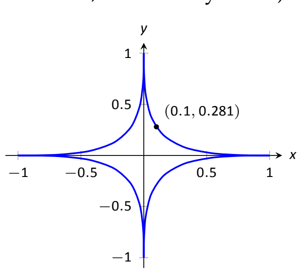
17) $\dfrac{\sin(x) + y}{\cos(y) + x} = 1$

In exercises 18–22, find the equation of the tangent line to the graph of the implicitly defined function at the given points.

18) $x^{2/5} + y^{2/5} = 1$

(a) At $(1, 0)$.

(b) At $(0.1, 0.281)$ (which does not *exactly* lie on the curve, but is very close).

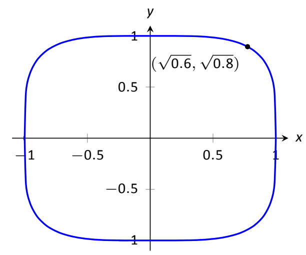


19) $x^4 + y^4 = 1$

(a) At $(1, 0)$.

(b) At $(\sqrt{0.6}, \sqrt{0.8})$.

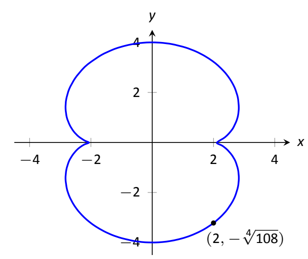
(c) At $(0, 1)$.



20) $(x^2 + y^2 - 4)^3 = 108y^2$

(a) At $(0, 4)$.

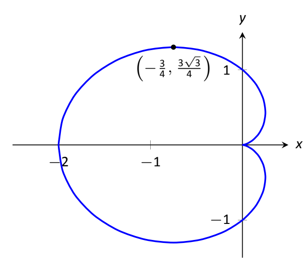
(b) At $(2, -\sqrt[4]{108})$.



21) $(x^2 + y^2 + x)^2 = x^2 + y^2$

(a) At $(0, 1)$.

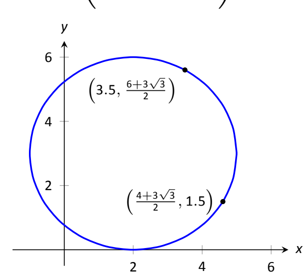
(b) At $\left(-\dfrac{3}{4}, \dfrac{3\sqrt{3}}{4}\right)$.



22) $(x-2)^2 + (y-3)^2 = 9$

(a) At $\left(\dfrac{7}{2}, \dfrac{6 + 3\sqrt{3}}{2}\right)$.

(b) At $\left(\dfrac{4 + 3\sqrt{3}}{2}, \dfrac{3}{2}\right)$.



In exercises 23–26, compute the second derivative of the implicitly defined function.

23) $x^4 + y^2 + y = 7$

24) $x^{2/5} + y^{2/5} = 1$

25) $\cos(x) + \sin(y) = 1$

26) $\dfrac{x}{y} = 10$

27) Consider the curve given by the equation $2y^3 + y^2 - y^5 = x^4 - 2x^3 + x^2$. Find all points at which the tangent line to the curve is horizontal or vertical.

28) For the curve given by the equation $\sin(x + y) + \cos(x - y) = 1$, find the equation of the tangent line to the curve at the point $(\frac{\pi}{2}, \frac{\pi}{2})$.

2.8 *Derivatives of inverse functions*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- What is the derivative of the natural logarithm function?
- What are the derivatives of the inverse trigonometric functions $\arcsin(x)$ and $\arctan(x)$?
- If g is the inverse of a differentiable function f, how is g' computed in terms of f, f', and g?

Introduction

Much of mathematics centers on the notion of function. Indeed, throughout our study of calculus, we are investigating the behavior of functions, often doing so with particular emphasis on how fast the output of the function changes in response to changes in the input. Because each function represents a process, a natural question to ask is whether or not the particular process can be reversed. That is, if we know the output that results from the function, can we determine the input that led to it? Connected to this question, we now also ask: if we know how fast a particular process is changing, can we determine how fast the inverse process is changing?

As we have noted, one of the most important functions in all of mathematics is the natural exponential function $f(x) = e^x$. Because the natural logarithm, $g(x) = \ln(x)$, is the inverse of the natural exponential function, the natural logarithm is similarly important. One of our goals in this section is to learn how to differentiate the logarithm function, and thus expand our library of basic functions with known derivative formulas. First, we investigate a more familiar setting to refresh some of the basic concepts surrounding functions and their inverses.

Preview Activity 2.8

The equation $y = \frac{5}{9}(x-32)$ relates a temperature given in x degrees Fahrenheit to the corresponding temperature y measured in degrees Celcius.

(a) Solve the equation $y = \frac{5}{9}(x-32)$ for x to write x (Fahrenheit temperature) in terms of y (Celcius temperature).

(b) Let $C(x) = \frac{5}{9}(x-32)$ be the function that takes a Fahrenheit temperature as input and produces the Celcius temperature as output. In addition, let $F(y)$ be the function that converts a temperature given in y degrees Celcius to the temperature $F(y)$ measured in degrees Fahrenheit. Use your work in (a) to write a formula for $F(y)$.

(c) Next consider the new function defined by $p(x) = F(C(x))$. Use the formulas for F and C to determine an expression for $p(x)$ and simplify this expression as much as possible. What do you observe?

(d) Now, let $r(y) = C(F(y))$. Use the formulas for F and C to determine an expression for $r(y)$ and simplify this expression as much as possible. What do you observe?

(e) What is the value of $C'(x)$? of $F'(y)$? How do these values appear to be related?

Basic facts about inverse functions

A function $f : A \to B$ is a rule that associates each element in the set A to one and only one element in the set B. We call A the *domain* of f and B the *codomain* of f. If there exists a function $g : B \to A$ such that $g(f(a)) = a$ for every possible choice of a in the set A and $f(g(b)) = b$ for every b in the set B, then we say that g is the *inverse* of f. We often use the notation f^{-1} (read "f-inverse") to denote the inverse of f. Perhaps the most essential thing to observe about the inverse function is that it undoes the work of f. Indeed, if $y = f(x)$, then

$$f^{-1}(y) = f^{-1}(f(x)) = x,$$

and this leads us to another key observation: writing $y = f(x)$ and $x = f^{-1}(y)$ say the exact same thing. The only difference between the two equations is one of perspective – one is solved for x, while the other is solved for y.

Here we briefly remind ourselves of some key facts about inverse functions. For a function $f : A \to B$,

- f has an inverse if and only if f is one-to-one[8] and onto[9];
- provided f^{-1} exists, the domain of f^{-1} is the codomain of f, and the codomain of f^{-1} is the domain of f;
- $f^{-1}(f(x)) = x$ for every x in the domain of f and $f(f^{-1}(y)) = y$ for every y in the codomain of f;
- $y = f(x)$ if and only if $x = f^{-1}(y)$.

The last stated fact reveals a special relationship between the graphs of f and f^{-1}. In particular, if we consider $y = f(x)$ and a point (x, y) that lies on the graph of f, then it is also true that $x = f^{-1}(y)$, which means that the point (y, x) lies on the graph of f^{-1}. This shows us that the graphs of f and f^{-1} are the reflections of one another across the line $y = x$, since reflecting across $y = x$ is precisely the geometric action that swaps the coordinates in an ordered pair. In Figure 2.38, we see this exemplified for the function $y = f(x) = 2^x$ and its inverse, with the points $(-1, \frac{1}{2})$ and $(\frac{1}{2}, -1)$ highlighting the reflection of the curves across $y = x$.

[8] A function f is *one-to-one* provided that no two distinct inputs lead to the same output.

[9] A function f is *onto* provided that every possible element of the codomain can be realized as an output of the function for some choice of input from the domain.

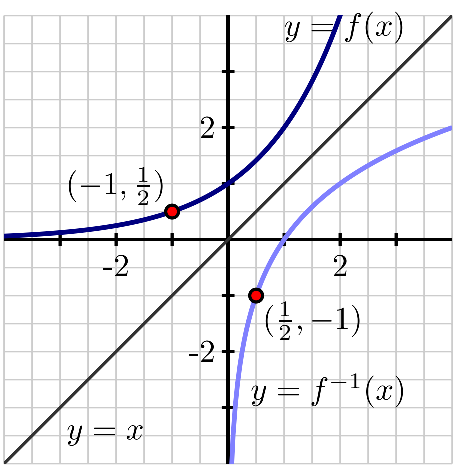


Figure 2.38: A graph of a function $y = f(x)$ along with its inverse, $y = f^{-1}(x)$.

To close our review of important facts about inverses, we recall that the natural exponential function $y = f(x) = e^x$ has an inverse function, and its inverse is the natural logarithm, $x = f^{-1}(y) = \ln(y)$. Indeed, writing $y = e^x$ is interchangeable with $x = \ln(y)$, plus $\ln(e^x) = x$ for every real number x and $e^{\ln(y)} = y$ for every positive real number y.

The derivative of the natural logarithm function

In what follows, we determine a formula for the derivative of $g(x) = \ln(x)$. To do so, we take advantage of the fact that we know the derivative of the natural exponential function, which is the inverse of g. In particular, we know that writing $g(x) = \ln(x)$ is equivalent to writing $e^{g(x)} = x$. Now we differentiate both sides of this most recent equation. In particular, we observe that

$$\frac{d}{dx}\left[e^{g(x)}\right] = \frac{d}{dx}[x].$$

The righthand side is simply 1; applying the chain rule to the left side, we find that

$$e^{g(x)}g'(x) = 1.$$

Since our goal is to determine $g'(x)$, we solve for $g'(x)$, so

$$g'(x) = \frac{1}{e^{g(x)}}.$$

Finally, we recall that since $g(x) = \ln(x)$, $e^{g(x)} = e^{\ln(x)} = x$, and thus

$$g'(x) = \frac{1}{x}.$$

Natural Logarithm

For all positive real numbers x,

$$\frac{d}{dx}[\ln(x)] = \frac{1}{x}.$$

This rule for the natural logarithm function now joins our list of other basic derivative rules that we have already established. There are two particularly interesting things to note about the fact that $\frac{d}{dx}[\ln(x)] = \frac{1}{x}$. One is that this rule is restricted to only apply to positive values of x, as these are the only values for which the original function is defined. The other is that for the first time in our work, differentiating a basic function of a particular type has led to a function of a very different nature: the derivative of the natural logarithm is not another logarithm, nor even an exponential function, but rather a rational one.

Derivatives of logarithms may now be computed in concert with all of the rules known to date. For instance, if $f(t) = \ln(t^2+1)$, then by the chain rule, $f'(t) = \frac{1}{t^2+1} \cdot 2t$.

Example 1

Find the derivatives of the following functions.

1. $y = \ln(4x^3 - 2x^2)$
2. $g(x) = x \ln x - x$.

Solution.

1. Recognize that $y = \ln(4x^3 - 2x^2)$ is the composition of $f(x) = \ln x$ and $g(x) = 4x^3 - 2x^2$. Also, recall that

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

This leads us to:

$$y' = \frac{1}{4x^3-2x^2} \cdot (12x^2 - 4x) = \frac{12x^2-4x}{4x^3-2x^2} = \frac{4x(3x-1)}{2x(2x^2-x)} = \frac{2(3x-1)}{2x^2-x}.$$

2. We use the Product Rule to find

$$\frac{d}{dx}\Big(x \ln x\Big) = x \cdot 1/x + 1 \cdot \ln x = 1 + \ln x$$

Using this result, we compute

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

Activity 2.8–1

For each function given below, find its derivative.

(a) $h(x) = x^2 \ln(x)$

(b) $p(t) = \dfrac{\ln(t)}{e^t + 1}$

(c) $s(y) = \ln(\cos(y) + 2)$

(d) $z(x) = \tan(\ln(x))$

(e) $m(z) = \ln(\ln(z))$

In addition to the important rule we have derived for the derivative of the natural log functions, there are additional interesting connections to note between the graphs of $f(x) = e^x$ and $f^{-1}(x) = \ln(x)$.

In Figure 2.39, we are reminded that since the natural exponential function has the property that its derivative is itself, the slope of the tangent to $y = e^x$ is equal to the height of the curve at that point. For instance, at the point $A = (\ln(0.5), 0.5)$, the slope of the tangent line is $m_A = 0.5$, and at $B = (\ln(5), 5)$, the tangent line's slope is $m_B = 5$. At the corresponding points A' and B' on the graph of the natural logarithm function (which come from reflecting across the line $y = x$), we know that the slope of the tangent line is the reciprocal of the x-coordinate of the point (since $\frac{d}{dx}[\ln(x)] = \frac{1}{x}$). Thus, with $A' = (0.5, \ln(0.5))$, we have $m_{A'} = \frac{1}{0.5} = 2$, and at $B' = (5, \ln(5))$, $m_{B'} = \frac{1}{5}$.

In particular, we observe that $m_{A'} = \frac{1}{m_A}$ and $m_{B'} = \frac{1}{m_B}$. This is not a coincidence, but in fact holds for any curve $y = f(x)$ and its inverse, provided the inverse exists. One rationale for why this is the case is due to the reflection across $y = x$: in so doing, we essentially change the roles of x and y, thus reversing the rise and run, which leads to the slope of the inverse function at the reflected point being the reciprocal of the slope of the original function. At the close of this section, we will also look at how the chain rule provides us with an algebraic formulation of this general phenomenon.

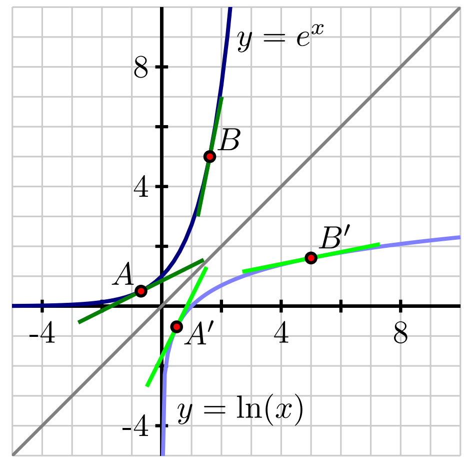


Figure 2.39: A graph of the function $y = e^x$ along with its inverse, $y = \ln(x)$, where both functions are viewed using the input variable x.

Inverse trigonometric functions and their derivatives

Trigonometric functions are periodic, so they fail to be one-to-one, and thus do not have inverses. However, if we restrict the domain of each trigonometric function, we can force the function to be one-to-one. For instance, consider the sine function on the domain $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

Because no output of the sine function is repeated on this interval, the function is one-to-one and thus has an inverse. In particular, if we view $f(x) = \sin(x)$ as having domain $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and codomain $[-1, 1]$, then there exists an inverse function f^{-1} such that

$$f^{-1} : [-1, 1] \to [-\frac{\pi}{2}, \frac{\pi}{2}].$$

We call f^{-1} the *arcsine* (or inverse sine) function and write $f^{-1}(y) = \arcsin(y)$. It is especially important to remember that writing

$$y = \sin(x) \ \text{ and } \ x = \arcsin(y)$$

say the exact same thing. We often read "the arcsine of y" as "the angle whose sine is y." For example, we say that $\frac{\pi}{6}$ is the angle whose sine is $\frac{1}{2}$, which can be written more concisely as $\arcsin(\frac{1}{2}) = \frac{\pi}{6}$, which is equivalent to writing $\sin(\frac{\pi}{6}) = \frac{1}{2}$.

Next, we determine the derivative of the arcsine function. Letting $h(x) = \arcsin(x)$, our goal is to find $h'(x)$. Since $h(x)$ is the

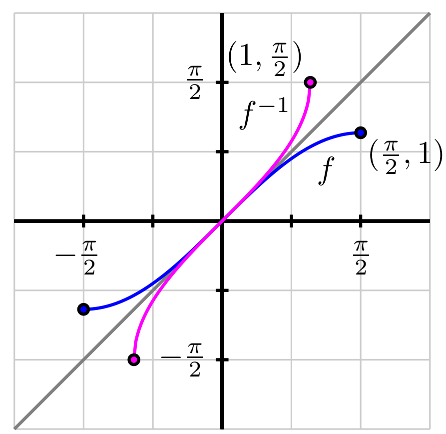


Figure 2.40: A graph of $f(x) = \sin(x)$ (in blue), restricted to the domain $[-\frac{\pi}{2}, \frac{\pi}{2}]$, along with its inverse, $f^{-1}(x) = \arcsin(x)$ (in magenta).

angle whose sine is x, it is equivalent to write

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

Differentiating both sides of the previous equation, we have

$$\frac{d}{dx}[\sin(h(x))] = \frac{d}{dx}[x],$$

and by the fact that the righthand side is simply 1 and by the chain rule applied to the left side,

$$\cos(h(x))h'(x) = 1.$$

Solving for $h'(x)$, it follows that

$$h'(x) = \frac{1}{\cos(h(x))}.$$

Finally, we recall that $h(x) = \arcsin(x)$, so the denominator of $h'(x)$ is the function $\cos(\arcsin(x))$, or in other words, "the cosine of the angle whose sine is x." A bit of right triangle trigonometry allows us to simplify this expression considerably.

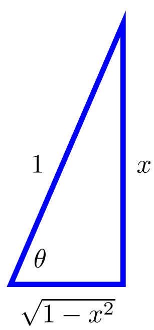


Figure 2.41: The right triangle that corresponds to the angle $\theta = \arcsin(x)$.

Let's say that $\theta = \arcsin(x)$, so that θ is the angle whose sine is x. From this, it follows that we can picture θ as an angle in a right triangle with hypotenuse 1 and a vertical leg of length x, as shown in Figure 2.41. The horizontal leg must be $\sqrt{1-x^2}$, by the Pythagorean Theorem. Now, note particularly that $\theta = \arcsin(x)$ since $\sin(\theta) = x$, and recall that we want to know a different expression for $\cos(\arcsin(x))$. From the figure, $\cos(\arcsin(x)) = \cos(\theta) = \sqrt{1-x^2}$.

Thus, returning to our earlier work where we established that if $h(x) = \arcsin(x)$, then $h'(x) = \frac{1}{\cos(\arcsin(x))}$, we have now shown that

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

Derivative of Inverse Sine

For all real numbers x such that $-1 < x < 1$,

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

Function	Domain	Range	Inverse Function	Domain	Range
$\sin x$	$[-\pi/2, \pi/2]$	$[-1, 1]$	$\sin^{-1} x$	$[-1, 1]$	$[-\pi/2, \pi/2]$
$\cos x$	$[0, \pi]$	$[-1, 1]$	$\cos^{-1}(x)$	$[-1, 1]$	$[0, \pi]$
$\tan x$	$(-\pi/2, \pi/2)$	$(-\infty, \infty)$	$\tan^{-1}(x)$	$(-\infty, \infty)$	$(-\pi/2, \pi/2)$
$\csc x$	$[-\pi/2, 0) \cup (0, \pi/2]$	$(-\infty, -1] \cup [1, \infty)$	$\csc^{-1} x$	$(-\infty, -1] \cup [1, \infty)$	$[-\pi/2, 0) \cup (0, \pi/2]$
$\sec x$	$[0, \pi/2) \cup (\pi/2, \pi]$	$(-\infty, -1] \cup [1, \infty)$	$\sec^{-1}(x)$	$(-\infty, -1] \cup [1, \infty)$	$[0, \pi/2) \cup (\pi/2, \pi]$
$\cot x$	$(0, \pi)$	$(-\infty, \infty)$	$\cot^{-1}(x)$	$(-\infty, \infty)$	$(0, \pi)$

Figure 2.42: Domains and ranges of the trigonometric and inverse trigonometric functions.

Using similar techniques, we can find the derivatives of all the inverse trigonometric functions. In Figure 2.42 we show the restrictions of the domains of the standard trigonometric functions that allow them to be invertible.

Activity 2.8–2

The following prompts in this activity will lead you to develop the derivative of the inverse tangent function.

(a) Let $r(x) = \arctan(x)$. Use the relationship between the arctangent and tangent functions to rewrite this equation using only the tangent function.

(b) Differentiate both sides of the equation you found in (a). Solve the resulting equation for $r'(x)$, writing $r'(x)$ as simply as possible in terms of a trigonometric function evaluated at $r(x)$.

(c) Recall that $r(x) = \arctan(x)$. Update your expression for $r'(x)$ so that it only involves trigonometric functions and the independent variable x.

(d) Introduce a right triangle with angle θ so that $\theta = \arctan(x)$. What are the three sides of the triangle?

(e) In terms of only x and 1, what is the value of $\cos(\arctan(x))$?

(f) Use the results of your work above to find an expression involving only 1 and x for $r'(x)$.

In Activity 2.8–2, we will developed the derivative of the inverse tangent function. The derivatives for the other inverse trigonometric functions can be established similarly.

Derivatives of Inverse Trigonometric Functions

The inverse trigonometric functions are differentiable on their domains (as listed in Figure 2.42) and their derivatives are as follows:

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

2) $\frac{d}{dx}\big(\operatorname{arcsec}(x)\big) = \frac{1}{|x|\sqrt{x^2-1}}$

3) $\frac{d}{dx}\big(\arctan(x)\big) = \frac{1}{1+x^2}$

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

5) $\frac{d}{dx}\big(\operatorname{arccsc}(x)\big) = -\frac{1}{|x|\sqrt{x^2-1}}$

6) $\frac{d}{dx}\big(\operatorname{arccot}(x)\big) = -\frac{1}{1+x^2}$

Note how the last three derivatives are merely the opposites of the first three, respectively. Because of this, the first three are used almost exclusively throughout this text.

With these rules added to our library of derivatives of basic functions, we can differentiate even more functions using derivative shortcuts. In Activity 2.8–3, we see each of these rules at work.

Activity 2.8–3

Determine the derivative of each of the following functions.

(a) $f(x) = x^3 \arctan(x) + e^x \ln(x)$

(b) $p(t) = 2^{t \arcsin(t)}$

(c) $h(z) = (\arcsin(5z) + \arctan(4-z))^{27}$

(d) $s(y) = \cot(\arctan(y))$

(e) $m(v) = \ln(\sin^2(v) + 1)$

(f) $g(w) = \arctan\left(\frac{\ln(w)}{1+w^2}\right)$

The link between the derivative of a function and the derivative of its inverse

In Figure 2.39, we saw an interesting relationship between the slopes of tangent lines to the natural exponential and natural logarithm functions at points that corresponded to reflection across the line $y = x$. In particular, we observed that for a point such as $(\ln(2), 2)$ on the graph of $f(x) = e^x$, the slope of the tangent line at this point is $f'(\ln(2)) = 2$, while at the corresponding point $(2, \ln(2))$ on the graph of $f^{-1}(x) = \ln(x)$, the slope of the tangent line at this point is $(f^{-1})'(2) = \frac{1}{2}$, which is the reciprocal of $f'(\ln(2))$.

That the two corresponding tangent lines having slopes that are reciprocals of one another is not a coincidence. If we consider the general setting of a differentiable function f with differentiable inverse g such that $y = f(x)$ if and only if $x = g(y)$, then we know that $f(g(x)) = x$ for every x in the domain of f^{-1}. Differentiating both sides of this equation with respect to x, we have

$$\frac{d}{dx}[f(g(x))] = \frac{d}{dx}[x],$$

and by the chain rule,

$$f'(g(x))g'(x) = 1.$$

Solving for $g'(x)$, we have

$$g'(x) = \frac{1}{f'(g(x))}.$$

Here we see that the slope of the tangent line to the inverse function g at the point $(x, g(x))$ is precisely the reciprocal of the slope of the tangent line to the original function f at the point $(g(x), f(g(x))) = (g(x), x)$.

To see this more clearly, consider the graph of the function $y = f(x)$ shown in Figure 2.43, along with its inverse $y = g(x)$. Given a point (a, b) that lies on the graph of f, we know that (b, a) lies on the graph of g; said differently, $f(a) = b$ and $g(b) = a$. Now, applying the rule that

$$g'(x) = \frac{1}{f'(g(x))}$$

to the value $x = b$, we have

$$g'(b) = \frac{1}{f'(g(b))} = \frac{1}{f'(a)},$$

which is precisely what we see in the figure: the slope of the tangent line to g at (b, a) is the reciprocal of the slope of the tangent line to f at (a, b), since these two lines are reflections of one another across the line $y = x$.

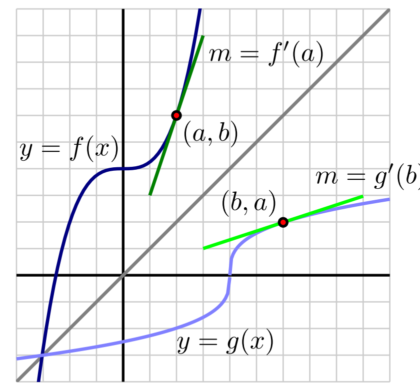


Figure 2.43: A graph of function $y = f(x)$ along with its inverse, $y = g(x) = f^{-1}(x)$. Observe that the slopes of the two tangent lines are reciprocals of one another.

Derivative of an inverse function

Suppose that f is a differentiable function with inverse g and that (a, b) is a point that lies on the graph of f at which $f'(a) \neq 0$. Then

$$g'(b) = \frac{1}{f'(a)}.$$

More generally, for any x in the domain of g', we have

$$g'(x) = \frac{1}{f'(g(x))}.$$

The rules we derived for $\ln(x)$, $\arcsin(x)$, and $\arctan(x)$ are all just specific examples of this general property of the derivative of an inverse function. For example, with $g(x) = \ln(x)$ and $f(x) = e^x$, it follows that

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{e^{\ln(x)}} = \frac{1}{x}.$$

Example 2

Let $f(x) = \tan\left(x^3 + x + \frac{\pi}{4}\right)$ with domain restricted so that f is one-to-one. Find $(f^{-1})'(1)$.

Solution. We begin by finding the derivative of f using the chain rule.

$$f'(x) = \sec^2\left(x^3 + x + \frac{\pi}{4}\right) \cdot (3x^2+1) = (3x^2+1)\sec^2\left(x^3 + x + \frac{\pi}{4}\right)$$

Next, we apply the derivative of an inverse rule.

$$(f^{-1})'(b) = \frac{1}{f'(f^{-1}(b))} = \frac{1}{(3a^2+1)\sec^2\left(a^3 + a + \frac{\pi}{4}\right)}$$

Notice $f(0) = 1$. Therefore, $f^{-1}(1) = 0$. Finally we have the result

$$(f^{-1})'(1) = \frac{1}{(3(0)^2+1)\sec^2\left(0^3 + 0 + \frac{\pi}{4}\right)} = \frac{1}{2}$$

Logarithmic Differentiation

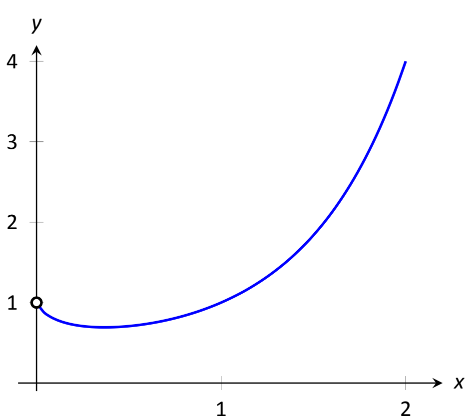


Figure 2.44: A plot of $y = x^x$.

Consider the function $y = x^x$; it is graphed in Figure 2.44. It is well–defined for $x > 0$ and we might be interested in finding equations of lines tangent and normal to its graph. How do we take its derivative?

The function is not a power function: it has a "power" of x, not a constant. It is not an exponential function: it has a "base" of x, not a constant.

A differentiation technique known as *logarithmic differentiation* becomes useful here. The basic principle is this: take the natural log of both sides of an equation $y = f(x)$, then use implicit differentiation to find y'. We demonstrate this in the following example.

Example 3

Given $y = x^x$, use logarithmic differentiation to find y'.

Solution. As suggested above, we start by taking the natural log of both sides then applying implicit differentiation.

$$\begin{aligned} y &= x^x \\ \ln(y) &= \ln(x^x) && \text{(apply logarithm rule)} \\ \ln(y) &= x\ln(x) && \text{(now use implicit differentiation)} \\ \frac{d}{dx}\Big(\ln(y)\Big) &= \frac{d}{dx}\Big(x\ln(x)\Big) \\ \frac{y'}{y} &= \ln(x) + x\cdot\frac{1}{x} \\ \frac{y'}{y} &= \ln(x) + 1 \\ y' &= y\big(\ln(x)+1\big) \text{(now substitute } y = x^x) \\ y' &= x^x\big(\ln(x)+1\big). \end{aligned}$$

To "test" our answer, let's use it to find the equation of the tangent line at $x = 1.5$. The point on the graph our tangent line must pass through is $(1.5, 1.5^{1.5}) \approx (1.5, 1.837)$. Using the equation for y', we find the slope as

$$y' = 1.5^{1.5}\big(\ln(1.5)+1\big) \approx 1.837(1.405) \approx 2.582.$$

Thus the equation of the tangent line is $y = 1.6833(x - 1.5) + 1.837$. Figure 2.45 graphs $y = x^x$ along with this tangent line.

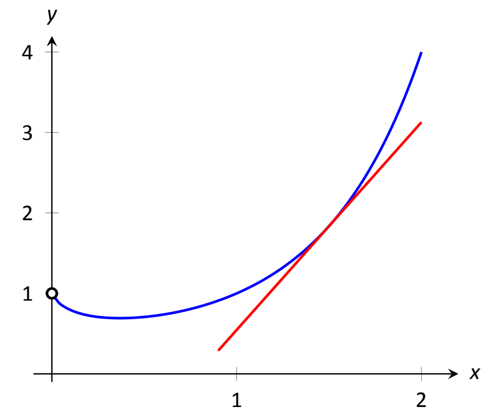


Figure 2.45: A plot of $y = x^x$.

However, is it possible, or even wise, to use Logarithmic Differentation at other times? Consider the function

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

If we used logarithmic differentiation, then it would certainly be possible to make the function much simpler to work with if we broke it apart using the laws of logarithms. Let's try it!

Example 4

Given $y = \dfrac{\sqrt{x}(x^2+1)^3}{\sin(x)}$, use logarithmic differentiation to find y'.

Solution. As before, we start by taking the natural log of both sides then applying implicit differentiation.

$$\begin{aligned}
y &= \frac{\sqrt{x}(x^2+1)^3}{\sin(x)} \\
\ln y &= \ln\left(\frac{\sqrt{x}(x^2+1)^3}{\sin(x)}\right) \quad \text{(apply logarithm rule)} \\
\ln(y) &= \ln\left(\sqrt{x}(x^2+1)^3\right) - \ln(\sin(x)) \\
\ln(y) &= \ln\sqrt{x} + \ln(x^2+1)^3 - \ln(\sin(x)) \\
\ln(y) &= \frac{1}{2}\ln(x) + 3\ln(x^2+1) - \ln(\sin(x)) \\
\frac{d}{dx}\Big(\ln(y)\Big) &= \frac{d}{dx}\left(\frac{1}{2}\ln(x) + 3\ln(x^2+1) - \ln(\sin(x))\right) \\
\frac{y'}{y} &= \frac{1}{2}\cdot\frac{1}{x} + 3\cdot\frac{1}{x^2+1}\cdot 2x - \frac{1}{\sin(x)}\cdot\cos(x) \\
\frac{y'}{y} &= \frac{1}{2x} + \frac{6x}{x^2+1} - \cot(x) \\
y' &= y\left(\frac{1}{2x} + \frac{6x}{x^2+1} - \cot(x)\right) \\
y' &= \frac{\sqrt{x}(x^2+1)^3}{\sin(x)}\left(\frac{1}{2x} + \frac{6x}{x^2+1} - \cot(x)\right).
\end{aligned}$$

So it seems that we can use logarithmic differentiation on functions that aren't of the form $f(x)^{g(x)}$ to break apart a complex function into one of many pieces whereby differentiating each piece is easier.

Proof of the Power Rule

Now that we have inverse relationships and the derivatives of logarithmic and exponential functions, we can finally prove the Power Rule for Differentiation.

Power Rule

For any nonzero real number, if $f(x) = x^n$, then

$$f'(x) = nx^{n-1}.$$

Proof: By the inverse relationships, we can write

$$\frac{d}{dx}[x^n] \quad = \quad \frac{d}{dx}[e^{\ln(x^n)}] \quad = \quad \frac{d}{dx}[e^{n\cdot\ln(x)}],$$

which we can differentiate as

$$\frac{d}{dx}[e^{n\cdot\ln(x)}] = e^{n\cdot\ln(x)}\cdot n\cdot\frac{1}{x}.$$

We can use the inverse relationships to back-substitute

$$\frac{dy}{dx} = x^n \cdot n \cdot \frac{1}{x} \Rightarrow \frac{dy}{dx} = n \cdot \frac{x^n}{x} = n \cdot x^{n-1}.$$

We conclude this section and chapter by restating many of the rules we have used to find derivatives throughout this chapter. The following is intended to be a reference for future work.

Glossary of Derivatives of Elementary Functions

Let u and v be differentiable functions, and let a, c and n be a real numbers, $a > 0$, $n \neq 0$.

1) $\frac{d}{dx}\big(cu\big) = cu'$
2) $\frac{d}{dx}\big(u \cdot v\big) = uv' + u'v$
3) $\frac{d}{dx}\big(u(v)\big) = u'(v)v'$
4) $\frac{d}{dx}\big(x\big) = 1$
5) $\frac{d}{dx}\big(e^x\big) = e^x$
6) $\frac{d}{dx}\big(\ln(x)\big) = \frac{1}{x}$
7) $\frac{d}{dx}\big(\sin(x)\big) = \cos(x)$
8) $\frac{d}{dx}\big(\csc(x)\big) = -\csc(x)\cot(x)$
9) $\frac{d}{dx}\big(\tan(x)\big) = \sec^2(x)$
10) $\frac{d}{dx}\big(\arcsin(x)\big) = \frac{1}{\sqrt{1-x^2}}$
11) $\frac{d}{dx}\big(\operatorname{arccsc}(x)\big) = -\frac{1}{|x|\sqrt{x^2-1}}$
12) $\frac{d}{dx}\big(\arctan(x)\big) = \frac{1}{1+x^2}$
13) $\frac{d}{dx}\big(u \pm v\big) = u' \pm v'$
14) $\frac{d}{dx}\big(\frac{u}{v}\big) = \frac{u'v-uv'}{v^2}$
15) $\frac{d}{dx}\big(c\big) = 0$
16) $\frac{d}{dx}\big(x^n\big) = nx^{n-1}$
17) $\frac{d}{dx}\big(a^x\big) = \ln(a) \cdot a^x$
18) $\frac{d}{dx}\big(\log_a(x)\big) = \frac{1}{\ln(a)} \cdot \frac{1}{x}$
19) $\frac{d}{dx}\big(\cos(x)\big) = -\sin(x)$
20) $\frac{d}{dx}\big(\sec(x)\big) = \sec(x)\tan(x)$
21) $\frac{d}{dx}\big(\cot(x)\big) = -\csc^2(x)$
22) $\frac{d}{dx}\big(\arccos(x)\big) = -\frac{1}{\sqrt{1-x^2}}$
23) $\frac{d}{dx}\big(\operatorname{arcsec}(x)\big) = \frac{1}{|x|\sqrt{x^2-1}}$
24) $\frac{d}{dx}\big(\operatorname{arccot}(x)\big) = -\frac{1}{1+x^2}$

Summary

In this section, we encountered the following important ideas:

- For all positive real numbers x, $\frac{d}{dx}[\ln(x)] = \frac{1}{x}$.
- For all real numbers x such that $-1 \leq x \leq 1$, $\frac{d}{dx}[\arcsin(x)] = \frac{1}{\sqrt{1-x^2}}$. In addition, for all real numbers x, $\frac{d}{dx}[\arctan(x)] = \frac{1}{1+x^2}$.
- If g is the inverse of a differentiable function f, then for any point x in the domain of g', $g'(x) = \frac{1}{f'(g(x))}$.

Exercises

Terms and Concepts

1) T/F: Every function has an inverse.

2) In your own words explain what it means for a function to be "one to one."

3) If $(1,10)$ lies on the graph of $y=f(x)$, what can be said about the graph of $y=f^{-1}(x)$?

4) If $(1,10)$ lies on the graph of $y=f(x)$ and $f'(1)=5$, what can be said about $y=f^{-1}(x)$?

Problems

In exercises 5–8, verify that the given functions are inverses.

5) $f(x)=2x+6$ and $g(x)=\frac{1}{2}x-3$

6) $f(x)=x^2+6x+11$, $x\geq 3$ and
$g(x)=\sqrt{x-2}-3$, $x\geq 2$

7) $f(x)=\dfrac{3}{x-5}$, $x\neq 5$ and
$g(x)=\dfrac{3+5x}{x}$, $x\neq 0$

8) $f(x)=\dfrac{x+1}{x-1}$, $x\neq 1$ and $g(x)=f(x)$

In exercises 9–14, an invertible function $f(x)$ is given along with a point that lies on its graph. Evaluate $\left(f^{-1}\right)'(x)$ at the indicated value.

9) $f(x)=5x+10$
Point= $(2,20)$
Evaluate $\left(f^{-1}\right)'(20)$

10) $f(x)=x^2-2x+4$, $x\geq 1$
Point= $(3,7)$
Evaluate $\left(f^{-1}\right)'(7)$

11) $f(x)=\sin 2x$, $-\pi/4\leq x\leq\pi/4$
Point= $(\pi/6,\sqrt{3}/2)$
Evaluate $\left(f^{-1}\right)'(\sqrt{3}/2)$

12) $f(x)=x^3-6x^2+15x-2$
Point= $(1,8)$
Evaluate $\left(f^{-1}\right)'(8)$

13) $f(x)=\dfrac{1}{1+x^2}$, $x\geq 0$
Point= $(1,1/2)$
Evaluate $\left(f^{-1}\right)'(1/2)$

14) $f(x)=6e^{3x}$
Point= $(0,6)$
Evaluate $\left(f^{-1}\right)'(6)$

In exercises 15–24, compute the derivative of the given function.

15) $h(t)=\arcsin(2t)$

16) $f(t)=\operatorname{arcsec}(2t)$

17) $g(x)=\arctan(2x)$

18) $f(x)=x\arcsin(x)$

19) $g(t)=\sin(t)\arccos(t)$

20) $f(t)=\ln(te^t)$

21) $h(x)=\dfrac{\arcsin(x)}{\arccos(x)}$

22) $g(x)=\arctan(\sqrt{x})$

23) $f(x)=\operatorname{arcsec}(1/x)$

24) $f(x)=\sin(\arcsin(x))$

25) $f(x)=\ln(2\arctan(x)+3\arcsin(x)+5)$

26) $r(z)=\arctan(\ln(\arcsin(z)))$

27) $q(t)=\arctan^2(3t)\arcsin^4(7t)$

28) $g(v)=\ln\left(\dfrac{\arctan(v)}{\arcsin(v)+v^2}\right)$

In exercises 29–34, use logarithmic differentiation to find $\frac{dy}{dx}$, then find the equation of the tangent line at the indicated x-value.

29) $y=(1+x)^{1/x},\quad x=1$

30) $y=(2x)^{x^2},\quad x=1$

31) $y=\dfrac{x^x}{x+1},\quad x=1$

32) $y=x^{\sin(x)+2},\quad x=\pi/2$

33) $y=\dfrac{x+1}{x+2},\quad x=1$

34) $y=\dfrac{(x+1)(x+2)}{(x+3)(x+4)},\quad x=0$

In exercises 25–24, find the equation of the tangent line to the graph of the function at the indicated value.

35) $f(x)=\arcsin(x)$ at $x=\frac{\sqrt{2}}{2}$

36) $f(x)=\arccos(2x)$ at $x=\frac{\sqrt{3}}{4}$

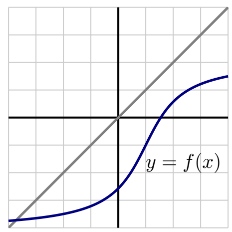
37) Let $f(x)=\frac{1}{4}x^3+4$.

(a) Sketch a graph of $y=f(x)$ and explain why f is an invertible function.

(b) Let g be the inverse of f and determine a formula for g.

(c) Compute $f'(x)$, $g'(x)$, $f'(2)$, and $g'(6)$. What is the special relationship between $f'(2)$ and $g'(6)$? Why?

38) Consider the graph of $y = f(x)$ provided below and use it to answer the following questions.



(a) Use the provided graph to estimate the value of $f'(1)$.

(b) Sketch an approximate graph of $y = f^{-1}(x)$. Label at least three distinct points on the graph that correspond to three points on the graph of f.

(c) Based on your work in (a), what is the value of $(f^{-1})'(-1)$? Why?

39) Let $h(x) = x + \sin(x)$.

(a) Sketch a graph of $y = h(x)$ and explain why h must be invertible.

(b) Explain why it does not appear to be algebraically possible to determine a formula for h^{-1}.

(c) Observe that the point $(\frac{\pi}{2}, \frac{\pi}{2} + 1)$ lies on the graph of $y = h(x)$. Determine the value of $(h^{-1})'(\frac{\pi}{2} + 1)$.

Chapter 3
Applications of Derivatives

3.1 Related rates

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- If two quantities that are related, such as the radius and volume of a spherical balloon, are both changing as implicit functions of time, how are their rates of change related? That is, how does the relationship between the values of the quantities affect the relationship between their respective derivatives with respect to time?

Introduction

In most of our study of derivatives so far, we have worked in settings where one quantity (often called y) depends explicitly on another (say x), and in some way we have been interested in the instantaneous rate at which y changes with respect to x, leading us to compute $\frac{dy}{dx}$. These settings emphasize how the derivative enables us to quantify how the quantity y is changing as x changes at a given x-value.

We are next going to consider situations where multiple quantities are related to one another and changing, but where each quantity can be considered an implicit function of the variable t, which represents time. Through knowing how the quantities are related, we will be interested in determining how their respective rates of change with respect to time are related. For example, suppose that air is being pumped into a spherical balloon in such a way that its volume increases at a constant rate of 20 cubic inches per second. It makes sense that since the balloon's volume and radius are related, by knowing how fast the volume is changing, we ought to be able to relate this rate to how fast the radius is changing. More specifically, can we find how fast the radius of the balloon is increasing at the moment the balloon's diameter is 12 inches?

The following preview activity leads you through the steps to answer this question.

Preview Activity 3.1

A spherical balloon is being inflated at a constant rate of 20 cubic inches per second. How fast is the radius of the balloon changing at the instant the balloon's diameter is 12 inches? Is the radius changing more rapidly when $d = 12$ or when $d = 16$? Why?

(a) Draw several spheres with different radii, and observe that as volume changes, the radius, diameter, and surface area of the balloon also change.

(b) Recall that the volume of a sphere of radius r is $V = \frac{4}{3}\pi r^3$. Note well that in the setting of this problem, *both* V and r are changing as time t changes, and thus both V and r may be viewed as implicit functions of t, with respective derivatives $\frac{dV}{dt}$ and $\frac{dr}{dt}$.
Differentiate both sides of the equation $V = \frac{4}{3}\pi r^3$ with respect to t (using the chain rule on the right) to find a formula for $\frac{dV}{dt}$ that depends on both r and $\frac{dr}{dt}$.

(c) At this point in the problem, by differentiating we have "related the rates" of change of V and r. Recall that we are given in the problem that the balloon is being inflated at a constant *rate* of 20 cubic inches per second. Is this rate the value of $\frac{dr}{dt}$ or $\frac{dV}{dt}$? Why?

(d) From part (c), we know the value of $\frac{dV}{dt}$ at every value of t. Next, observe that when the diameter of the balloon is 12, we know the value of the radius. In the equation $\frac{dV}{dt} = 4\pi r^2 \frac{dr}{dt}$, substitute these values for the relevant quantities and solve for the remaining unknown quantity, which is $\frac{dr}{dt}$. How fast is the radius changing at the instant $d = 12$?

(e) How is the situation different when $d = 16$? When is the radius changing more rapidly, when $d = 12$ or when $d = 16$?

Related Rates Problems

In problems where two or more quantities can be related to one another, and all of the variables involved can be viewed as implicit functions of time, t, we are often interested in how the rates of change of the individual quantities with respect to time are themselves related; we call these *related rates* problems. Often these problems involve identifying one or more key underlying geometric relationships to relate the variables involved. Once we have an equation establishing the fundamental relationship among variables, we differentiate implicitly with respect to time to find connections among the rates of change.

For example, consider the situation where sand is being dumped by a conveyor belt on a pile so that the sand forms a right circular cone, as pictured in Figure 3.1.

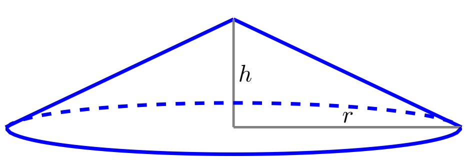


Figure 3.1: A conical pile of sand.

As sand falls from the conveyor belt onto the top of the pile, obviously several features of the sand pile will change: the volume of the pile will grow, the height will increase, and the radius will get bigger, too. All of these quantities are related to one another, and the rate at which each is changing is related to the rate at which sand falls from the conveyor.

The first key steps in any related rates problem involve identifying which variables are changing and how they are related. In the current problem involving a conical pile of sand, we observe that the radius and height of the pile are related to the volume of the pile by the standard equation for the volume of a cone,

$$V = \frac{1}{3}\pi r^2 h.$$

Viewing each of V, r, and h as functions of t, we can differentiate implicitly to determine an equation that relates their respective rates of change. Taking the derivative of each side of the equation with respect to t,

$$\frac{d}{dt}[V] = \frac{d}{dt}\left[\frac{1}{3}\pi r^2 h\right].$$

On the left, $\frac{d}{dt}[V]$ is simply $\frac{dV}{dt}$. On the right, the situation is more complicated, as both r and h are implicit functions of t, hence we have to use the product and chain rules. Doing so, we find that

$$\begin{aligned}\frac{dV}{dt} &= \frac{d}{dt}\left[\frac{1}{3}\pi r^2 h\right] \\ &= \frac{1}{3}\pi r^2 \frac{d}{dt}[h] + \frac{1}{3}\pi h \frac{d}{dt}[r^2] \\ &= \frac{1}{3}\pi r^2 \frac{dh}{dt} + \frac{1}{3}\pi h 2r \frac{dr}{dt}\end{aligned}$$

Note particularly how we are using ideas from Section 2.7 on implicit differentiation. There we found that when y is an implicit function of x, $\frac{d}{dx}[y^2] = 2y\frac{dy}{dx}$. The exact same thing is occurring here when we compute $\frac{d}{dt}[r^2] = 2r\frac{dr}{dt}$.

With our arrival at the equation

$$\frac{dV}{dt} = \frac{1}{3}\pi r^2 \frac{dh}{dt} + \frac{2}{3}\pi r h \frac{dr}{dt},$$

we have now related the rates of change of V, h, and r. If we are given sufficient information, we may then find the value of one or more of these rates of change at one or more points in time. Say, for instance, that we know the following: (a) sand falls from the conveyor in such a way that the height of the pile is always half the radius, and (b) sand falls from the conveyor belt at a constant rate of 10 cubic feet per minute. With this information given, we can answer questions such as: how fast is the height of the sandpile changing at the moment the radius is 4 feet?

The information that the height is always half the radius tells us that for all values of t, $h = \frac{1}{2}r$. Differentiating with respect to t, it follows that $\frac{dh}{dt} = \frac{1}{2}\frac{dr}{dt}$. These relationships enable us to relate $\frac{dV}{dt}$ exclusively to just one of r or h. Substituting the expressions involving r and $\frac{dr}{dt}$ for h and $\frac{dh}{dt}$, we now have that

$$\frac{dV}{dt} = \frac{1}{3}\pi r^2 \cdot \frac{1}{2}\frac{dr}{dt} + \frac{2}{3}\pi r \cdot \frac{1}{2}r \cdot \frac{dr}{dt}. \tag{3.1}$$

Since sand falls from the conveyor at the constant rate of 10 cubic feet per minute, this tells us the value of $\frac{dV}{dt}$, the rate at which the volume of the sand pile changes. In particular, $\frac{dV}{dt} = 10$ ft^3/min. Furthermore, since we are interested in how fast the height of the pile is changing at the instant $r = 4$, we use the value $r = 4$ along with $\frac{dV}{dt} = 10$ in Equation (3.1), and hence find that

$$10 = \frac{1}{3}\pi 4^2 \cdot \frac{1}{2}\left.\frac{dr}{dt}\right|_{r=4} + \frac{2}{3}\pi 4 \cdot \frac{1}{2}4 \cdot \left.\frac{dr}{dt}\right|_{r=4} = \frac{8}{3}\pi \left.\frac{dr}{dt}\right|_{r=4} + \frac{16}{3}\pi \left.\frac{dr}{dt}\right|_{r=4}.$$

With only the value of $\left.\frac{dr}{dt}\right|_{r=4}$ remaining unknown, we solve for $\left.\frac{dr}{dt}\right|_{r=4}$ and find that $10 = 8\pi \left.\frac{dr}{dt}\right|_{r=4}$, so that

$$\left.\frac{dr}{dt}\right|_{r=4} = \frac{10}{8\pi} \approx 0.39789$$

feet per second. Because we were interested in how fast the height of the pile was changing at this instant, we want to know $\frac{dh}{dt}$ when $r = 4$. Since $\frac{dh}{dt} = \frac{1}{2}\frac{dr}{dt}$ for all values of t, it follows

$$\left.\frac{dh}{dt}\right|_{r=4} = \frac{5}{8\pi} \approx 0.19894 \text{ ft/min}.$$

Note particularly how we distinguish between the notations $\frac{dr}{dt}$ and $\left.\frac{dr}{dt}\right|_{r=4}$. The former represents the rate of change of r with respect to t at an arbitrary value of t, while the latter is the rate of change of r with respect to t at a particular moment, in fact the moment $r = 4$. While we don't know the exact value of t, because information is provided about the value of r, it is important to distinguish that we are using this more specific data.

The relationship between h and r, with $h = \frac{1}{2}r$ for all values of t, enables us to transition easily between questions involving r and h. Indeed, had we known this information at the problem's outset, we could have immediately simplified our work. Using $h = \frac{1}{2}r$, it follows that since $V = \frac{1}{3}\pi r^2 h$, we can write V solely in terms of r to have

$$V = \frac{1}{3}\pi r^2 \left(\frac{1}{2}h\right) = \frac{1}{6}\pi r^3.$$

From this last equation, differentiating with respect to t implies

$$\frac{dV}{dt} = \frac{1}{2}\pi r^2,$$

from which the same conclusions made earlier about $\frac{dr}{dt}$ and $\frac{dh}{dt}$ can be made.

Our work with the sandpile problem above is similar in many ways to our approach in Preview Activity 3.1, and these steps are typical of most related rates problems. In certain ways, they also resemble work we will do in Applied Optimization problems in Section **??**, and here we summarize the main approach for consideration in subsequent problems.

- Identify the quantities in the problem that are changing and choose clearly defined variable names for them. Draw one or more figures that clearly represent the situation.
- Determine all rates of change that are known or given and identify the rate(s) of change to be found.
- Find an equation that relates the variables whose rates of change are known to those variables whose rates of change are to be found.
- Differentiate implicitly with respect to t to relate the rates of change of the involved quantities.
- Evaluate the derivatives and variables at the information relevant to the instant at which a certain rate of change is sought. Use proper notation to identify when a derivative is being evaluated at a particular instant, such as $\left.\frac{dr}{dt}\right|_{r=4}$.

In the first step of identifying changing quantities and drawing a picture, it is important to think about the dynamic ways in which the involved quantities change. Sometimes a sequence of pictures can be helpful; for some already-drawn pictures that can be easily modified as applets built in Geogebra, see the following links[1] which represent

- how a circular oil slick's area grows as its radius increases: `http://gvsu.edu/s/9n`;

[1] We again refer to the work of Prof. Marc Renault of Shippensburg University, found at `http://gvsu.edu/s/5p`.

- how the location of the base of a ladder and its height along a wall change as the ladder slides: `http://gvsu.edu/s/9o`;
- how the water level changes in a conical tank as it fills with water at a constant rate `http://gvsu.edu/s/9p` (compare the problem in Activity 3.2–1); and
- how a skateboarder's shadow changes as he moves past a lamppost: `http://gvsu.edu/s/9q`.

Drawing well-labeled diagrams and envisioning how different parts of the figure change is a key part of understanding related rates problems and being successful at solving them.

Example 1

Water streams out of a faucet at a rate of 2 in^3/s onto a flat surface at a constant rate, forming a circular puddle that is 1/8 in deep.

1) At what rate is the area of the puddle growing?
2) At what rate is the radius of the circle growing?

Solution.

1) We can answer this question two ways: using "common sense" or related rates. The common sense method states that the volume of the puddle is growing by 2 in^3/s, where

$$\text{volume of puddle} = \text{area of circle} \times \text{depth}.$$

Since the depth is constant at 1/8 in, the area must be growing by 16 in^2/s.

This approach reveals the underlying related–rates principle. Let V and A represent the Volume and Area of the puddle. We know $V = A \times \frac{1}{8}$. Take the derivative of both sides with respect to t, employing implicit differentiation.

$$\begin{aligned} V &= \frac{1}{8}A \\ \frac{d}{dt}(V) &= \frac{d}{dt}\left(\frac{1}{8}A\right) \\ \frac{dV}{dt} &= \frac{1}{8}\frac{dA}{dt} \end{aligned}$$

Since $\frac{dV}{dt} = 2$, we know $2 = \frac{1}{8}\frac{dA}{dt}$, and hence $\frac{dA}{dt} = 16$. The area is growing by 16 in^2/s.

2) To start, we need an equation that relates what we know to the radius. We just learned something about the surface area of the circular puddle, and we know $A = \pi r^2$. We should be able to learn about the rate at which the radius is growing with this information.

Implicitly derive both sides of $A = \pi r^2$ with respect to t:

$$\begin{aligned} A &= \pi r^2 \\ \frac{d}{dt}\big(A\big) &= \frac{d}{dt}\big(\pi r^2\big) \\ \frac{dA}{dt} &= 2\pi r\frac{dr}{dt} \end{aligned}$$

Our work above told us that $\frac{dA}{dt} = 16\ \text{in}^2/\text{s}$. Solving for $\frac{dr}{dt}$, we have

$$\frac{dr}{dt} = \frac{8}{\pi r}.$$

Note how our answer is not a number, but rather a function of r. In other words, *the rate at which the radius is growing depends on how big the circle already is.* If the circle is very large, adding 2 in^3 of water will not make the circle much bigger at all. If the circle dime–sized, adding the same amount of water will make a radical change in the radius of the circle.

In some ways, our problem was (intentionally) ill–posed. We need to specify a current radius in order to know a rate of change. When the puddle has a radius of 10 in, the radius is growing at a rate of

$$\frac{dr}{dt} = \frac{8}{10\pi} = \frac{4}{5\pi} \approx 0.25\text{in/s}.$$

Example 2

Radar guns measure the rate of distance change between the gun and the object it is measuring. For instance, a reading of "55 mph" means the object is moving away from the gun at a rate of 55 miles per hour, whereas a measurement of "-25 mph" would mean that the object is approaching the gun at a rate of 25 miles per hour.

If the radar gun is moving (say, attached to a police car) then radar readouts are only immediately understandable if the gun and the object are moving along the same line. If a police officer is traveling 60 mph and gets a readout of 15 mph, he knows that the car ahead of him is moving away at a rate of 15 miles an hour, meaning the car is traveling 75 mph. (This straight–line principle is one reason officers park on the side of the highway and try to shoot straight back down the road. It gives the most accurate reading.)

Suppose an officer is driving due north at 60 mph and sees a car moving due east, as shown in Figure 3.2. Using his radar gun, he measures a reading of 80 mph. By using landmarks, he believes both he and the other car are about 1/2 mile from the intersection of their two roads.

If the speed limit on the other road is 55 mph, is the other driver speeding?

Solution. Using the diagram in Figure 3.2, let's label what we know about the situation. As both the police officer and other driver are 1/2 mile from the intersection, we have $A = 1/2$, $B = 1/2$, and through the Pythagorean Theorem, $C = 1/\sqrt{2} \approx 0.707$.

We know the police officer is traveling at 60 mph; that is, $\frac{dA}{dt} = 60$.

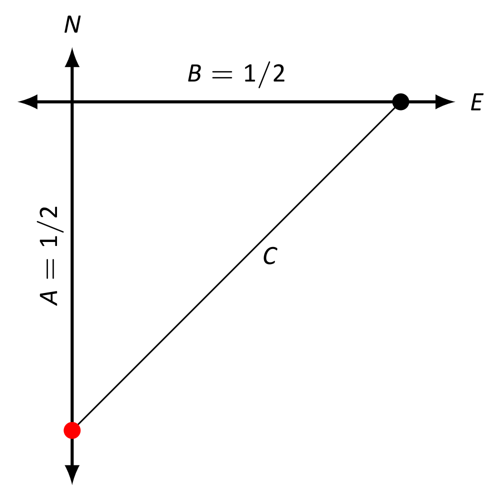


Figure 3.2: A sketch of a police car (at bottom) attempting to measure the speed of a car (at right) in Example 2.

Example 2 is both interesting and impractical. It highlights the difficulty in using radar in a non–linear fashion, and explains why "in real life" the police officer would follow the other driver to determine their speed, and not pull out pencil and paper. The principles here are important, though. Many automated vehicles make judgments about other moving objects based on perceived distances and radar–like measurements using related–rates ideas.

The radar measurement is $\frac{dC}{dt} = 80$. We want to find $\frac{dB}{dt}$.

We need an equation that contains relates B to A and/or C. The Pythagorean Theorem seems like a good choice: $A^2 + B^2 = C^2$. Differentiate both sides with respect to t:

$$\begin{aligned} A^2 + B^2 &= C^2 \\ \frac{d}{dt}\left(A^2 + B^2\right) &= \frac{d}{dt}\left(C^2\right) \\ 2A\frac{dA}{dt} + 2B\frac{dB}{dt} &= 2C\frac{dC}{dt} \end{aligned}$$

We have values for everything except $\frac{dB}{dt}$. Solving for this we have

$$\frac{dB}{dt} = \frac{C\frac{dC}{dt} - A\frac{dA}{dt}}{B} \approx 53.12\text{mph}.$$

The other driver does not appear to be speeding.

Example 3

A camera is placed on a tripod 10 ft from the side of a road. The camera is to turn to track a car that is to drive by at 100 mph for a promotional video. The video's planners want to know what kind of motor the tripod should be equipped with in order to properly track the car as it passes by. Figure 3.3 shows the proposed setup. How fast must the camera be able to turn to track the car?

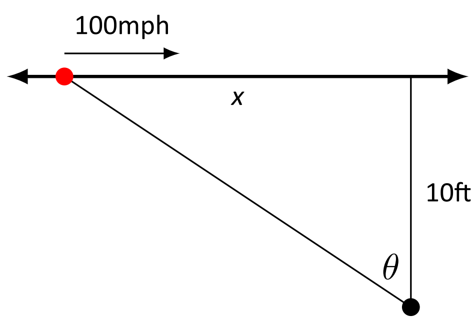


Figure 3.3: Tracking a speeding car (at left) with a rotating camera.

Solution. We seek information about how fast the camera is to *turn*; therefore, we need an equation that will relate an angle θ to the position of the camera and the speed and position of the car.

Figure 3.3 suggests we use a trigonometric equation. Letting x represent the distance the car is from the point on the road directly in front of the camera, we have

$$\tan\theta = \frac{x}{10}. \tag{3.2}$$

As the car is moving at 100 mph, we have $\frac{dx}{dt} = 100$ mph. We need to convert the measurements to common units; rewrite 100 mph in terms of ft/s:

$$\frac{dx}{dt} = 100\frac{\text{m}}{\text{h}} = 100\frac{\text{m}}{\text{h}} \cdot 5280\frac{\text{f}}{\text{m}} \cdot \frac{1}{3600}\frac{\text{h}}{\text{s}} = 146.\overline{6}\text{ft/s}.$$

Now take the derivative of both sides of Equation (3.2) using implicit differentiation:

$$\begin{aligned} \tan\theta &= \frac{x}{10} \\ \frac{d}{dt}\big(\tan\theta\big) &= \frac{d}{dt}\left(\frac{x}{10}\right) \\ \sec^2\theta\,\frac{d\theta}{dt} &= \frac{1}{10}\frac{dx}{dt} \\ \frac{d\theta}{dt} &= \frac{\cos^2\theta}{10}\frac{dx}{dt} \end{aligned} \tag{3.3}$$

We want to know the fastest the camera has to turn. Common sense tells us this is when the car is directly in front of the camera (i.e., when

$\theta = 0$). Our mathematics bears this out. In Equation (3.3) we see this is when $\cos^2\theta$ is largest; this is when $\cos\theta = 1$, or when $\theta = 0$.

With $\frac{dx}{dt} \approx 146.67$ ft/s, we have

$$\frac{d\theta}{dt} = \frac{1 \text{ rad}}{10 \text{ ft}} 146.67\text{ft/s} = 14.667 \text{ radians/s}.$$

What does this number mean? Recall that 1 circular revolution goes through 2π radians, thus 14.667 rad/s means $14.667/(2\pi) \approx 2.33$ revolutions per second.

Activity 3.1–1

A water tank has the shape of an inverted circular cone (point down) with a base of radius 6 feet and a depth of 8 feet. Suppose that water is being pumped into the tank at a constant instantaneous rate of 4 cubic feet per minute.

(a) Draw a picture of the conical tank, including a sketch of the water level at a point in time when the tank is not yet full. Introduce variables that measure the radius of the water's surface and the water's depth in the tank, and label them on your figure.

(b) Say that r is the radius and h the depth of the water at a given time, t. What equation relates the radius and height of the water, and why?

(c) Determine an equation that relates the volume of water in the tank at time t to the depth h of the water at that time.

(d) Through differentiation, find an equation that relates the instantaneous rate of change of water volume with respect to time to the instantaneous rate of change of water depth at time t.

(e) Find the instantaneous rate at which the water level is rising when the water in the tank is 3 feet deep.

(f) When is the water rising most rapidly: at $h = 3$, $h = 4$, or $h = 5$?

Recognizing familiar geometric configurations is one way that we relate the changing quantities in a given problem. For instance, while the problem in Activity 3.2–1 is centered on a conical tank, one of the most important observations is that there are two key right triangles present. In another setting, a right triangle might be indicative of an opportunity to take advantage of the Pythagorean Theorem to relate the legs of the triangle. But in the conical tank, the fact that the water at any time fills a portion of the tank in such a way that the ratio of radius to depth is constant turns out to be the most important relationship with which to work. That enables us to write r in terms of h and reduce the overall problem to one where the volume of water depends simply on h, and hence to subsequently relate $\frac{dV}{dt}$ and $\frac{dh}{dt}$. In other situations where a changing angle is involved, a right triangle may offer the opportunity to find relationships among various parts of the triangle using trigonometric functions.

Activity 3.1–2

A television camera is positioned 4000 feet from the base of a rocket launching pad. The angle of elevation of the camera has to change at the correct rate in order to keep the rocket in sight. In addition, the auto-focus of the camera has to take into account the increasing distance between the camera and the rocket. We assume that the rocket rises vertically. (A similar problem is discussed and pictured dynamically at `http://gvsu.edu/s/9t`. Exploring the applet at the link will be helpful to you in answering the questions that follow.)

(a) Draw a figure that summarizes the given situation. What parts of the picture are changing? What parts are constant? Introduce appropriate variables to represent the quantities that are changing.

(b) Find an equation that relates the camera's angle of elevation to the height of the rocket, and then find an equation that relates the instantaneous rate of change of the camera's elevation angle to the instantaneous rate of change of the rocket's height (where all rates of change are with respect to time).

(c) Find an equation that relates the distance from the camera to the rocket to the rocket's height, as well as an equation that relates the instantaneous rate of change of distance from the camera to the rocket to the instantaneous rate of change of the rocket's height (where all rates of change are with respect to time).

(d) Suppose that the rocket's speed is 600 ft/sec at the instant it has risen 3000 feet. How fast is the distance from the television camera to the rocket changing at that moment? If the camera is following the rocket, how fast is the camera's angle of elevation changing at that same moment?

(e) If from an elevation of 3000 feet onward the rocket continues to rise at 600 feet/sec, will the rate of change of distance with respect to time be greater when the elevation is 4000 feet than it was at 3000 feet, or less? Why?

In addition to being able to find instantaneous rates of change at particular points in time, we are often able to make more general observations about how particular rates themselves will change over time. For instance, when a conical tank (point down) is filling with water at a constant rate, we naturally intuit that the depth of the water should increase more slowly over time. Note how carefully we need to speak: we mean to say that while the depth, h, of the water is increasing, its rate of change $\frac{dh}{dt}$ is decreasing (both as a function of t and as a function of h). These observations may often be made by taking the general equation that relates the various rates and solving for one of them, and doing this without substituting any particular values for known variables or rates. For instance, in the conical tank problem in Activity 3.2–1, we established that

$$\frac{dV}{dt} = \frac{1}{16}\pi h^2 \frac{dh}{dt},$$

and hence

$$\frac{dh}{dt} = \frac{16}{\pi h^2}\frac{dV}{dt}.$$

Provided that $\frac{dV}{dt}$ is constant, it is immediately apparent that as h gets larger, $\frac{dh}{dt}$ will get smaller, while always remaining positive. Hence, the depth of the water is increasing at a decreasing rate.

Activity 3.1–3

As pictured in the applet at `http://gvsu.edu/s/9q`, a skateboarder who is 6 feet tall rides under a 15 foot tall lamppost at a constant rate of 3 feet per second. We are interested in understanding how fast his shadow is changing at various points in time.

(a) Draw an appropriate right triangle that represents a snapshot in time of the skateboarder, lamppost, and his shadow. Let x denote the horizontal distance from the base of the lamppost to the skateboarder and s represent the length of his shadow. Label these quantities, as well as the skateboarder's height and the lamppost's height on the diagram.

(b) Observe that the skateboarder and the lamppost represent parallel line segments in the diagram, and thus similar triangles are present. Use similar triangles to establish an equation that relates x and s.

(c) Use your work in (b) to find an equation that relates $\frac{dx}{dt}$ and $\frac{ds}{dt}$.

(d) At what rate is the length of the skateboarder's shadow increasing at the instant the skateboarder is 8 feet from the lamppost?

(e) As the skateboarder's distance from the lamppost increases, is his shadow's length increasing at an increasing rate, increasing at a decreasing rate, or increasing at a constant rate?

(f) Which is moving more rapidly: the skateboarder or the tip of his shadow? Explain, and justify your answer.

As we progress further into related rates problems, less direction will be provided. In the first three activities of this section, we have been provided with guided instruction to build a solution in a step by step way. For the closing activity and the following exercises, most of the detailed work is left to the reader.

Activity 3.1–4

A baseball diamond is $90'$ square. A batter hits a ball along the third base line runs to first base. At what rate is the distance between the ball and first base changing when the ball is halfway to third base, if at that instant the ball is traveling 100 feet/sec? At what rate is the distance between the ball and the runner changing at the same instant, if at the same instant the runner is 1/8 of the way to first base running at 30 feet/sec?

Summary

In this section, we encountered the following important ideas:

- When two or more related quantities are changing as implicit functions of time, their rates of change can be related by implicitly differentiating the equation that relates the quantities themselves. For instance, if the sides of a right triangle are all changing as functions of time, say having lengths x, y, and z, then these quantities are related by the Pythagorean Theorem: $x^2 + y^2 = z^2$. It follows by implicitly differentiating with respect to t that their rates are related by the equation

$$2x\frac{dx}{dt} + 2y\frac{dy}{dt} = 2z\frac{dz}{dt},$$

so that if we know the values of x, y, and z at a particular time, as well as two of the three rates, we can deduce the value of the third.

Exercises

Terms and Concepts

1) T/F: Implicit differentiation is often used when solving “related rates” type problems.

Problems

2) Water flows onto a flat surface at a rate of $5\text{cm}^3/\text{s}$ forming a circular puddle 10mm deep. How fast is the radius growing when the radius is:

   (a) 1 cm?

   (b) 10 cm?

   (c) 100 cm?

3) A circular balloon is inflated with air flowing at a rate of $10\text{cm}^3/\text{s}$. How fast is the radius of the balloon increasing when the radius is:

   (a) 1 cm?

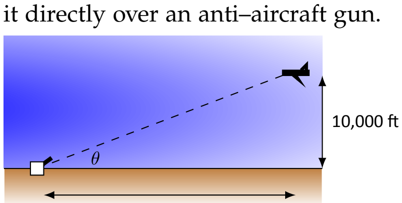
   (b) 10 cm?

   (c) 100 cm?

4) Consider the traffic situation introduced in Example 2. How fast is the “other car” traveling if the officer and the other car are each 1/2 mile from the intersection, the officer is traveling 50mph, and the radar reading is 70mph?

5) Consider the traffic situation introduced in Example 2. How fast is the “other car” traveling if the officer and the other car are each 1 mile from the intersection, the officer is traveling 60mph, and the radar reading is 80mph?

6) An F-22 aircraft is flying at 500mph with an elevation of 10,000ft on a straight–line path that will take it directly over an anti–aircraft gun.



   How fast must the gun be able to turn to accurately track the aircraft when the plane is:

   (a) 1 mile away?

   (b) 1/5 mile away?

   (c) Directly overhead?

7) An F-22 aircraft is flying at 500mph with an elevation of 100ft on a straight–line path that will take it directly over an anti–aircraft gun as in Exercise 6) (note the lower elevation here).

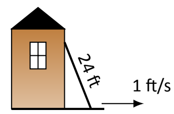
   How fast must the gun be able to turn to accurately track the aircraft when the plane is:

   (a) 1000 feet away?

   (b) 100 feet away?

   (c) Directly overhead?

8) A 24ft. ladder is leaning against a house while the base is pulled away at a constant rate of 1ft/s.



   At what rate is the top of the ladder sliding down the side of the house when the base is:

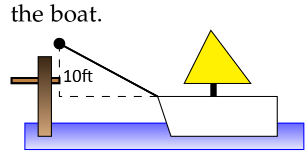
   (a) 1 foot from the house?

   (b) 10 feet from the house?

   (c) 23 feet from the house?

   (d) 24 feet from the house?

9) A boat is being pulled into a dock at a constant rate of 30ft/min by a winch located 10ft above the deck of the boat.



   At what rate is the boat approaching the dock when the boat is:

   (a) 50 feet out?

   (b) 15 feet out?

   (c) 1 foot from the dock?

   (d) What happens when the length of rope pulling in the boat is less than 10 feet long?

10) An inverted cylindrical cone, 20ft deep and 10ft across at the top, is being filled with water at a rate of $10\text{ft}^3/\text{min}$. At what rate is the water rising in the tank when the depth of the water is:

   (a) 1 foot?

   (b) 10 feet?

   (c) 19 feet?

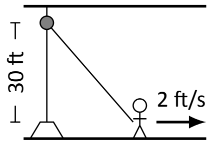
   How long will the tank take to fill when starting at empty?

11) A hot air balloon lifts off from ground rising vertically. From 100 feet away, a 5′ woman tracks the path of the balloon. When her sightline with the balloon makes a 45° angle with the horizontal, she notes the angle is increasing at about $5^\circ/\text{min}$.

   (a) What is the elevation of the balloon?

   (b) How fast is it rising?

12) A rope, attached to a weight, goes up through a pulley at the ceiling and back down to a worker. The man holds the rope at the same height as the connection point between rope and weight.



Suppose the man stands directly next to the weight (i.e., a total rope length of 60 ft) and begins to walk away at a rate of 2ft/s. How fast is the weight rising when the man has walked:

(a) 10 feet?

(b) 40 feet?

How far must the man walk to raise the weight all the way to the pulley?

13) Consider the situation described in Exercise 12). Suppose the man starts 40ft from the weight and begins to walk away at a rate of 2ft/s.

(a) How long is the rope?

(b) How fast is the weight rising after the man has walked 10 feet?

(c) How fast is the weight rising after the man has walked 40 feet?

(d) How far must the man walk to raise the weight all the way to the pulley?

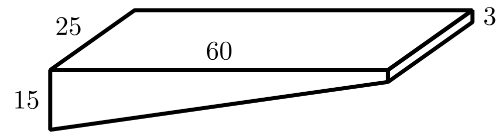
14) A company that produces landscaping materials is dumping sand into a conical pile. The sand is being poured at a rate of $5\text{ft}^3/\text{sec}$; the physical properties of the sand, in conjunction with gravity, ensure that the cone's height is roughly $2/3$ the length of the diameter of the circular base.

How fast is the cone rising when it has a height of 30 feet?

15) A baseball diamond is a square with sides 90 feet long. Suppose a baseball player is advancing from second to third base at the rate of 24 feet per second, and an umpire is standing on home plate. Let θ be the angle between the third baseline and the line of sight from the umpire to the runner. How fast is θ changing when the runner is 30 feet from third base?

16) Sand is being dumped off a conveyor belt onto a pile in such a way that the pile forms in the shape of a cone whose radius is always equal to its height. Assuming that the sand is being dumped at a rate of 10 cubic feet per minute, how fast is the height of the pile changing when there are 1000 cubic feet on the pile?

17) A swimming pool is 60 feet long and 25 feet wide. Its depth varies uniformly from 3 feet at the shallow end to 15 feet at the deep end, as shown below.



Suppose the pool has been emptied and is now being filled with water at a rate of 800 cubic feet per minute. At what rate is the depth of water (measured at the deepest point of the pool) increasing when it is 5 feet deep at that end? Over time, describe how the depth of the water will increase: at an increasing rate, at a decreasing rate, or at a constant rate. Explain.

3.2 *Using derivatives to identify extreme values of a function*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- What are the critical values of a function f and how are they connected to identifying the most extreme values the function achieves?
- How does the first derivative of a function reveal important information about the behavior of the function, including the function's extreme values?
- How can the second derivative of a function be used to help identify extreme values of the function?

Introduction

In many different settings, we are interested in knowing where a function achieves its least and greatest values. These can be important in applications – say to identify a point at which maximum profit or minimum cost occurs – or in theory to understand how to characterize the behavior of a function or a family of related functions. Consider the simple and familiar example of a parabolic function such as $s(t) = -16t^2 + 32t + 48$ (shown at left in Figure 3.4) that represents the height of an object tossed vertically: its maximum value occurs at the vertex of the parabola and represents the highest value that the object reaches. Moreover, this maximum value identifies an especially important point on the graph, the point at which the curve changes from increasing to decreasing.

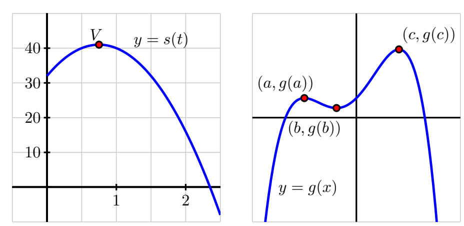


Figure 3.4: At left, $s(t) = -16t^2 + 24t + 32$ whose vertex is $(\frac{3}{4}, 41)$; at right, a function g that demonstrates several high and low points.

More generally, for any function we consider, we can investigate where its lowest and highest points occur in comparison to points nearby or to all possible points on the graph. Given a function f, we say that $f(c)$ is a *global* or *absolute maximum* provided that $f(c) \geq f(x)$ for all x in the domain of f, and similarly call $f(c)$ a *global* or *absolute minimum* whenever $f(c) \leq f(x)$ for all x in the domain of f. For instance, for the function g given at right in Figure 3.4, g has a global maximum of $g(c)$, but g does not appear to have a global minimum, as the graph of g seems to decrease without bound. We note that the point $(c, g(c))$ marks a fundamental change in the behavior of g, where g changes from increasing to decreasing; similar things happen at both $(a, g(a))$ and $(b, g(b))$, although these points are not global mins or maxes.

For any function f, we say that $f(c)$ is a *local maximum* or *relative maximum* provided that $f(c) \geq f(x)$ for all x near c, while $f(c)$ is called a *local* or *relative minimum* whenever $f(c) \leq f(x)$ for all x near c. Any maximum or minimum may be called an *extreme value* of f. For example, in Figure 3.4, g has a relative minimum of $g(b)$ at the point $(b, g(b))$ and a relative maximum of $g(a)$ at $(a, g(a))$. We have already identified the global maximum of g as $g(c)$; this global maximum can also be considered a relative maximum.

We would like to use fundamental calculus ideas to help us identify and classify key function behavior, including the location of relative extremes. Of course, if we are given a graph of a function, it is often straightforward to locate these important behaviors visually. We investigate this situation in the following preview activity.

Preview Activity 3.2

Consider the function h given by the graph in Figure 3.5. Use the graph to answer each of the following questions.

(a) Identify all values c for which $h(c)$ is a local maximum of h.

(b) Identify all values c for which $h(c)$ is a local minimum of h.

(c) Does h have a global maximum? If so, what is the value of this global maximum?

(d) Does h have a global minimum? If so, what is its value?

(e) Identify all values of c for which $h'(c) = 0$.

(f) Identify all values of c for which $h'(c)$ does not exist.

(g) True or false: every relative maximum and minimum of h occurs at a point where $h'(c)$ is either zero or does not exist.

(h) True or false: at every point where $h'(c)$ is zero or does not exist, h has a relative maximum or minimum.

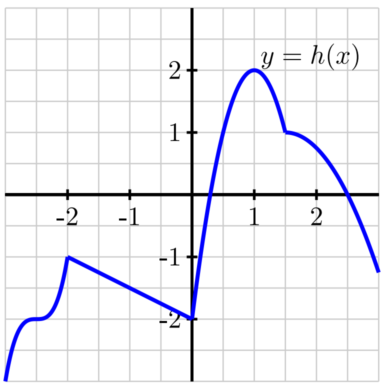


Figure 3.5: The graph of a function h on the interval $[-3, 3]$.

Increasing/Decreasing test

Before we continue examining extreme values of a function, let's recall a topic we discussed that we will use throughout this section. In Section 2.3, we discussed that whether a function is increasing or decreasing depends precisely on the value of the derivative at a point. Repeating the idea here, a function is increasing at $x = a$ if and only if $f'(a) > 0$ and decreasing at $x = a$ if and only if $f'(a) < 0$. This can be expanded to intervals in the domain of the function which we state in the following concept.

Test for Increasing/Decreasing Functions

Let f be a continuous function on $[a, b]$ and differentiable on (a, b).

1. If $f'(c) > 0$ for all c in (a, b), then f is increasing on $[a, b]$.

2. If $f'(c) < 0$ for all c in (a,b), then f is decreasing on $[a,b]$.
3. If $f'(c) = 0$ for all c in (a,b), then f is constant on $[a,b]$.

If $f'(x) = 0$ for a finite number of x values as oppossed to the entire interval (a,b)[2], the function is neither increasing or decreasing at those values. The values where $f'(x) = 0$ are very important in identifying extreme values and require more examination.

[2] If f is constant on $[a,b]$, then the graph of f is a horizontal line and therefore its rate of change is 0 for values in (a,b).

Critical values and the first derivative test

If a function has a relative extreme value at a point $(c, f(c))$, the function must change its behavior at c regarding whether it is increasing or decreasing before or after the point.

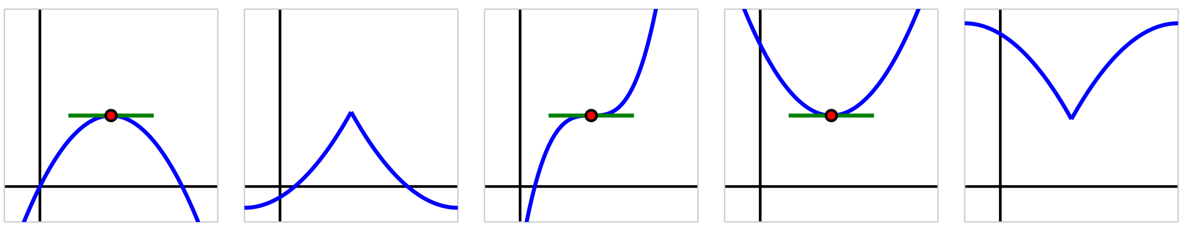


Figure 3.6: From left to right, a function with a relative maximum where its derivative is zero; a function with a relative maximum where its derivative is undefined; a function with neither a maximum nor a minimum at a point where its derivative is zero; a function with a relative minimum where its derivative is zero; and a function with a relative minimum where its derivative is undefined.

For example, if a continuous function has a relative maximum at c, such as those pictured in the two leftmost functions in Figure 3.6, then it is both necessary and sufficient that the function change from being increasing just before c to decreasing just after c. In the same way, a continuous function has a relative minimum at c if and only if the function changes from decreasing to increasing at c. See, for instance, the two functions pictured at right in Figure 3.6. There are only two possible ways for these changes in behavior to occur: either $f'(c) = 0$ or $f'(c)$ is undefined.

Because these values of c are so important, we call them *critical values*. More specifically, we say that a function f has a *critical value* at $x = c$ provided that $f'(c) = 0$ or $f'(c)$ is undefined. Critical values provide us with the only possible locations where the function f may have relative extremes. Note that not every critical value produces a maximum or minimum; in the middle graph of Figure 3.6, the function pictured there has a horizontal tangent line at the noted point, but the function is increasing before and increasing after, so the critical value does not yield a location where the function is greater than every value nearby, nor less than every value nearby.

The *first derivative test* summarizes how sign changes in the first derivative indicate the presence of a local maximum or minimum for a given function.

First Derivative Test

If p is a critical value of a continuous function f that is differentiable near p (except possibly at $x = p$), then f has a relative maximum at p if and only if f' changes sign from positive to negative at p, and f has a relative minimum at p if and only if f' changes sign from negative to positive at p.

We consider an example to show one way the first derivative test can be used to identify the relative extreme values of a function.

Example 1

Let f be a function whose derivative is given by the formula $f'(x) = e^{-2x}(3-x)(x+1)^2$. Determine all critical values of f and decide whether a relative maximum, relative minimum, or neither occurs at each.

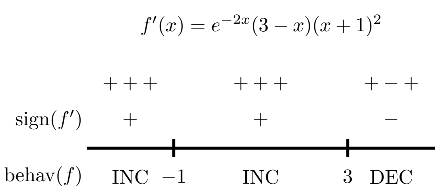


Figure 3.7: The first derivative sign chart for a function f whose derivative is given by the formula $f'(x) = e^{-2x}(3-x)(x+1)^2$.

Solution. Since we already have $f'(x)$ written in factored form, it is straightforward to find the critical values of f. Since $f'(x)$ is defined for all values of x, we need only determine where $f'(x) = 0$. From the equation

$$e^{-2x}(3-x)(x+1)^2 = 0$$

and the zero product property, it follows that $x = 3$ and $x = -1$ are critical values of f. (Note particularly that there is no value of x that makes $e^{-2x} = 0$.)

Next, to apply the first derivative test, we'd like to know the sign of $f'(x)$ at values near the critical values. Because the critical values are the only locations at which f' can change sign, it follows that the sign of the derivative is the same on each of the intervals created by the critical values: for instance, the sign of f' must be the same for every value of $x < -1$. We create a first derivative sign chart to summarize the sign of f' on the relevant intervals along with the corresponding behavior of f.

The first derivative sign chart in Figure 3.7 comes from thinking about the sign of each of the terms in the factored form of $f'(x)$ at one selected point in the interval under consideration. For instance, for $x < -1$, we could consider $x = -2$ and determine the sign of e^{-2x}, $(3-x)$, and $(x+1)^2$ at the value $x = -2$. We note that both e^{-2x} and $(x+1)^2$ are positive regardless of the value of x, while $(3-x)$ is also positive at $x = -2$. Hence, each of the three terms in f' is positive, which we indicate by writing "$+++$." Taking the product of three positive terms obviously results in a value that is positive, which we denote by the "$+$" in the interval to the left of $x = -1$ indicating the overall sign of f'. And, since f' is positive on that interval, we further know that f is increasing, which we summarize by writing "INC" to represent the corresponding behavior of f. In a similar way, we find that f' is positive and f is increasing on $-1 < x < 3$, and f' is negative and f is decreasing for $x > 3$.

Now, by the first derivative test, to find relative extremes of f we look for critical value at which f' changes sign. In this example, f' only changes sign at $x = 3$, where f' changes from positive to negative, and thus f has a relative maximum at $x = 3$. While f has a critical value at $x = -1$, since f is increasing both before and after $x = -1$, f has neither a minimum nor a maximum at $x = -1$.

Example 2

Find the intervals on which f is increasing and decreasing, and use the First Derivative Test to determine the relative extrema of f, where

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

Solution. We start by calculating f' using the Quotient Rule. We find

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

We need to find the critical values of f; we want to know when $f'(x) = 0$ and when f' is not defined. That latter is straightforward: when the denominator of f' is 0, f' is undefined. That occurs when $x = 1$. $f'(x) = 0$ when the numerator of f' is 0. That occurs when $x^2 - 2x - 3 = (x-3)(x+1) = 0$; i.e., when $x = -1, 3$.

We have found that f has three critical numbers, dividing the real number line into 4 subintervals:

$$(-\infty,-1), \quad (-1,1), \quad (1,3) \quad \text{and} \quad (3,\infty).$$

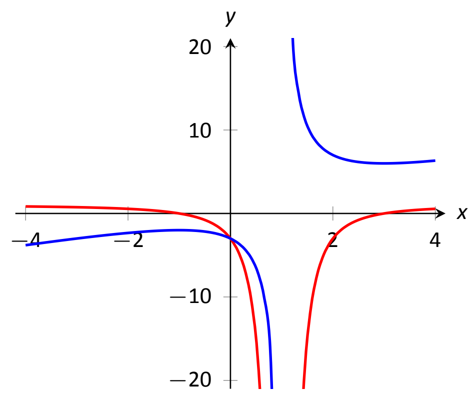


Figure 3.8: A graph of $f(x)$ in Example 2, showing where f is increasing and decreasing.

Pick a number p from each subinterval and test the sign of f′ at p to determine whether f is increasing or decreasing on that interval. Again, we do well to avoid complicated computations; notice that the denominator of f' is *always* positive so we can ignore it during our work.

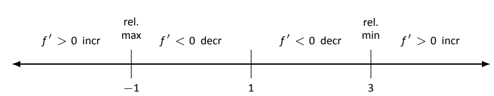


Figure 3.9: Number line for f in Example 2.

Interval 1, $(-\infty,-1)$: Choosing a very small number (i.e., a negative number with a large magnitude) p returns $p^2 - 2p - 3$ in the numerator of f'; that will be positive. Hence f is increasing on $(-\infty,-1)$.

Interval 2, $(-1,1)$: Choosing 0 seems simple: $f'(0) = -3 < 0$. We conclude f is decreasing on $(-1,1)$.

Interval 3, $(1,3)$: Choosing 2 seems simple: $f'(2) = -3 < 0$. Again, f is decreasing.

Interval 4, $(3,\infty)$: Choosing a very large number p from this subinterval will give a positive numerator and (of course) a positive denominator. So f is increasing on $(3,\infty)$.

In summary, f is increasing on the set $(-\infty,-1) \cup (3,\infty)$ and is decreasing on the set $(-1,1) \cup (1,3)$. Since at $x = -1$, the sign of f' switched from positive to negative, then $f(-1)$ is a relative maximum of f. At $x = 3$, the sign of f' switched from negative to positive, meaning

$f(3)$ is a relative minimum. At $x = 1$, f is not defined, so there is no relative extrema at $x = 1$.

This is summarized in the number line shown in Figure 3.9. Also, Figure 3.8 shows a graph of f, confirming our calculations. This figure also shows f' in red, again demonstrating that f is increasing when $f' > 0$ and decreasing when $f' < 0$.

Activity 3.2–1

Suppose that $g(x)$ is a function continuous for every value of $x \neq 2$ whose first derivative is $g'(x) = \dfrac{(x+4)(x-1)^2}{x-2}$. Further, assume that it is known that g has a vertical asymptote at $x = 2$.

(a) Determine all critical values of g.

(b) By developing a carefully labeled first derivative sign chart, decide whether g has as a local maximum, local minimum, or neither at each critical value.

(c) Does g have a global maximum? global minimum? Justify your claims.

(d) What is the value of $\lim_{x\to\infty} g'(x)$? What does the value of this limit tell you about the long-term behavior of g?

(e) Sketch a possible graph of $y = g(x)$.

The second derivative test

Recall that the second derivative of a function tells us several important things about the behavior of the function itself. For instance, if f'' is positive on an interval, then we know that f' is increasing on that interval and, consequently, that f is concave up,[3] which also tells us that throughout the interval the tangent line to $y = f(x)$ lies below the curve at every point. In this situation where we know that $f'(p) = 0$, it turns out that the sign of the second derivative determines whether f has a local minimum or local maximum at the critical value p.

[3] See Section 2.3 for the definition of concavity.

In Figure 3.10, we see the four possibilities for a function f that has a critical point p at which $f'(p) = 0$, provided $f''(p)$ is not zero on an interval including p (except possibly at p). On either side of the critical point, f'' can be either positive or negative, and hence f can be either concave up or concave down. In the first two graphs, f does not change concavity at p, and in those situations, f has either a local minimum or local maximum. In particular, if $f'(p) = 0$ and $f''(p) < 0$, then we know f is concave down at p with a horizontal tangent line, and this guarantees f has a local maximum there. This fact, along with the corresponding statement for when $f''(p)$ is positive, is stated in the *second derivative test*.

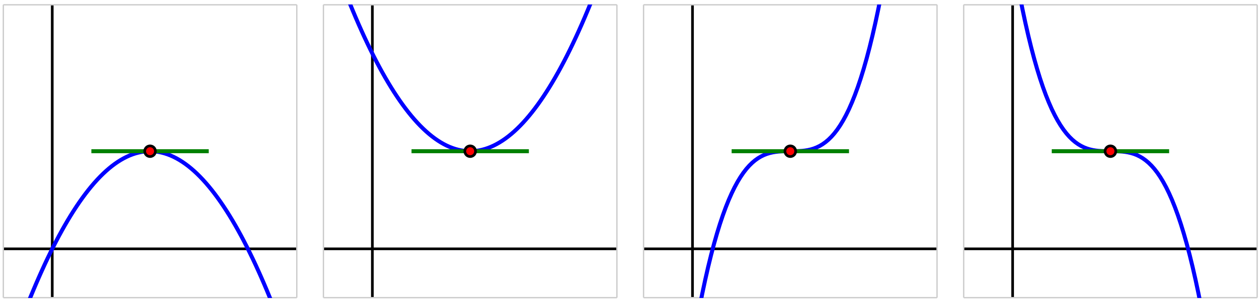


Figure 3.10: Four possible graphs of a function f with a horizontal tangent line at a critical value.

Second Derivative Test

If p is a critical value of a continuous function f such that $f'(p) = 0$ and $f''(p) \neq 0$, then f has a relative maximum at p if and only if $f''(p) < 0$, and f has a relative minimum at p if and only if $f''(p) > 0$.

In the event that $f''(p) = 0$, the second derivative test is inconclusive. That is, the test doesn't provide us any information. This is because if $f''(p) = 0$, it is possible that f has a local minimum, local maximum, or neither.[4]

[4] Consider the functions $f(x) = x^4$, $g(x) = -x^4$, and $h(x) = x^3$ at the critical point $p = 0$.

Just as a first derivative sign chart reveals all of the increasing and decreasing behavior of a function, we can construct a second derivative sign chart that demonstrates all of the important information involving concavity.

Example 3

Let $f(x)$ be a function whose first derivative is $f'(x) = 3x^4 - 9x^2$. Construct both first and second derivative sign charts for f, fully discuss where f is increasing and decreasing and concave up and concave down, identify all relative extreme values, and sketch a possible graph of f.

Solution. Since we know $f'(x) = 3x^4 - 9x^2$, we can find the critical values of f by solving $3x^4 - 9x^2 = 0$. Factoring, we observe that

$$0 = 3x^2(x^2 - 3) = 3x^2(x + \sqrt{3})(x - \sqrt{3}),$$

so that $x = 0, \pm\sqrt{3}$ are the three critical values of f. It then follows that the first derivative sign chart for f is given in Figure 3.11.

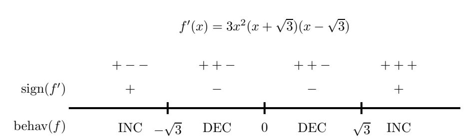


Figure 3.11: The first derivative sign chart for f when $f'(x) = 3x^4 - 9x^2 = 3x^2(x^2 - 3)$ in Example 3.

Thus, f is increasing on the intervals $(-\infty, -\sqrt{3})$ and $(\sqrt{3}, \infty)$, while f is decreasing on $(-\sqrt{3}, 0)$ and $(0, \sqrt{3})$. Note particularly that by the first derivative test, this information tells us that f has a local maximum at $x = -\sqrt{3}$ and a local minimum at $x = \sqrt{3}$. While f also has a critical value at $x = 0$, neither a maximum nor minimum occurs there since f' does not change sign at $x = 0$.

Next, we move on to investigate concavity. Differentiating $f'(x) = 3x^4 - 9x^2$, we see that $f''(x) = 12x^3 - 18x$. Since we are interested in knowing the intervals on which f'' is positive and negative, we first find where $f''(x) = 0$. Observe that

$$0 = 12x^3 - 18x = 12x\left(x^2 - \frac{3}{2}\right) = 12x\left(x + \sqrt{\frac{3}{2}}\right)\left(x - \sqrt{\frac{3}{2}}\right),$$

which implies that $x = 0, \pm\sqrt{\frac{3}{2}}$. Building a sign chart for f'' in the exact same way we do so for f, we see the result shown in Figure 3.12.

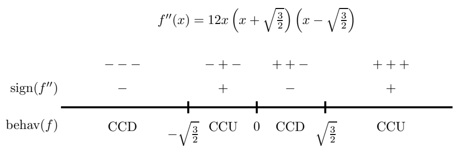


Figure 3.12: The second derivative sign chart for f when $f'(x) = 3x^4 - 9x^2 = 3x^2(x^2 - 3)$ in Example 3.

Therefore, f is concave down on the intervals $\left(-\infty, -\sqrt{\frac{3}{2}}\right)$ and $\left(0, \sqrt{\frac{3}{2}}\right)$, and concave up on $\left(0, \sqrt{\frac{3}{2}}\right)$ and $\left(\sqrt{\frac{3}{2}}, \infty\right)$.

Putting all of the above information together, we now see a complete and accurate possible graph of f in Figure 3.13.

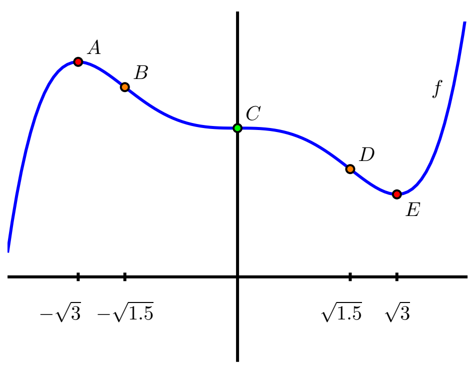


Figure 3.13: A possible graph of the function f in Example 3.

The point $A = (-\sqrt{3}, f(-\sqrt{3}))$ is a local maximum, as f is increasing prior to A and decreasing after; similarly, the point $E = (\sqrt{3}, f(\sqrt{3}))$ is a local minimum. Note, too, that f is concave down at A and concave up at B, which is consistent both with our second derivative sign chart and the second derivative test. At points B and D, concavity changes, as we saw in the results of the second derivative sign chart in Figure 3.12. Finally, at point C, f has a critical value with a horizontal tangent line, but neither a maximum nor a minimum occurs there since f is decreasing both before and after C. It is also the case that concavity changes at C.

While we completely understand where f is increasing and decreasing, where f is concave up and concave down, and where f has relative extremes, we do not know any specific information about the y-coordinates of points on the curve. For instance, while we know that f has a local maximum at $x = -\sqrt{3}$, we don't know the value of that maximum because we do not know $f(-\sqrt{3})$. Any vertical translation of our sketch of f in Figure 3.13 would satisfy the given criteria for f.

Points B, C, and D in Figure 3.13 are locations at which the concavity of f changes. We give a special name to any such point: if p is a value in the domain of a continuous function f at which f changes concavity, then we say that $(p, f(p))$ is an *inflection point* of f. Just as we look for locations where f changes from increasing to decreasing at points where $f'(p) = 0$ or $f'(p)$ is undefined, so too we find where $f''(p) = 0$ or $f''(p)$ is undefined to see if there are points of inflection at these locations.

It is important at this point in our study to remind ourselves of the big picture that derivatives help to paint: the sign of the first derivative f' tells us *whether* the function f is increasing or decreasing, while the sign of the second derivative f'' tells us *how* the function f is increasing or decreasing.

Activity 3.2–2

Suppose that g is a function whose second derivative, g'', is given by the following graph.

(a) Find all points of inflection of g.

(b) Fully describe the concavity of g by making an appropriate sign chart.

(c) Suppose you are given that $g'(-1.67857351) = 0$. Is there is a local maximum, local minimum, or neither (for the function g) at this critical value of g, or is it impossible to say? Why?

(d) Assuming that $g''(x)$ is a polynomial (and that all important behavior of g'' is seen in the given graph, what degree polynomial do you think $g(x)$ is? Why?

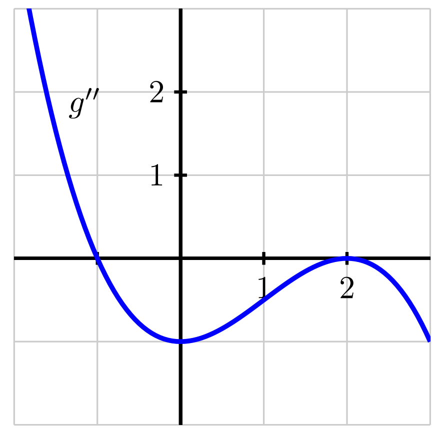


Figure 3.14: The graph of $y = g''(x)$.

We end this section with an application problem to help us better understand inflection points.

Example 4

The sales of a certain product over a three-year span are modeled by $S(t) = t^4 - 8t^2 + 20$, where t is the time in years, shown in Figure 3.15. Over the first two years, sales are decreasing. Find the point at which sales are decreasing at their greatest rate.

Solution. We want to maximize the rate of decrease, which is to say, we want to find where S' has a minimum. To do this, we find where S'' is 0. We find $S'(t) = 4t^3 - 16t$ and $S''(t) = 12t^2 - 16$. Setting $S''(t) = 0$ and solving, we get $t = \sqrt{4/3} \approx 1.16$ (we ignore the negative value of t since it does not lie in the domain of our function S).

This is both the inflection point and the point of maximum decrease. This is the point at which things first start looking up for the company. After the inflection point, it will still take some time before sales start to increase, but at least sales are not decreasing quite as quickly as they had been.

A graph of $S(t)$ and $S'(t)$ is given in Figure 3.15. When $S'(t) < 0$, sales are decreasing; note how at $t \approx 1.16$, $S'(t)$ is minimized. That is, sales are decreasing at the fastest rate at $t \approx 1.16$. On the interval of $(1.16, 2)$, S is decreasing but concave up, so the decline in sales is "leveling off."

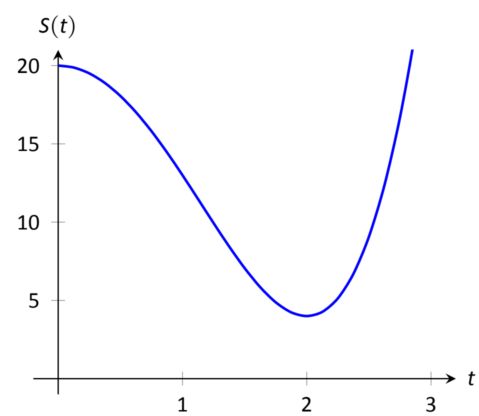


Figure 3.15: A graph of $S(t)$ in Example 4, modeling the sale of a product over time.

Curve Sketching

We have been learning how we can understand the behavior of a function based on its first and second derivatives. While we have been treating the properties of a function separately (increasing and decreasing, concave up and concave down, etc.), we combine them here to produce an accurate graph of the function without plotting lots of extraneous points.

Why bother? Graphing utilities are very accessible, whether on a computer, a hand–held calculator, or a smartphone. These resources are usually very fast and accurate. We will see that our method is not particularly fast – it will require time (but it is not *hard*). So again: why bother?

We are attempting to understand the behavior of a function f based on the information given by its derivatives. While all of a function's derivatives relay information about it, it turns out that "most" of the behavior we care about is explained by f' and f''. Understanding the interactions between the graph of f and f' and f'' is important. To gain this understanding, one might argue that all that is needed is to look at lots of graphs. This is true to a point, but is somewhat similar to stating that one understands how an engine works after looking only at pictures. It is true that the basic ideas will be conveyed, but "hands–on" access increases understanding.

The following Concept summarizes what we have learned so far that is applicable to sketching function graphs and gives a framework for putting that information together. It is followed by several examples.

Curve Sketching

To produce an accurate sketch a given function f, consider the following steps.

1) Find the domain of f.

2) Find the critical values of f.

3) Find the possible points of inflection of f.

4) Find the location of any vertical asymptotes of f–those values of a such that $\lim\limits_{x\to a^\pm} f(x) = \pm\infty$ (usually done in conjunction with item 1 above).

5) Consider the limits $\lim\limits_{x\to-\infty} f(x)$ and $\lim\limits_{x\to\infty} f(x)$ to determine the end behavior/horizontal asymptotes of the function.

6) Determine the intervals for which f is increasing or decreasing and concave up or down.

7) Evaluate f at each critical point and possible point of inflection. Plot these points on a set of axes, and connect them with curves exhibiting the proper behavior. Sketch asymptotes and x and y intercepts where applicable.

Example 5

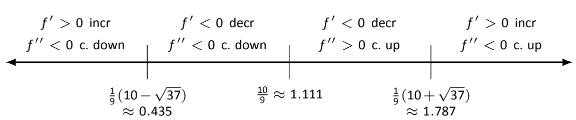
Sketch $f(x) = 3x^3 - 10x^2 + 7x + 5$.

Solution. We follow the steps outlined in the Curve Sketching Concept.

1) The domain of f is the entire real line; there are no values x for which $f(x)$ is not defined.
2) Find the critical values of f. We compute $f'(x) = 9x^2 - 20x + 7$. Use the Quadratic Formula to find the roots of f':

$$x = \frac{20 \pm \sqrt{(-20)^2 - 4(9)(7)}}{2(9)} = \frac{1}{9}\left(10 \pm \sqrt{37}\right) \Rightarrow x \approx 0.435, 1.787.$$

3) Find the possible points of inflection of f. Compute $f''(x) = 18x - 20$. We have
$$f''(x) = 0 \Rightarrow x = 10/9 \approx 1.111.$$
4) There are no vertical asymptotes.
5) We determine the end behavior using limits as x approaches $\pm$infinity.
$$\lim_{x\to-\infty} f(x) = -\infty \qquad \lim_{x\to\infty} f(x) = \infty.$$
We do not have any horizontal asymptotes.
6) We place the values $x = (10 \pm \sqrt{37})/9$ and $x = 10/9$ on a number line, as shown as shown below. We mark each subinterval as increasing or decreasing, concave up or down.



7) We plot the appropriate points on axes as shown in Figure 3.16-(a) and connect the points with straight lines. In Figure 3.16-(b) we adjust these lines to demonstrate the proper concavity. Our curve crosses the y axis at $y = 5$ and crosses the x axis near $x = -0.424$. In Figure 3.16-(c) we show a graph of f drawn with a computer program, verifying the accuracy of our sketch.

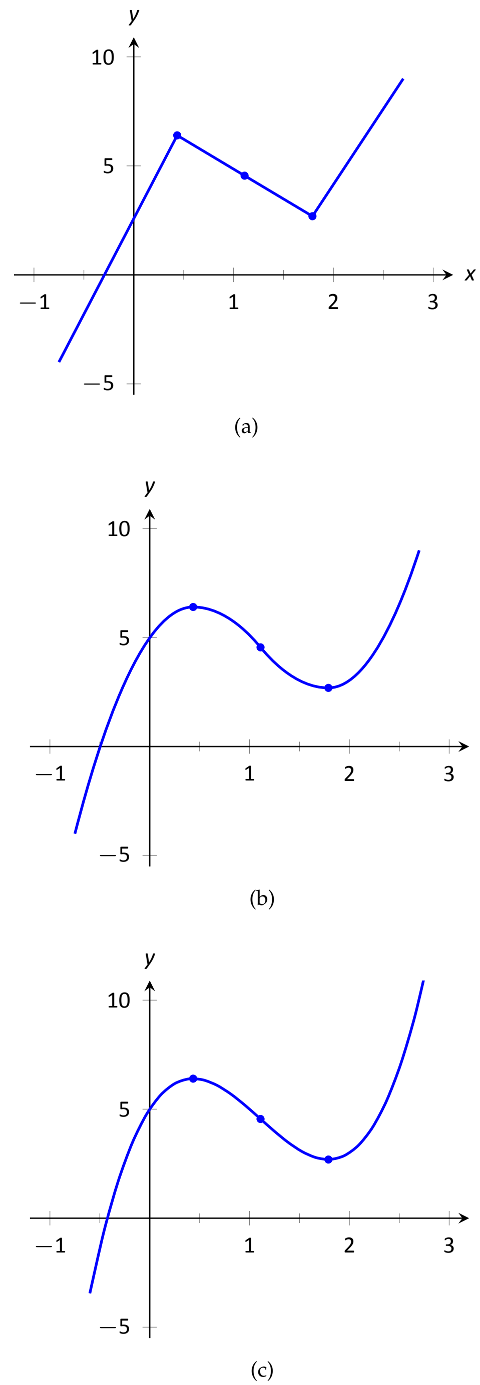


Figure 3.16: Sketching f in Example 5.

Example 6

Sketch $f(x) = \dfrac{x^2 - x - 2}{x^2 - x - 6}$.

Solution. We follow the steps outlined in the Curve Sketching Concept.

1) In determining the domain, we assume it is all real numbers and looks for restrictions. We find that at $x = -2$ and $x = 3$, $f(x)$ is not defined. So the domain of f is $D = \{\text{real numbers } x \mid x \neq -2, 3\}$.
2) To find the critical values of f, we first find $f'(x)$. Using the Quotient

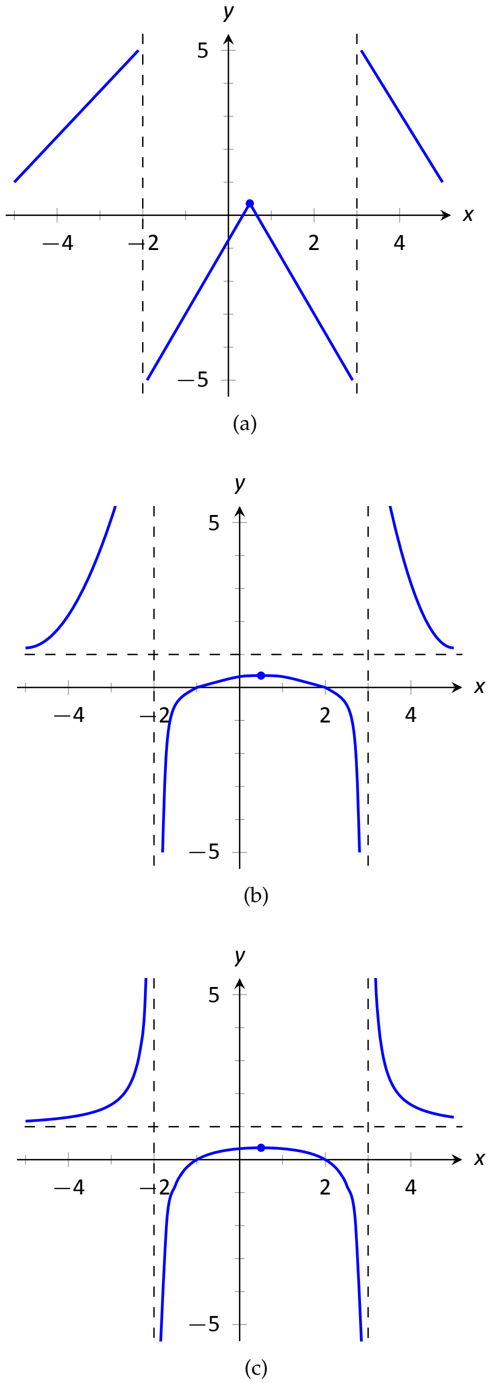


Figure 3.17: Sketching f in Example 6.

Rule, we find

$$f'(x) = \frac{-8x+4}{(x^2+x-6)^2} = \frac{-8x+4}{(x-3)^2(x+2)^2}.$$

We find that $f'(x) = 0$ when $x = 1/2$, and f' is undefined when $x = -2, 3$.

3) To find the possible points of inflection, we find $f''(x)$, again employing the Quotient Rule:

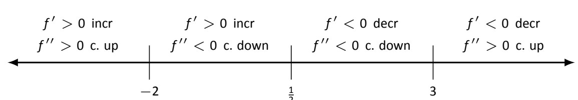
$$f''(x) = \frac{24x^2 - 24x + 56}{(x-3)^3(x+2)^3}.$$

We find that $f''(x)$ is never 0 (setting the numerator equal to 0 and solving for x, we find the only roots to this quadratic are complex) and f'' is undefined when $x = -2, 3$. Thus concavity will possibly only change at $x = -2$ and $x = 3$.

4) The vertical asymptotes of f are at $x = -2$ and $x = 3$, the places where f is undefined.

5) There is a horizontal asymptote of $y = 1$, as $\lim\limits_{x\to-\infty} f(x) = 1$ and $\lim\limits_{x\to\infty} f(x) = 1$.

6) We place the values $x = 1/2$, $x = -2$ and $x = 3$ on a number line as shown below. We mark in each interval whether f is increasing or decreasing, concave up or down. We see that f has a relative maximum at $x = 1/2$; concavity changes only at the vertical asymptotes.



7) In Figure 3.17-(a), we plot the points from the number line on a set of axes and connect the points with straight lines to get a general idea of what the function looks like (these lines effectively only convey increasing/decreasing information).

In Figure 3.17-(b), we adjust the graph with the appropriate concavity. We also show f crossing the x axis at $x = -1$ and $x = 2$.

Figure 3.17-(c) shows a computer generated graph of f, which verifies the accuracy of our sketch.

Example 7

Sketch $f(x) = \dfrac{5(x-2)(x+1)}{x^2+2x+4}$.

Solution. We again follow the steps outlined in the Curve Sketching Concept.

1) We assume that the domain of f is all real numbers and consider restrictions. The only restrictions come when the denominator is 0, but this never occurs. Therefore the domain of f is all real numbers, $\mathbb{R}$.

2) We find the critical values of f by setting $f'(x) = 0$ and solving for x.

We find

$$f'(x) = \frac{15x(x+4)}{(x^2+2x+4)^2} \quad \Rightarrow \quad f'(x) = 0 \text{ when } x = -4, 0.$$

3) We find the possible points of inflection by solving $f''(x) = 0$ for x. We find

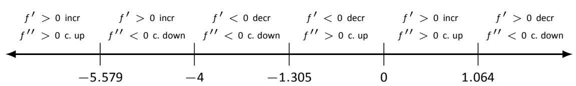
$$f''(x) = -\frac{30x^3 + 180x^2 - 240}{(x^2+2x+4)^3}.$$

The cubic in the numerator does not factor very "nicely." We instead approximate the roots at $x = -5.759$, $x = -1.305$ and $x = 1.064$.

4) There are no vertical asymptotes.

5) We have a horizontal asymptote of $y = 5$, as $\lim\limits_{x\to-\infty} f(x) = \lim\limits_{x\to\infty} f(x) = 5$.

6) We place the critical points and possible points on a number line as shown below and mark each interval as increasing/decreasing, concave up/down appropriately.



7) In Figure 3.18-(a) we plot the significant points from the number line as well as the two roots of f, $x = -1$ and $x = 2$, and connect the points with straight lines to get a general impression about the graph. In Figure 3.18-(b), we add concavity. Figure 3.18-(c) shows a computer generated graph of f, affirming our results.

In each of our examples, we found significant points on the graph of f that corresponding to changes in increasing/decreasing or concavity. We connected these points with straight lines, then adjusted for concavity, and finished by showing a very accurate, computer generated graph.

Why are computer graphics so good? It is not because computers are "smarter" than we are. Rather, it is largely because computers are much faster at computing than we are. In general, computers graph functions much like most students do when first learning to draw graphs: they plot equally spaced points, then connect the dots using lines. By using lots of points, the connecting lines are short and the graph looks smooth.

This does a fine job of graphing in most cases (in fact, this is the method used for many graphs in this text). However, in regions where the graph is very "curvy," this can generate noticeable sharp edges on the graph unless a large number of points are used. High quality computer algebra systems, such as *Mathematica*, use special algorithms to plot lots of points only where the graph is "curvy."

In Figure 3.19, a graph of $y = \sin(x)$ is given, generated by *Mathematica*. The small points represent each of the places *Mathematica* sampled the function. Notice how at the "bends" of

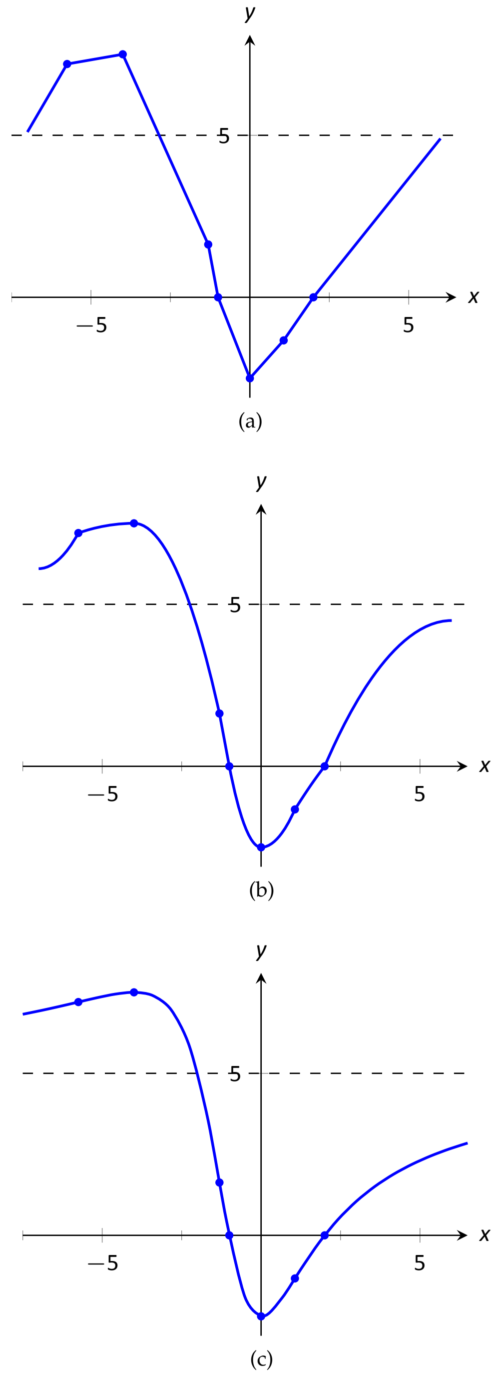


Figure 3.18: Sketching f in Example 7.

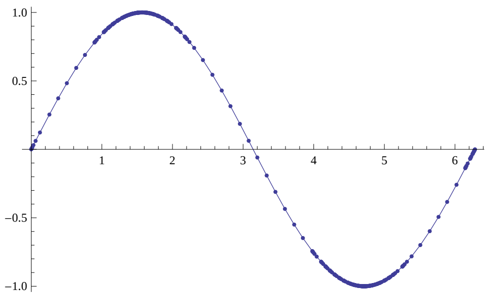


Figure 3.19: A graph of $y = \sin(x)$ generated by *Mathematica*.

$\sin(x)$, lots of points are used; where $\sin(x)$ is relatively straight, fewer points are used. (Many points are also used at the end-points to ensure the "end behavior" is accurate.)

How does *Mathematica* know where the graph is "curvy"? Calculus. When we study *curvature* in a later chapter, we will see how the first and second derivatives of a function work together to provide a measurement of "curviness." *Mathematica* employs algorithms to determine regions of "high curvature" and plots extra points there.

Summary

In this section, we encountered the following important ideas:

- The critical values of a continuous function f are the values of p for which $f'(p) = 0$ or $f'(p)$ does not exist. These values are important because they identify horizontal tangent lines or corner points on the graph, which are the only possible locations at which a local maximum or local minimum can occur.
- Given a differentiable function f, whenever f' is positive, f is increasing; whenever f' is negative, f is decreasing. The first derivative test tells us that at any point where f changes from increasing to decreasing, f has a local maximum, while conversely at any point where f changes from decreasing to increasing f has a local minimum.
- Given a twice differentiable function f, if we have a horizontal tangent line at $x = p$ and $f''(p)$ is nonzero, then the fact that f'' tells us the concavity of f will determine whether f has a maximum or minimum at $x = p$. In particular, if $f'(p) = 0$ and $f''(p) < 0$, then f is concave down at p and f has a local maximum there, while if $f'(p) = 0$ and $f''(p) > 0$, then f has a local minimum at p. If $f'(p) = 0$ and $f''(p) = 0$, then the second derivative does not tell us whether f has a local extreme at p or not.

Exercises

Terms and Concepts

1) In your own words describe what it means for a function to be increasing.
2) What does a decreasing function "look like"?
3) Sketch a graph of a function on $[0,2]$ that is increasing but not strictly increasing.
4) Give an example of a function describing a situation where it is "bad" to be increasing and "good" to be decreasing.
5) A function f has derivative $f'(x) = (\sin(x+2))e^{x^2+1}$, where $f'(x) > 1$ for all x. Is f increasing, decreasing, or can we not tell from the given information?
6) Sketch a graph of a function $f(x)$ that is:
   (a) Increasing, concave up on $(0,1)$,
   (b) increasing, concave down on $(1,2)$,
   (c) decreasing, concave down on $(2,3)$ and
   (d) increasing, concave down on $(3,4)$.
7) Is is possible for a function to be increasing and concave down on $(0,\infty)$ with a horizontal asymptote of $y=1$? If so, give a sketch of such a function.
8) Is is possible for a function to be increasing and concave up on $(0,\infty)$ with a horizontal asymptote of $y=1$? If so, give a sketch of such a function.

Problems

In exercises 9–18, a function is given.

(a) Find the critical numbers of f.

(b) Determine the intervals on which f is increasing and decreasing.

(c) Use the First Derivative Test to determine whether each critical point is a local maximum, local minimum, or neither.

(d) Use the Second Derivative Test to determine whether each critical point is a local maximum, local minimum, or if the test fails.

9) $f(x) = x^2 + 2x - 3$
10) $f(x) = x^3 + 3x^2 + 3$
11) $f(x) = 2x^3 + x^2 - x + 3$
12) $f(x) = x^3 - 3x^2 + 3x - 1$
13) $f(x) = \dfrac{1}{x^2 - 2x + 2}$
14) $f(x) = \dfrac{x^2 - 4}{x^2 - 1}$
15) $f(x) = \dfrac{x}{x^2 - 2x - 8}$
16) $f(x) = \dfrac{(x-2)^{2/3}}{x}$
17) $f(x) = \sin x \cos x$ on $(-\pi, \pi)$
18) $f(x) = x^5 - 5x$

In exercises 19–18, a function is given.

(a) Find the possible points of inflection of f.

(b) Determine the intervals on which f is concave up and concave down.

(c) Determine whether each possible point of inflection is in fact a point of inflection.

19) $f(x) = x^2 - 2x + 1$
20) $f(x) = -x^2 - 5x + 7$
21) $f(x) = x^3 - x + 1$
22) $f(x) = 2x^3 - 3x^2 + 9x + 5$
23) $f(x) = \dfrac{x^4}{4} + \dfrac{x^3}{3} - 2x + 3$
24) $f(x) = -3x^4 + 8x^3 + 6x^2 - 24x + 2$
25) $f(x) = x^4 - 4x^3 + 6x^2 - 4x + 1$
26) $f(x) = \dfrac{1}{x^2 + 1}$
27) $f(x) = \dfrac{x}{x^2 - 1}$
28) $f(x) = \sin x + \cos x$ on $(-\pi, \pi)$
29) $f(x) = x^2 e^x$
30) $f(x) = x^2 \ln x$
31) $f(x) = e^{-x^2}$

In exercises 32–45, sketch a graph of the given function using the steps outlined in the Curve Sketching Concept.

32) $f(x) = x^3 - 2x^2 + 4x + 1$
33) $f(x) = -x^3 + 5x^2 - 3x + 2$
34) $f(x) = x^3 + 3x^2 + 3x + 1$
35) $f(x) = x^3 - x^2 - x + 1$
36) $f(x) = (x-2)\ln(x-2)$
37) $f(x) = (x-2)^2 \ln(x-2)$
38) $f(x) = \dfrac{x^2 - 4}{x^2}$
39) $f(x) = \dfrac{x^2 - 4x + 3}{x^2 - 6x + 8}$
40) $f(x) = \dfrac{x^2 - 2x + 1}{x^2 - 6x + 8}$
41) $f(x) = x\sqrt{x+1}$
42) $f(x) = x^2 e^x$
43) $f(x) = \sin x \cos x$ on $[-\pi, \pi]$
44) $f(x) = (x-3)^{2/3} + 2$
45) $f(x) = \dfrac{(x-1)^{2/3}}{x}$

3.3 *Global Optimization*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- What are the differences between finding relative extreme values and global extreme values of a function?
- How is the process of finding the global maximum or minimum of a function over the function's entire domain different from determining the global maximum or minimum on a restricted domain?
- For a function that is guaranteed to have both a global maximum and global minimum on a closed, bounded interval, what are the possible points at which these extreme values occur?

Introduction

We have seen that we can use the first derivative of a function to determine where the function is increasing or decreasing, and the second derivative to know where the function is concave up or concave down. Each of these approaches provides us with key information that helps us determine the overall shape and behavior of the graph, as well as whether the function has a relative minimum or relative maximum at a given critical value. Remember that the difference between a relative maximum and a global maximum is that there is a relative minimum of f at $x = p$ if $f(p) \geq f(x)$ for all x near p, while there is a global maximum at p if $f(p) \geq f(x)$ for all x in the domain of f.

For instance, in Figure 3.20, we see a function f that has a global maximum at $x = c$ and a relative maximum at $x = a$, since $f(c)$ is greater than $f(x)$ for every value of x, while $f(a)$ is only greater than the value of $f(x)$ for x near a. Since the function appears to decrease without bound, f has no global minimum, though clearly f has a relative minimum at $x = b$.

Our emphasis in this section is on finding the global extreme values of a function (if they exist). In so doing, we will either be interested in the behavior of the function over its entire domain or on some restricted portion. The former situation is familiar and similar to work that we did in the two preceding sections of the text. We explore this through a particular example in the following preview activity.

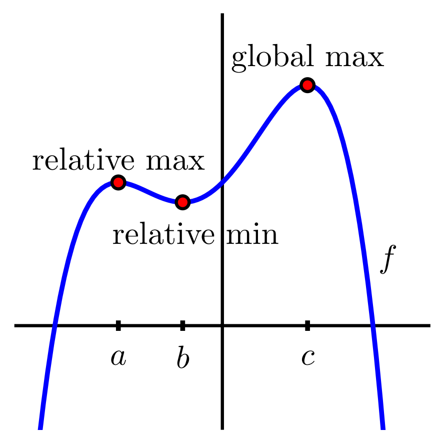


Figure 3.20: A function f with a global maximum, but no global minimum.

Preview Activity 3.3

Let $f(x) = 2 + \dfrac{3}{1+(x+1)^2}$.

(a) Determine all of the critical values of f.

(b) Construct a first derivative sign chart for f and thus determine all intervals on which f is increasing or decreasing.

(c) Does f have a global maximum? If so, why, and what is its value and where is the maximum attained? If not, explain why.

(d) Determine $\lim_{x\to\infty} f(x)$ and $\lim_{x\to-\infty} f(x)$.

(e) Explain why $f(x) > 2$ for every value of x.

(f) Does f have a global minimum? If so, why, and what is its value and where is the minimum attained? If not, explain why.

Global Optimization

For the functions in Figure 3.20 and Preview Activity 3.3, we were interested in finding the global minimum and global maximum on the entire domain, which turned out to be $(-\infty, \infty)$ for each. At other times, our perspective on a function might be more focused due to some restriction on its domain. For example, rather than considering $f(x) = 2 + \frac{3}{1+(x+1)^2}$ for every value of x, perhaps instead we are only interested in those x for which $0 \leq x \leq 4$, and we would like to know which values of x in the interval $[0, 4]$ produce the largest possible and smallest possible values of f. We are accustomed to critical values playing a key role in determining the location of extreme values of a function; now, by restricting the domain to an interval, it makes sense that the endpoints of the interval will also be important to consider, as we see in the following activity. When limiting ourselves to a particular interval, we will often refer to the *absolute* maximum or minimum value, rather than the *global* maximum or minimum.

Activity 3.3–1

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

(a) Find all critical values of g that lie in the interval $-2 \leq x \leq 3$.

(b) Use a graphing utility to construct the graph of g on the interval $-2 \leq x \leq 3$.

(c) From the graph, determine the x-values at which the absolute minimum and absolute maximum of g occur on the interval $[-2, 3]$.

(d) How do your answers change if we instead consider the interval $-2 \leq x \leq 2$?

(e) What if we instead consider the interval $-2 \leq x \leq 1$?

In Activity 3.3–1, we saw how the absolute maximum and absolute minimum of a function on a closed, bounded interval $[a, b]$, depend not only on the critical values of the function, but also on the selected values of a and b. These observations demonstrate several important facts that hold much more generally. First, we state an important result called the Extreme Value Theorem.

The Extreme Value Theorem

If f is a continuous function on a closed interval $[a, b]$, then f attains both an absolute minimum and absolute maximum on $[a, b]$. That is, for some value x_m such that $a \leq x_m \leq b$, it follows that $f(x_m) \leq f(x)$ for all x in $[a, b]$. Similarly, there is a value x_M in $[a, b]$ such that $f(x_M) = M$ for all x in $[a, b]$. Letting $m = f(x_m)$ and $M = f(x_M)$, it follows that $m \leq f(x) \leq M$ for all x in $[a, b]$.

The Extreme Value Theorem tells us that provided a function is continuous, on any closed interval $[a, b]$ the function has to achieve both an absolute minimum and an absolute maximum. Note, however, that this result does not tell us where these extreme values occur, but rather only that they must exist. As seen in the examples of Activity 3.3–1, it is apparent that the only possible locations for relative extremes are either the endpoints of the interval or at a critical value (the latter being where a relative miniumum or maximum could occur, which is a potential location for an absolute extreme). Thus, we have the following approach to finding the absolute maximum and minimum of a continuous function f on the interval $[a, b]$:

- find all critical values of f that lie in the interval;
- evaluate the function f at each critical value in the interval and at each endpoint of the interval;
- from among the noted function values, the smallest is the absolute minimum of f on the interval, while the largest is the absolute maximum.

Example 1

Find the extreme values of $f(x) = 2x^3 + 3x^2 - 12x$ on $[0, 3]$, graphed in Figure 3.21.

Solution. We follow the steps outlined. We first evaluate f at the endpoints:

$$f(0) = 0 \quad \text{and} \quad f(3) = 45.$$

Next, we find the critical values of f on $[0, 3]$. $f'(x) = 6x^2 + 6x - 12 = 6(x + 2)(x - 1)$; therefore the critical values of f are $x = -2$ and $x = 1$. Since $x = -2$ does not lie in the interval $[0, 3]$, we ignore it. Evaluating f at the only critical number in our interval gives: $f(1) = -7$.

The table in Table 3.1 gives f evaluated at the "important" x values in $[0, 3]$. We can easily see the maximum and minimum values of f: the maximum value is 45 and the minimum value is -7.

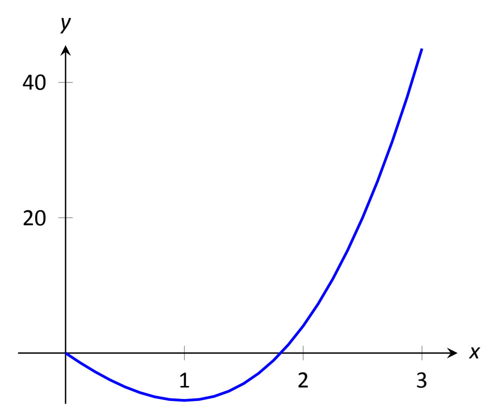


Figure 3.21: A graph of $f(x) = 2x^3 + 3x^2 - 12x$ on $[0, 3]$ as in Example 1.

x	$f(x)$
0	0
1	-7
3	45

Table 3.1: Finding the extreme values of f in Example 1.

Example 2

Find the maximum and minimum values of f on $[-4,2]$, where

$$f(x) = \begin{cases} (x-1)^2 & x \le 0 \\ x+1 & x > 0 \end{cases}.$$

Solution. Here f is piecewise–defined, but we can still apply our approach. Evaluating f at the endpoints gives:

$$f(-4) = 25 \quad \text{and} \quad f(2) = 3.$$

We now find the critical numbers of f. We have to define f' in a piecewise manner; it is

$$f'(x) = \begin{cases} 2(x-1) & x < 0 \\ 1 & x > 0 \end{cases}.$$

Note that while f is defined for all of $[-4,2]$, f' is not, as the derivative of f does not exist when $x=0$. (From the left, the derivative approaches -2; from the right the derivative is 1.) Thus one critical number of f is $x = 0$.

We now set $f'(x) = 0$. When $x > 0$, $f'(x)$ is never 0. When $x < 0$, $f'(x)$ is also never 0. (We may be tempted to say that $f'(x) = 0$ when $x = 1$. However, this is nonsensical, for we only consider $f'(x) = 2(x-1)$ when $x < 0$, so we will ignore a solution that says $x = 1$.

So we have three important x values to consider: $x = -4, 2$ and 0. Evaluating f at each gives, respectively, 25, 3 and 1, shown in Table 3.2. Thus the absolute minimum of f is 1; the absolute maximum of f is 25. Our answer is confirmed by the graph of f in Figure 3.22.

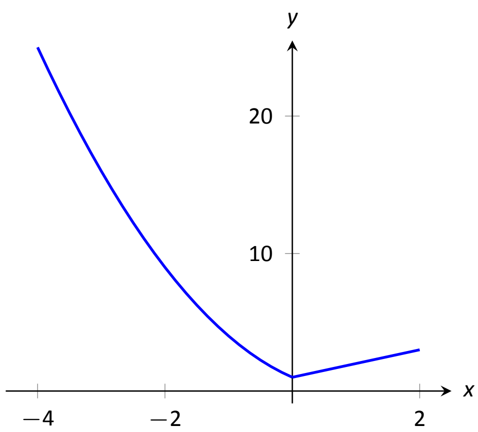


Figure 3.22: A graph of $f(x)$ on $[-4,2]$ as in Example 2.

x	$f(x)$
-4	25
0	1
2	3

Table 3.2: Finding the extreme values of f in Example 2.

Example 3

Find the extrema of $f(x) = \cos(x^2)$ on $[-2,2]$.

Solution. We again our approach to find extrema. Evaluating f at the endpoints of the interval gives: $f(-2) = f(2) = \cos(4) \approx -0.6536$. We now find the critical values of f.

Applying the Chain Rule, we find $f'(x) = -2x\sin(x^2)$. Set $f'(x) = 0$ and solve for x to find the critical values of f.

We have $f'(x) = 0$ when $x = 0$ and when $\sin(x^2) = 0$. In general, $\sin(t) = 0$ when $t = \ldots -2\pi, -\pi, 0, \pi, \ldots$ Thus $\sin(x^2) = 0$ when $x^2 = 0, \pi, 2\pi, \ldots$ (x^2 is always positive so we ignore $-\pi$, etc.) So $\sin(x^2) = 0$ when $x = 0, \pm\sqrt{\pi}, \pm\sqrt{2\pi}, \ldots$. The only values to fall in the given interval of $[-2,2]$ are $-\sqrt{\pi}$ and $\sqrt{\pi}$, approximately ± 1.77.

We again construct a table of important values in Table 3.3. In this example we have 5 values to consider: $x = 0, \pm 2, \pm\sqrt{\pi}$.

From the table it is clear that the maximum value of f on $[-2,2]$ is 1; the minimum value is -1. The graph in Figure 3.23 confirms our results.

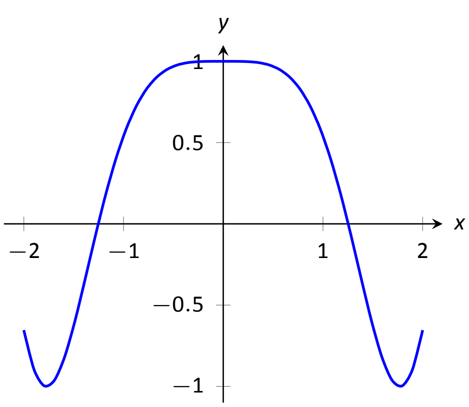


Figure 3.23: A graph of $f(x) = \cos(x^2)$ on $[-2,2]$ as in Example 3.

x	$f(x)$
-2	-0.65
$-\sqrt{\pi}$	-1
0	1
$\sqrt{\pi}$	-1
2	-0.65

Table 3.3: Finding the extreme values of $f(x) = \cos(x^2)$ in Example 3.

Activity 3.3–2

Find the *exact* absolute maximum and minimum of each function on the stated interval.

(a) $h(x) = xe^{-x}$, $[0, 3]$

(b) $p(t) = \sin(t) + \cos(t)$, $[-\frac{\pi}{2}, \frac{\pi}{2}]$

(c) $q(x) = \frac{x^2}{x-2}$, $[3, 7]$

(d) $f(x) = 4 - e^{-(x-2)^2}$, $(-\infty, \infty)$

(e) $h(x) = xe^{-ax}$, $[0, \frac{2}{a}]$ $(a > 0)$

(f) $f(x) = b - e^{-(x-a)^2}$, $(-\infty, \infty)$, $a, b > 0$

One of the big lessons in finding absolute extreme values is the realization that the interval we choose has nearly the same impact on the problem as the function under consideration. Consider, for instance, the function pictured in Figure 3.24. In sequence, from left to right, as we see the interval under consideration change from $[-2, 3]$ to $[-2, 2]$ to $[-2, 1]$, we move from having two critical values in the interval with the absolute minimum at one critical value and the absolute maximum at the right endpoint, to still having both critical numbers in the interval but then with the absolute minimum and maximum at the two critical values, to finally having just one critical value in the interval with the absolute maximum at one critical value and the absolute minimum at one endpoint. It is particularly essential to always remember to only consider the critical values that lie within the interval.

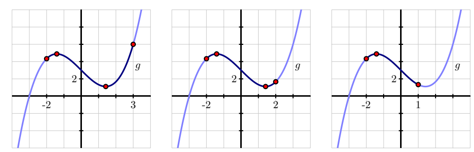


Figure 3.24: A function g considered on three different intervals.

Moving towards applications

In Section 3.4, we will focus almost exclusively on applied optimization problems: problems where we seek to find the absolute maximum or minimum value of a function that represents some physical situation. We conclude this current section with an example of one such problem because it highlights the role that a closed, bounded domain can play in finding absolute extrema. In addition, these problems often involve considerable preliminary work to develop the function which is to be optimized, and this example demonstrates that process.

Example 4

A 20 cm piece of wire is cut into two pieces. One piece is used to form a square and the other an equilateral triangle. How should the wire be cut to maximize the total area enclosed by the square and triangle? to minimize the area?

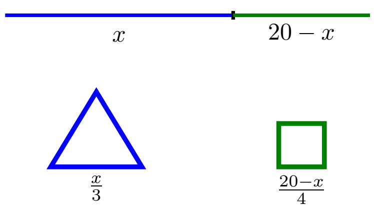


Figure 3.25: A 20 cm piece of wire cut into two pieces, one of which forms an equilateral triangle, the other which yields a square.

Solution. We begin by constructing a picture that exemplifies the given situation. The primary variable in the problem is where we decide to cut the wire. We thus label that point x, and note that the remaining portion of the wire then has length $20 - x$

As shown in Figure 3.25, we see that the x cm of the wire that are used to form the equilateral triangle result in a triangle with three sides of length $\frac{x}{3}$. For the remaining $20 - x$ cm of wire, the square that results will have each side of length $\frac{20-x}{4}$.

At this point, we note that there are obvious restrictions on x: in particular, $0 \le x \le 20$. In the extreme cases, all of the wire is being used to make just one figure. For instance, if $x = 0$, then all 20 cm of wire are used to make a square that is 5×5.

Now, our overall goal is to find the absolute minimum and absolute maximum areas that can be enclosed. We note that the area of the triangle is $A_\triangle = \frac{1}{2}bh = \frac{1}{2} \cdot \frac{x}{3} \cdot \frac{x\sqrt{3}}{6}$, since the height of an equilateral triangle is $\sqrt{3}$ times half the length of the base. Further, the area of the square is $A_\square = s^2 = \left(\frac{20-x}{4}\right)^2$. Therefore, the total area function is

$$A(x) = \frac{\sqrt{3}x^2}{36} + \left(\frac{20-x}{4}\right)^2.$$

Again, note that we are only considering this function on the restricted domain $[0, 20]$ and we seek its absolute minimum and absolute maximum.

Differentiating $A(x)$, we have

$$A'(x) = \frac{\sqrt{3}x}{18} + 2\left(\frac{20-x}{4}\right)\left(-\frac{1}{4}\right) = \frac{\sqrt{3}}{18}x + \frac{1}{8}x - \frac{5}{2}.$$

Setting $A'(x) = 0$, it follows that $x = \frac{180}{4\sqrt{3}+9} \approx 11.3007$ is the only critical value of A, and we note that this lies within the interval $[0, 20]$.

Evaluating A at the critical value and endpoints, we see that

- $A\left(\frac{180}{4\sqrt{3}+9}\right) = \frac{\sqrt{3}(\frac{180}{4\sqrt{3}+9})^2}{4} + \left(\frac{20 - \frac{180}{4\sqrt{3}+9}}{4}\right)^2 \approx 10.8741$
- $A(0) = 25$
- $A(20) = \frac{\sqrt{3}}{36}(400) = \frac{100}{9}\sqrt{3} \approx 19.2450$

Thus, the absolute minimum occurs when $x \approx 11.3007$ and results in the minimum area of approximately 10.8741 square centimeters, while the absolute maximum occurs when we invest all of the wire in the square (and none in the triangle), resulting in 25 square centimeters of area. These results are confirmed by a plot of $y = A(x)$ on the interval $[0, 20]$, as shown in Figure 3.26.

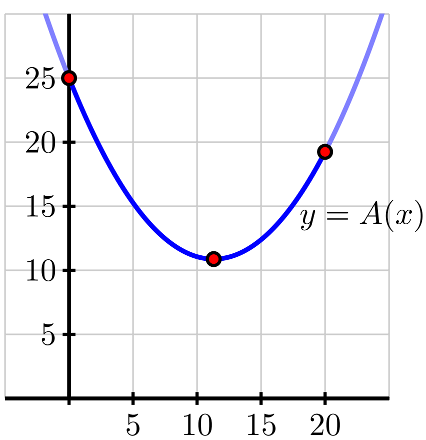


Figure 3.26: A plot of the area function from Example 4.

Activity 3.3–3

A piece of cardboard that is 10×15 (each measured in inches) is being made into a box without a top. To do so, squares are cut from each corner of the box and the remaining sides are folded up. If the box needs to be at least 1 inch deep and no more than 3 inches deep, what is the maximum possible volume of the box? what is the minimum volume? Justify your answers using calculus.

(a) Draw a labeled diagram that shows the given information. What variable should we introduce to represent the choice we make in creating the box? Label the diagram appropriately with the variable, and write a sentence to state what the variable represents.

(b) Determine a formula for the function V (that depends on the variable in (a)) that tells us the volume of the box.

(c) What is the domain of the function V? That is, what values of x make sense for input? Are there additional restrictions provided in the problem?

(d) Determine all critical values of the function V.

(e) Evaluate V at each of the endpoints of the domain and at any critical values that lie in the domain.

(f) What is the maximum possible volume of the box? the minimum?

The approaches shown in Example 4 and experienced in Activity 3.3–3 include standard steps that we undertake in almost every applied optimization problem: we draw a picture to demonstrate the situation, introduce one or more variables to represent quantities that are changing, work to find a function that models the quantity to be optimized, and then decide an appropriate domain for that function. Once that work is done, we are in the familiar situation of finding the absolute minimum and maximum of a function over a particular domain, at which time we apply the calculus ideas that we have been studying to this point in Chapter 3.

Summary

In this section, we encountered the following important ideas:

- To find relative extreme values of a function, we normally use a first derivative sign chart and classify all of the function's critical values. If instead we are interested in absolute extreme values, we first decide whether we are considering the entire domain of the function or a particular interval.
- In the case of finding global extremes over the function's entire domain, we again use a first or second derivative sign chart in an effort to make overall conclusions about whether or not the function can have a absolute maximum or minimum. If we are working to find absolute extremes on a restricted interval, then we first identify all critical values of the function that lie in the interval.
- For a continuous function on a closed, bounded interval, the only possible points at which absolute extreme values occur are the critical values and the endpoints. Thus, to find said absolute extremes, we simply evaluate the function at each endpoint and each critical value in the interval, and then we compare the results to decide which is largest (the absolute maximum) and which is smallest (the absolute minimum).

Exercises

Terms and Concepts

1) Describe what an "extreme value" of a function is in your own words.

2) Sketch the graph of a function f on $(-1, 1)$ that has both a maximum and minimum value.

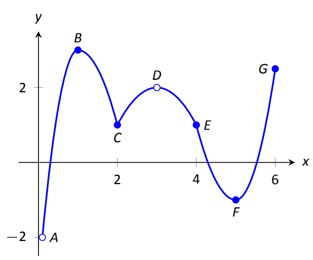
3) Describe the difference between and absolute and relative maximum in your own words.

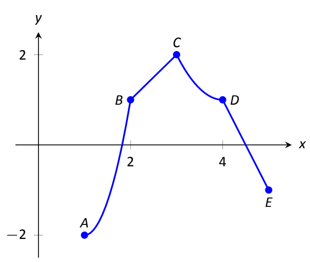
4) Sketch the graph of a function f where f has a relative maximum at $x = 1$ and $f'(1)$ is undefined.

5) T/F: If c is a critical value of a function f, then f has either a relative maximum or relative minimum at $x = c$.

Problems

In exercises 6–7, identify each of the marked points as an absolute maximum or minimum, a local maximum or minimum, or none of the above.

6) 

7) 

In exercises 8–17, find the absolute extreme values of the given function on the specified interval.

8) $f(x) = x^2 + x + 4$ on $[-1, 2]$.

9) $f(x) = x^3 - \dfrac{9}{2}x^2 - 30x + 3$ on $[0, 6]$.

10) $f(x) = 3\sin(x)$ on $[\pi/4, 2\pi/3]$.

11) $f(x) = x^2\sqrt{4 - x^2}$ on $[-2, 2]$.

12) $f(x) = x + \dfrac{3}{x}$ on $[1, 5]$.

13) $f(x) = \dfrac{x^2}{x^2 + 5}$ on $[-3, 5]$.

14) $f(x) = e^x \cos(x)$ on $[0, \pi]$.

15) $f(x) = e^x \sin(x)$ on $[0, \pi]$.

16) $f(x) = \dfrac{\ln(x)}{x}$ on $[1, 4]$.

17) $f(x) = x^{2/3} - x$ on $[0, 2]$.

18) Based on the given information about each function, decide whether the function has global maximum, a global minimum, neither, both, or that it is not possible to say without more information. Assume that each function is twice differentiable and defined for all real numbers, unless noted otherwise. In each case, write one sentence to explain your conclusion.

(a) f is a function such that $f''(x) < 0$ for every x.

(b) g is a function with two critical values a and b (where $a < b$), and $g'(x) < 0$ for $x < a$, $g'(x) < 0$ for $a < x < b$, and $g'(x) > 0$ for $x > b$.

(c) h is a function with two critical values a and b (where $a < b$), and $h'(x) < 0$ for $x < a$, $h'(x) > 0$ for $a < x < b$, and $h'(x) < 0$ for $x > b$. In addition, $\lim_{x\to\infty} h(x) = 0$ and $\lim_{x\to-\infty} h(x) = 0$.

(d) p is a function differentiable everywhere except at $x = a$ and $p''(x) > 0$ for $x < a$ and $p''(x) < 0$ for $x > a$.

19) For each of the functions described below (each continuous on $[a, b]$), state the location of the function's absolute maximum and absolute minimum on the interval $[a, b]$, or say there is not enough information provided to make a conclusion. Assume that any critical values mentioned in the problem statement represent all of the critical numbers the function has in $[a, b]$. In each case, write one sentence to explain your answer.

(a) $f'(x) \le 0$ for all x in $[a, b]$

(b) g has a critical value at c such that $a < c < b$ and $g'(x) > 0$ for $x < c$ and $g'(x) < 0$ for $x > c$

(c) $h(a) = h(b)$ and $h''(x) < 0$ for all x in $[a, b]$

(d) $p(a) > 0$, $p(b) < 0$, and for the critical value c such that $a < c < b$, $p'(x) < 0$ for $x < c$ and $p'(x) > 0$ for $x > c$

20) Let $s(t) = 3\sin(2(t - \frac{\pi}{6})) + 5$. Find the exact absolute maximum and minimum of s on the provided intervals by testing the endpoints and finding and evaluating all relevant critical values of s.

(a) $[\frac{\pi}{6}, \frac{7\pi}{6}]$

(b) $[0, \frac{\pi}{2}]$

(c) $[0, 2\pi]$

(d) $[\frac{\pi}{3}, \frac{5\pi}{6}]$

3.4 *Applied Optimization*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- In a setting where a situation is described for which optimal parameters are sought, how do we develop a function that models the situation and use calculus to find the desired maximum or minimum?

Introduction

Near the conclusion of Section 3.3, we considered two examples of optimization problems where determining the function to be optimized was part of a broader question. In Example **??**, we sought to use a single piece of wire to build two geometric figures (an equilateral triangle and square) and to understand how various choices for how to cut the wire led to different values of the area enclosed. One of our conclusions was that in order to maximize the total combined area enclosed by the triangle and square, all of the wire must be used to make a square. In the subsequent Activity 3.3–3, we investigated how the volume of a box constructed from a piece of cardboard by removing squares from each corner and folding up the sides depends on the size of the squares removed.

Both of these problems exemplify situations where there is not a function explicitly provided to optimize. Rather, we first worked to understand the given information in the problem, drawing a figure and introducing variables, and then sought to develop a formula for a function that models the quantity (area or volume, in the two examples, respectively) to be optimized. Once the function was established, we then considered what domain was appropriate on which to pursue the desired absolute minimum or maximum (or both). At this point in the problem, we are finally ready to apply the ideas of calculus to determine and justify the absolute minimum or maximum. Thus, what is primarily different about problems of this type is that the problem-solver must do considerable work to introduce variables and develop the correct function and domain to represent the described situation.

Throughout what follows in the current section, the primary emphasis is on the reader solving problems. Initially, some substantial guidance is provided, with the problems progressing to require greater independence as we move along.

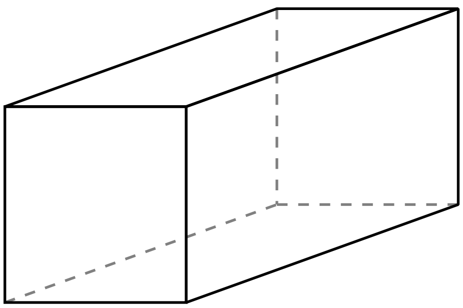


Figure 3.27: A rectangular parcel with a square end.

Preview Activity 3.4

According to U.S. postal regulations, the girth plus the length of a parcel sent by mail may not exceed 108 inches, where by "girth" we mean the perimeter of the smallest end. What is the largest possible volume of a rectangular parcel with a square end that can be sent by mail? What are the dimensions of the package of largest volume?

(a) Let x represent the length of one side of the square end and y the length of the longer side. Label these quantities appropriately on the image shown in Figure 3.27.

(b) What is the quantity to be optimized in this problem? Find a formula for this quantity in terms of x and y.

(c) The problem statement tells us that the parcel's girth plus length may not exceed 108 inches. In order to maximize volume, we assume that we will actually need the girth plus length to equal 108 inches. What equation does this produce involving x and y?

(d) Solve the equation you found in (c) for one of x or y (whichever is easier).

(e) Now use your work in (b) and (d) to determine a formula for the volume of the parcel so that this formula is a function of a single variable.

(f) Over what domain should we consider this function? Note that both x and y must be positive; how does the constraint that girth plus length is 108 inches produce intervals of possible values for x and y?

(g) Find the absolute maximum of the volume of the parcel on the domain you established in (f) and hence also determine the dimensions of the box of greatest volume. Justify that you've found the maximum using calculus.

More applied optimization problems

Many of the steps in Preview Activity 3.4 are ones that we will execute in any applied optimization problem. We briefly summarize those here to provide an overview of our approach in subsequent questions.

- Draw a picture and introduce variables. It is essential to first understand what quantities are allowed to vary in the problem and then to represent those values with variables. Constructing a figure with the variables labeled is almost always an essential first step. Sometimes drawing several diagrams can be especially helpful to get a sense of the situation. A nice example of this can be seen at `http://gvsu.edu/s/99`, where the choice of where to bend a piece of wire into the shape of a rectangle determines both the rectangle's shape and area.
- Identify the quantity to be optimized as well as any key relationships among the variable quantities. Essentially this step

involves writing equations that involve the variables that have been introduced: one to represent the quantity whose minimum or maximum is sought, and possibly others that show how multiple variables in the problem may be interrelated.

- Determine a function of a single variable that models the quantity to be optimized; this may involve using other relationships among variables to eliminate one or more variables in the function formula. For example, in Preview Activity 3.4, we initially found that $V = x^2y$, but then the additional relationship that $4x + y = 108$ (girth plus length equals 108 inches) allows us to relate x and y and thus observe equivalently that $y = 108 - 4x$. Substituting for y in the volume equation yields $V(x) = x^2(108 - 4x)$, and thus we have written the volume as a function of the single variable x.

- Decide the domain on which to consider the function being optimized. Often the physical constraints of the problem will limit the possible values that the independent variable can take on. Thinking back to the diagram describing the overall situation and any relationships among variables in the problem often helps identify the smallest and largest values of the input variable.

- Use calculus to identify the absolute maximum and/or minimum of the quantity being optimized. This always involves finding the critical values of the function first. Then, depending on the domain, we either construct a first derivative sign chart (for an open or unbounded interval) or evaluate the function at the endpoints and critical values (for a closed, bounded interval), using ideas we've studied so far in Chapter 3.

- Finally, we make certain we have answered the question: does the question seek the absolute maximum of a quantity, or the values of the variables that produce the maximum? That is, finding the absolute maximum volume of a parcel is different from finding the dimensions of the parcel that produce the maximum.

Example 1

A man has 100 feet of fencing, a large yard, and a small dog. He wants to create a rectangular enclosure for his dog with the fencing that provides the maximal area. What dimensions provide the maximal area?

Solution. One can likely guess the correct answer – that is great. We will proceed to show how calculus can provide this answer in a context that proves this answer is correct.

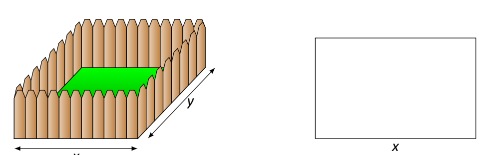


Figure 3.28: A sketch of the enclosure in Example 1

It helps to make a sketch of the situation. Our enclosure is sketched twice in Figure 3.28, either with green grass and nice fence boards or as a simple rectangle. Either way, drawing a rectangle forces us to realize that we need to know the dimensions of this rectangle so we can create an area function – after all, we are trying to maximize the area.

We let x and y denote the lengths of the sides of the rectangle. Clearly,

$$\text{Area} = xy.$$

We do not yet know how to handle functions with 2 variables; we need to reduce this down to a single variable. We know more about the situation: the man has 100 feet of fencing. By knowing the perimeter of the rectangle must be 100, we can create another equation:

$$\text{Perimeter} = 100 = 2x + 2y.$$

We now have 2 equations and 2 unknowns. In the latter equation, we solve for y:

$$y = 50 - x.$$

Now substitute this expression for y in the area equation:

$$\text{Area} = A(x) = x(50 - x).$$

Note we now have an equation of one variable; we can truly call the Area a function of x.

This function only makes sense when $0 \leq x \leq 50$, otherwise we get negative values of area. So we find the extreme values of $A(x)$ on the interval $[0, 50]$.

To find the critical points, we take the derivative of $A(x)$ and set it equal to 0, then solve for x.

$$\begin{aligned} A(x) &= x(50 - x) \\ &= 50x - x^2 \\ A'(x) &= 50 - 2x \end{aligned}$$

We solve $50 - 2x = 0$ to find $x = 25$; this is the only critical point. We evaluate $A(x)$ at the endpoints of our interval and at this critical point to find the extreme values; in this case, all we care about is the maximum.

Clearly $A(0) = 0$ and $A(50) = 0$, whereas $A(25) = 625\text{ft}^2$. This is the maximum. Since we earlier found $y = 50 - x$, we find that y is also 25. Thus the dimensions of the rectangular enclosure with perimeter of 100 ft. with maximum area is a square, with sides of length 25 ft.

Example 2

A woman has a 100 feet of fencing, a small dog, and a large yard that contains a stream (that is mostly straight). She wants to create a rectangular enclosure with maximal area that uses the stream as one side. (Apparently her dog won't swim away.) What dimensions provide the maximal area?

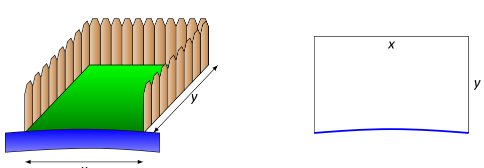


Figure 3.29: A sketch of the enclosure in Example 2

Solution. We will follow the outlined steps.

1) We are maximizing *area*. A sketch of the region will help; Figure 3.29

gives two sketches of the proposed enclosed area. A key feature of the sketches is to acknowledge that one side is not fenced.

2) We want to maximize the area; as in the example before,

$$\text{Area} = xy.$$

This is our fundamental equation. This defines area as a function of two variables, so we need another equation to reduce it to one variable.

We again appeal to the perimeter; here the perimeter is

$$\text{Perimeter} = 100 = x + 2y.$$

Note how this is different than in our previous example.

3) We now reduce the fundamental equation to a single variable. In the perimeter equation, solve for y: $y = 50 - 1/2x$. We can now write Area as

$$\text{Area} = A(x) = x(50 - 1/2x) = 50x - 1/2x^2.$$

Area is now defined as a function of one variable.

4) We want the area to be nonnegative. Since $A(x) = x(50 - 1/2x)$, we want $x \geq 0$ and $50 - 1/2x \geq 0$. The latter inequality implies that $x \leq 100$, so $0 \leq x \leq 100$.

5) We now find the extreme values. At the endpoints, the minimum is found, giving an area of 0.

Find the critical points. We have $A'(x) = 50 - x$; setting this equal to 0 and solving for x returns $x = 50$. This gives an area of

$$A(50) = 50(25) = 1250.$$

6) We earlier set $y = 50 - 1/2x$; thus $y = 25$. Thus our rectangle will have two sides of length 25 and one side of length 50, with a total area of 1250 ft^2.

Activity 3.4–1

A soup can in the shape of a right circular cylinder is to be made from two materials. The material for the side of the can costs \$0.015 per square inch and the material for the lids costs \$0.027 per square inch. Suppose that we desire to construct a can that has a volume of 16 cubic inches. What dimensions minimize the cost of the can?

(a) Draw a picture of the can and label its dimensions with appropriate variables.

(b) Use your variables to determine expressions for the volume, surface area, and cost of the can.

(c) Determine the total cost function as a function of a single variable. What is the domain on which you should consider this function?

(d) Find the absolute minimum cost and the dimensions that produce this value.

Familiarity with common geometric formulas is particularly helpful in problems like the one in Activity 3.4–1. Sometimes those involve perimeter, area, volume, or surface area. At other times, the constraints of a problem introduce right triangles (where the Pythagorean Theorem applies) or other functions whose formulas provide relationships among variables present.

Example 3

A power line needs to be run from an power station located on the beach to an offshore facility. Figure 3.30 shows the distances between the power station to the facility.

It costs \$50/ft. to run a power line along the land, and \$130/ft. to run a power line under water. How much of the power line should be run along the land to minimize the overall cost? What is the minimal cost?

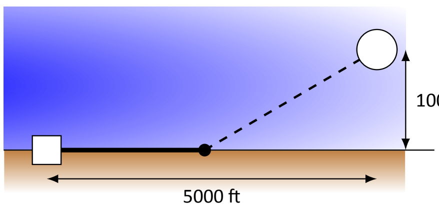


Figure 3.30: Running a power line from the power station to an offshore facility with minimal cost in Example 3

Solution. We will follow the strategy implicitly, without specifically numbering steps.

There are two immediate solutions that we could consider, each of which we will reject through "common sense." First, we could minimize the distance by directly connecting the two locations with a straight line. However, this requires that all the wire be laid underwater, the most costly option. Second, we could minimize the underwater length by running a wire all 5000 ft. along the beach, directly across from the offshore facility. This has the undesired effect of having the longest distance of all, probably ensuring a non–minimal cost.

The optimal solution likely has the line being run along the ground for a while, then underwater, as the figure implies. We need to label our unknown distances – the distance run along the ground and the distance run underwater. Recognizing that the underwater distance can be measured as the hypotenuse of a right triangle, we choose to label the distances as shown in Figure 3.31.

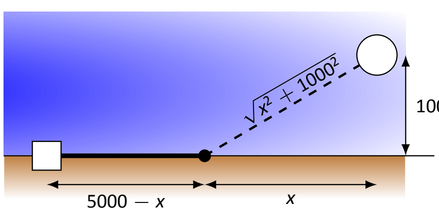


Figure 3.31: Labeling unknown distances in Example 3

By choosing x as we did, we make the expression under the square root simple. We now create the cost function.

$$\begin{array}{rcccc} \text{Cost} & = & \text{land cost} & + & \text{water cost} \\ & & \$50 \times \text{land distance} & + & \$130 \times \text{water distance} \\ & & 50(5000-x) & + & 130\sqrt{x^2+1000^2}. \end{array}$$

So we have $c(x) = 50(5000-x) + 130\sqrt{x^2+1000^2}$. This function only makes sense on the interval $[0, 5000]$. While we are fairly certain the endpoints will not give a minimal cost, we still evaluate $c(x)$ at each to verify.

$$c(0) = 380{,}000 \qquad c(5000) \approx 662{,}873.$$

We now find the critical values of $c(x)$. We compute $c'(x)$ as

$$c'(x) = -50 + \frac{130x}{\sqrt{x^2+1000^2}}.$$

Recognize that this is never undefined. Setting $c'(x) = 0$ and solving for x, we have:

$$\begin{aligned}
-50 + \frac{130x}{\sqrt{x^2+1000^2}} &= 0 \\
\frac{130x}{\sqrt{x^2+1000^2}} &= 50 \\
\frac{130^2x^2}{x^2+1000^2} &= 50^2 \\
130^2x^2 &= 50^2(x^2+1000^2) \\
130^2x^2 - 50^2x^2 &= 50^2 \cdot 1000^2 \\
(130^2-50^2)x^2 &= 50,000^2 \\
x^2 &= \frac{50,000^2}{130^2-50^2} \\
x &= \frac{50,000}{\sqrt{130^2-50^2}} \\
x &= \frac{50,000}{120} = 416\frac{2}{3}
\end{aligned}$$

Evaluating $c(x)$ at $x = 416.67$ gives a cost of about \$370,000. The distance the power line is laid along land is $5000 - 416.67 = 4583.33$ ft., and the underwater distance is $\sqrt{416.67^2 + 1000^2} \approx 1083$ ft.

Activity 3.4–2

A hiker starting at a point P on a straight road walks east towards point Q, which is on the road and 3 kilometers from point P. Two kilometers due north of point Q is a cabin. The hiker will walk down the road for a while, at a pace of 8 kilometers per hour. At some point Z between P and Q, the hiker leaves the road and makes a straight line towards the cabin through the woods, hiking at a pace of 3 kph, as pictured in Figure 3.32. In order to minimize the time to go from P to Z to the cabin, where should the hiker turn into the forest?

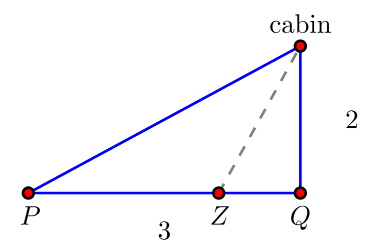


Figure 3.32: A hiker walks from P to Z to the cabin, as pictured.

In more geometric problems, we often use curves or functions to provide natural constraints. For instance, we could investigate which isosceles triangle that circumscribes a unit circle has the smallest area, which you can explore for yourself at `http://gvsu.edu/s/9b`. Or similarly, for a region bounded by a parabola, we might seek the rectangle of largest area that fits beneath the curve, as shown at `http://gvsu.edu/s/9c`. The next activity is similar to the latter problem.

Activity 3.4–3

Consider the region in the x-y plane that is bounded by the x-axis and the function $f(x) = 25 - x^2$. Construct a rectangle whose base lies on the x-axis and is centered at the origin, and whose sides extend vertically until they intersect the curve $y = 25 - x^2$. Which such rectangle has the maximum possible area? Which such rectangle has the greatest perimeter? Which has the greatest combined perimeter and area? (Challenge:

answer the same questions in terms of positive parameters a and b for the function $f(x) = b - ax^2$.)

Activity 3.4–4

A trough is being constructed by bending a 4×24 (measured in feet) rectangular piece of sheet metal. Two symmetric folds 2 feet apart will be made parallel to the longest side of the rectangle so that the trough has cross-sections in the shape of a trapezoid, as pictured in Figure 3.33. At what angle should the folds be made to produce the trough of maximum volume?

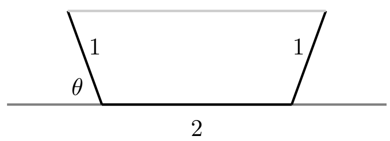


Figure 3.33: A cross-section of the trough formed by folding to an angle of θ.

Summary

In this section, we encountered the following important ideas:

- While there is no single algorithm that works in every situation where optimization is used, in most of the problems we consider, the following steps are helpful: draw a picture and introduce variables; identify the quantity to be optimized and find relationships among the variables; determine a function of a single variable that models the quantity to be optimized; decide the domain on which to consider the function being optimized; use calculus to identify the absolute maximum and/or minimum of the quantity being optimized.

Exercises

Terms and Concepts

1) T/F: An "optimization problem" is essentially an "extreme values" problem in a "story problem" setting.

2) T/F: This section teaches one to find the extreme values of function that have more than one variable.

Problems

3) Find the maximum product of two numbers (not necessarily integers) that have a sum of 100.

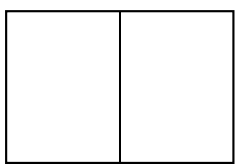
4) Find the minimum sum of two numbers whose product is 500.

5) Find the maximum sum of two numbers whose product is 500.

6) Find the maximum sum of two numbers, each of which is in $[0, 300]$ whose product is 500.

7) Find the maximal area of a right triangle with hypotenuse of length 1.

8) A rancher has 1000 feet of fencing in which to construct adjacent, equally sized rectangular pens. What dimensions should these pens have to maximize the enclosed area?

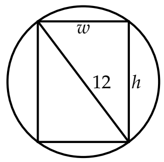


9) A standard soda can is roughly cylindrical and holds 355cm^3 of liquid. What dimensions should the cylinder be to minimize the material needed to produce the can? Based on your dimensions, determine whether or not the standard can is produced to minimize the material costs.

10) Find the dimensions of a cylindrical can with a volume of 206in^3 that minimizes the surface area.

    The "#10 can"is a standard sized can used by the restaurant industry that holds about 206in^3 with a diameter of 6 2/16in and height of 7in. Does it seem these dimensions where chosen with minimization in mind?

11) The strength S of a wooden beam is directly proportional to its cross sectional width w and the square of its height h; that is, $S = kwh^2$ for some constant k.



    Given a circular log with diameter of 12 inches, what sized beam can be cut from the log with maximum strength?

12) A power line is to be run to an offshore facility in the manner described in Example 3. The offshore facility is 2 miles at sea and 5 miles along the shoreline from the power plant. It costs \$50,000 per mile to lay a power line underground and \$80,000 to run the line underwater.

    How much of the power line should be run underground to minimize the overall costs?

13) A power line is to be run to an offshore facility in the manner described in Example 3. The offshore facility is 5 miles at sea and 2 miles along the shoreline from the power plant. It costs \$50,000 per mile to lay a power line underground and \$80,000 to run the line underwater.

    How much of the power line should be run underground to minimize the overall costs?

14) A woman throws a stick into a lake for her dog to fetch; the stick is 20 feet down the shore line and 15 feet into the water from there. The dog may jump directly into the water and swim, or run along the shore line to get closer to the stick before swimming. The dog runs about 22ft/s and swims about 1.5ft/s.

    How far along the shore should the dog run to minimize the time it takes to get to the stick?

15) A woman throws a stick into a lake for her dog to fetch; the stick is 15 feet down the shore line and 30 feet into the water from there. The dog may jump directly into the water and swim, or run along the shore line to get closer to the stick before swimming. The dog runs about 22ft/s and swims about 1.5ft/s.

    How far along the shore should the dog run to minimize the time it takes to get to the stick? *(Google "calculus dog" to learn more about a dog's ability to minimize times.)*

16) What are the dimensions of the rectangle with largest area that can be drawn inside the unit circle?

17) A rectangular box with a square bottom and closed top is to be made from two materials. The material for the side costs \$1.50 per square foot and the material for the bottom costs \$3.00 per square foot. If you are willing to spend \$15 on the box, what is the largest volume it can contain? Justify your answer completely using calculus.

18) A farmer wants to start raising cows, horses, goats, and sheep, and desires to have a rectangular pasture for the animals to graze in. However, no two different kinds of animals can graze together. In order to minimize the amount of fencing she will need, she has decided to enclose a large rectangular area and then divide it into four equally sized pens, or grazing areas. She has decided to purchase 7500 ft of fencing. What is the maximum possible area that each of the four pens will enclose?

19) Two vertical towers of heights 60 ft and 80 ft stand on level ground, with their bases 100 ft apart. A cable that is stretched from the top of one pole to some point on the ground between the poles, and then to the top of the other pole. What is the minimum possible length of cable required? Justify your answer completely using calculus.

20) A company is designing propane tanks that are cylindrical with hemispherical ends. Assume that the company wants tanks that will hold 1000 cubic feet of gas, and that the ends are more expensive to make, costing \$5 per square foot, while the cylindrical barrel between the ends costs \$2 per square foot. Use calculus to determine the minimum cost to construct such a tank.

3.5 *The Mean Value Theorem*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- In a setting where the average rate of change of a function f is known for a given interval, is there a value where the function has instantaneous rate of change equal to the average rate of change?

Introduction

We motivate this section with the following question: Suppose you leave your house and drive to your friend's house in a city 120 miles away, completing the trip in two hours. At any point during the trip do you necessarily have to be going 60 miles per hour?

In answering this question, it is clear that the *average* speed for the entire trip is 60 mph (i.e. 120 miles in 2 hours), but the question is whether or not your *instantaneous* speed is ever exactly 60 mph. More simply, does your speedometer ever read exactly 60 mph?. The answer, under some very reasonable assumptions, is "yes."

Let's now see why this situation is in a calculus text by translating it into mathematical symbols.

First assume that the function $y = f(t)$ gives the distance (in miles) traveled from your home at time t (in hours) where $0 \leq t \leq 2$. In particular, this gives $f(0) = 0$ and $f(2) = 120$. The slope of the secant line connecting the starting and ending points $(0, f(0))$ and $(2, f(2))$ is therefore

$$\frac{\Delta f}{\Delta t} = \frac{f(2) - f(0)}{2 - 0} = \frac{120 - 0}{2} = 60\,\text{mph}.$$

The slope at any point on the graph itself is given by the derivative $f'(t)$. So, since the answer to the question above is "yes," this means that at some time during the trip, the derivative takes on the value of 60 mph. Symbolically,

$$f'(c) = \frac{f(2) - f(0)}{2 - 0} = 60$$

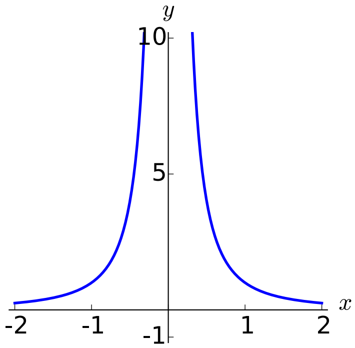
for some time $0 \leq c \leq 2$.

The Mean Value Theorem

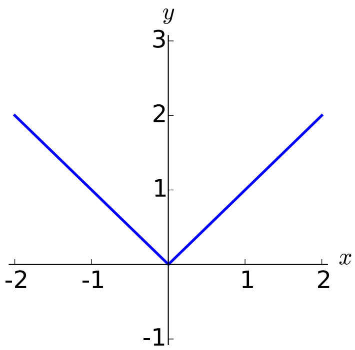
Can we generalize the velocity problem above to any function? In other words, given any function $y = f(x)$ and a range $a \leq$

$x \le b$, does the value of the derivative at some point between a and b have to match the slope of the secant line connecting the points $(a, f(a))$ and $(b, f(b))$? Or equivalently, does the equation $f'(c) = \frac{f(b)-f(a)}{b-a}$ have to hold for some $a < c < b$?

The following Activity will explore what properties f must possess in order for the equation $f'(c) = \frac{f(b)-f(a)}{b-a}$ have to hold for some $a < c < b$.



(a) $f_1(x) = 1/x^2$



(b) $f_2(x) = |x|$

Figure 3.34: The functions $f_1(x) = 1/x^2$ and $f_2(x) = |x|$ used in Activity 3.5–1

Activity 3.5–1

Consider functions

$$f_1(x) = \frac{1}{x^2} \quad \text{and} \quad f_2(x) = |x|$$

with $a = -1$ and $b = 1$ as shown in Figure 3.34-(a) and -(b), respectively.

(a) For $f_1(x)$, find the slope of the secant line connecting the points $(a, f(a))$ and $(b, f(b))$.

(b) Find a value c in (a, b) such that $f_1'(c)$ equals the slope of the secant line connecting the points $(a, f(a))$ and $(b, f(b))$. Was it possible to find c? If not, what went wrong?

(c) For $f_2(x)$, find the slope of the secant line connecting the points $(a, f(a))$ and $(b, f(b))$.

(d) Find a value c in (a, b) such that $f_2'(c)$ equals the slope of the secant line connecting the points $(a, f(a))$ and $(b, f(b))$. Was it possible to find c? If not, what went wrong?

So what went “wrong”? It may not be surprising to find that the discontinuity of f_1 and the corner of f_2 play a role. If our functions had been continuous and differentiable, would we have been able to find that special value c? This is our motivation for the following theorem.

The Mean Value Theorem of Differentiation

Let $y = f(x)$ be a continuous function on the closed interval $[a, b]$ and differentiable on the open interval (a, b). There exists a value c, $a < c < b$, such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

That is, there is a value c in (a, b) where the instantaneous rate of change of f at c is equal to the average rate of change of f on $[a, b]$.

Note that the reasons that the functions in Activity3.5–1 fail are indeed that f_1 has a discontinuity on the interval $[-1, 1]$ and

f_2 is not differentiable at the origin.

We will give a proof of the Mean Value Theorem below. To do so, we use another theorem, called Rolle's Theorem, stated here.

Rolle's Theorem

Let f be continuous on $[a,b]$ and differentiable on (a,b), where $f(a) = f(b)$. There is some c in (a,b) such that $f'(c) = 0$.

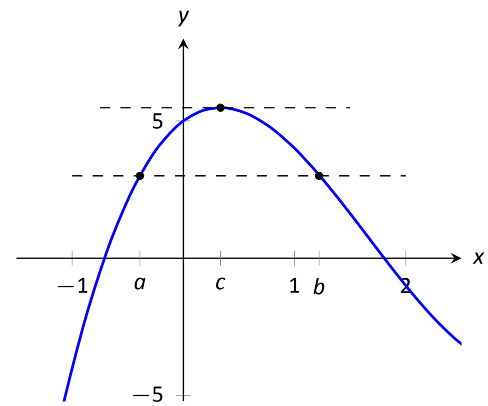


Figure 3.35: A graph of $f(x) = x^3 - 5x^2 + 3x + 5$, where $f(a) = f(b)$. Note the existence of c, where $a < c < b$, where $f'(c) = 0$.

Consider Figure 3.35 where the graph of a function f is given, where $f(a) = f(b)$. It should make intuitive sense that if f is differentiable (and hence, continuous) that there would be a value c in (a,b) where $f'(c) = 0$; that is, there would be a local maximum or minimum of f in (a,b). Rolle's Theorem guarantees at least one; there may be more.

Rolle's Theorem is really just a special case of the Mean Value Theorem. If $f(a) = f(b)$, then the *average* rate of change on (a,b) is 0, and the theorem guarantees some c where $f'(c) = 0$. We will prove Rolle's Theorem, then use it to prove the Mean Value Theorem.

Proof: (Rolle's Theorem) Let f be differentiable on (a,b) where $f(a) = f(b)$. We consider two cases.

Case 1: Consider the case when f is constant on $[a,b]$; that is, $f(x) = f(a) = f(b)$ for all x in $[a,b]$. Then $f'(x) = 0$ for all x in $[a,b]$, showing there is at least one value c in (a,b) where $f'(c) = 0$.

Case 2: Now assume that f is not constant on $[a,b]$. The Extreme Value Theorem guarantees that f has a maximal and minimal value on $[a,b]$, found either at the endpoints or at a critical value in (a,b). Since $f(a) = f(b)$ and f is not constant, it is clear that the maximum and minimum cannot *both* be found at the endpoints. Assume, without loss of generality, that the maximum of f is not found at the endpoints. Therefore there is a c in (a,b) such that $f(c)$ is the maximum value of f. Thus c must be a critical number of f; since f is differentiable, we have that $f'(c) = 0$, completing the proof of the theorem. □

Proof: (Mean Value Theorem) Define the function

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a}x.$$

We know g is differentiable on (a,b) and continuous on $[a,b]$ since f is. We can show $g(a) = g(b)$ (it is actually easier to show

$g(b) - g(a) = 0$, which suffices). We can then apply Rolle's theorem to guarantee the existence of $c \in (a, b)$ such that $g'(c) = 0$. But note that

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a};$$

hence

$$f'(c) = \frac{f(b) - f(a)}{b - a},$$

which is what we sought to prove. □

Going back to the very beginning of the section, we see that the only assumption we would need about our distance function $f(t)$ is that it be continuous and differentiable for t from 0 to 2 hours (both reasonable assumptions). By the Mean Value Theorem, we are guaranteed a time during the trip where our instantaneous speed is 60 mph. This fact is used in practice. Some law enforcement agencies monitor traffic speeds while in aircraft. They do not measure speed with radar, but rather by timing individual cars as they pass over lines painted on the highway whose distances apart are known. The officer is able to measure the *average* speed of a car between the painted lines; if that average speed is greater than the posted speed limit, the officer is assured that the driver exceeded the speed limit at some time.

Note that the Mean Value Theorem is an *existence* theorem. It states that a special value c *exists*, but it does not give any indication about how to find it. To do so, we must be able to solve equations, which sometimes can be very difficult. The following example demonstrates the process for finding a special value c.

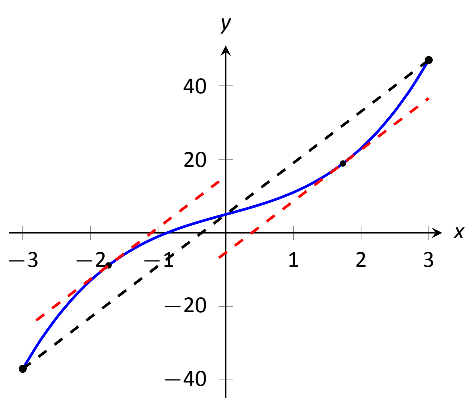


Figure 3.36: Demonstrating the Mean Value Theorem in Example 1

Example 1

Consider $f(x) = x^3 + 5x + 5$ on $[-3, 3]$. Find c in $[-3, 3]$ that satisfies the Mean Value Theorem.

Solution. The average rate of change of f on $[-3, 3]$ is:

$$\frac{f(3) - f(-3)}{3 - (-3)} = \frac{84}{6} = 14.$$

We want to find c such that $f'(c) = 14$. We find $f'(x) = 3x^2 + 5$. We set this equal to 14 and solve for x.

$$\begin{aligned} f'(x) &= 14 \\ 3x^2 + 5 &= 14 \\ x^2 &= 3 \\ x &= \pm\sqrt{3} \approx \pm 1.732 \end{aligned}$$

We have found 2 values c in $[-3,3]$ where the instantaneous rate of change is equal to the average rate of change; the Mean Value Theorem guaranteed at least one. In Figure 3.36 f is graphed with a dashed line representing the average rate of change; the lines tangent to f at $x = \pm\sqrt{3}$ are also given. Note how these lines are parallel (i.e., have the same slope) as the dashed line.

The Mean Value Theorem has practical use (for instance, the speed monitoring application mentioned before) in the sciences and other applications. However, it turns out that when we need the Mean Value Theorem in mathematics, existence is all we need, and it is mostly used to advance other theory.

Summary

In this section, we encountered the following important ideas:

- If f is a continuous function on the closed interval $[a,b]$ and differentiable on the open interval (a,b). There exists a value c, $a < c < b$, such that $f'(c) = \frac{f(b)-f(a)}{b-a}$, i.e. the instantaneous rate of change of f at c is equal to the average rate of change of f on $[a,b]$.

Exercises

Terms and Concepts

1) Explain in your own words what the Mean Value Theorem states.
2) Explain in your own words what Rolle's Theorem states.

Problems

In exercises 3–10, a function $f(x)$ and interval $[a,b]$ are given. Check if Rolle's Theorem can be applied, and if so, find c in $[a,b]$ such that $f'(c)=0$.

3) $f(x)=6$ on $[-1,1]$.
4) $f(x)=6x$ on $[-1,1]$.
5) $f(x)=x^2+x-6$ on $[-3,2]$.
6) $f(x)=x^2+x-2$ on $[-3,2]$.
7) $f(x)=x^2+x$ on $[-2,2]$.
8) $f(x)=\sin x$ on $[\pi/6,5\pi/6]$.
9) $f(x)=\cos x$ on $[0,\pi]$.
10) $f(x)=\dfrac{1}{x^2-2x+1}$ on $[0,2]$.

In exercises 11-20, a function $f(x)$ and interval $[a,b]$ are given Check if the Mean Value Theorem can be applied, and if so, find a value c guaranteed by the Mean Value Theorem.

11) $f(x)=x^2+3x-1$ on $[-2,2]$.
12) $f(x)=5x^2-6x+8$ on $[0,5]$.
13) $f(x)=\sqrt{9-x^2}$ on $[0,3]$.
14) $f(x)=\sqrt{25-x}$ on $[0,9]$.
15) $f(x)=\ln x$ on $[1,5]$.
16) $f(x)=\tan x$ on $[-\pi/4,\pi/4]$.
17) $f(x)=x^3-2x^2+x+1$ on $[-2,2]$.
18) $f(x)=2x^3-5x^2+6x+1$ on $[-5,2]$.
19) $f(x)=\sin^{-1} x$ on $[-1,1]$.
20) $f(x)=\dfrac{x^2-9}{x^2-1}$ on $[0,2]$.

3.6 *The Tangent Line Approximation*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- What is the formula for the general tangent line approximation to a differentiable function $y = f(x)$ at the point $(a, f(a))$?
- What is the principle of local linearity and what is the local linearization of a differentiable function f at a point $(a, f(a))$?
- How does knowing just the tangent line approximation tell us information about the behavior of the original function itself near the point of approximation? How does knowing the second derivative's value at this point provide us additional knowledge of the original function's behavior?

Introduction

Among all functions, linear functions are simplest. One of the powerful consequences of a function $y = f(x)$ being differentiable at a point $(a, f(a))$ is that, up close, the function $y = f(x)$ is locally linear and looks like its tangent line at that point. In certain circumstances, this allows us to approximate the original function f with a simpler function L that is linear: this can be advantageous when we have limited information about f or when f is computationally or algebraically complicated. We will explore all of these situations in what follows.

It is essential to recall that when f is differentiable at $x = a$, the value of $f'(a)$ provides the slope of the tangent line to $y = f(x)$ at the point $(a, f(a))$. By knowing both a point on the line and the slope of the line we are thus able to find the equation of the tangent line. Preview Activity 3.6 will refresh these concepts through a key example and set the stage for further study.

Preview Activity 3.6

Consider the function $y = g(x) = -x^2 + 3x + 2$.

(a) Use the limit definition of the derivative to compute a formula for $y = g'(x)$.

(b) Determine the slope of the tangent line to $y = g(x)$ at the value $x = 2$.

(c) Compute $g(2)$.

(d) Find an equation for the tangent line to $y = g(x)$ at the point $(2, g(2))$. Write your result in point-slope form. Recall that a line with slope m that passes through (x_0, y_0) has equation $y - y_0 = m(x - x_0)$, and this is the *point-slope form* of the equation.

(e) On the axes provided in Figure 3.37, sketch an accurate, labeled graph of $y = g(x)$ along with its tangent line at the point $(2, g(2))$.

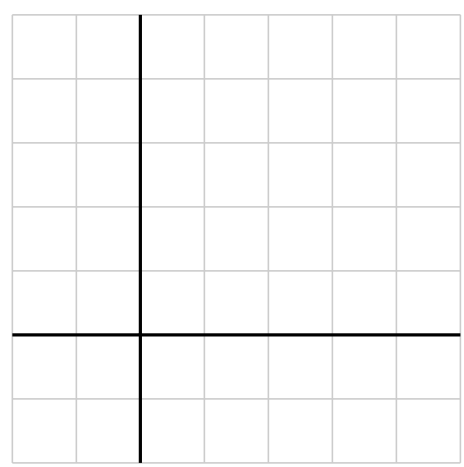


Figure 3.37: Axes for plotting $y = g(x)$ and its tangent line to the point $(2, g(2))$.

The tangent line

Given a function f that is differentiable at $x = a$, we know that we can determine the slope of the tangent line to $y = f(x)$ at $(a, f(a))$ by computing $f'(a)$. The resulting tangent line through $(a, f(a))$ with slope $m = f'(a)$ has its equation in point-slope form given by

$$y - f(a) = f'(a)(x - a),$$

which we can also express as $y = f'(a)(x - a) + f(a)$. Note well: there is a major difference between $f(a)$ and $f(x)$ in this context. The former is a constant that results from using the given fixed value of a, while the latter is the general expression for the rule that defines the function. The same is true for $f'(a)$ and $f'(x)$: we must carefully distinguish between these expressions. Each time we find the tangent line, we need to evaluate the function and its derivative at a fixed a-value.

In Figure 3.38, we see a labeled plot of the graph of a function f and its tangent line at the point $(a, f(a))$. Notice how when we zoom in we see the local linearity of f more clearly highlighted as the function and its tangent line are nearly indistinguishable up close. This can also be seen dynamically in the java applet at `http://gvsu.edu/s/6J`.

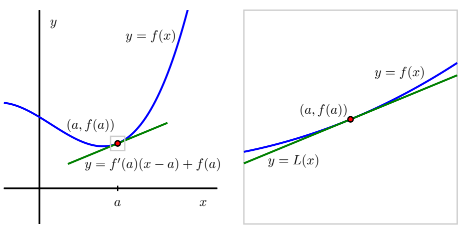


Figure 3.38: A function $y = f(x)$ and its tangent line at the point $(a, f(a))$: at left, from a distance, and at right, up close. At right, we label the tangent line function by $y = L(x)$ and observe that for x near a, $f(x) \approx L(x)$.

The local linearization

A slight change in perspective and notation will enable us to be more precise in discussing how the tangent line to $y = f(x)$ at $(a, f(a))$ approximates f near $x = a$. Taking the equation for the tangent line and solving for y, we observe that the tangent line is given by

$$y = f'(a)(x - a) + f(a)$$

and moreover that this line is itself a function of x. Replacing the variable y with the expression $L(x)$, we call

$$L(x) = f'(a)(x - a) + f(a)$$

the *local linearization of* f at the point $(a, f(a))$. In this notation, it is particularly important to observe that $L(x)$ is nothing more than a new name for the tangent line, and that for x close to a, we have that $f(x) \approx L(x)$.

Say, for example, that we know that a function $y = f(x)$ has its tangent line approximation given by $L(x) = 3 - 2(x - 1)$ at the point $(1, 3)$, but we do not know anything else about the function f. If we are interested in estimating a value of $f(x)$ for x near 1, such as $f(1.2)$, we can use the fact that $f(1.2) \approx L(1.2)$ and hence

$$f(1.2) \approx L(1.2) = 3 - 2(1.2 - 1) = 3 - 2(0.2) = 2.6.$$

Again, much of the new perspective here is only in notation since $y = L(x)$ is simply a new name for the tangent line function. In light of this new notation and our observations above, we note that since $L(x) = f(a) + f'(a)(x - a)$ and $L(x) \approx f(x)$ for x near a, it also follows that we can write

$$f(x) \approx f(a) + f'(a)(x - a) \text{ for } x \text{ near } a.$$

The next activity explores some additional important properties of the local linearization $y = L(x)$ to a function f at given a-value.

Activity 3.6–1

Suppose it is known that for a given differentiable function $y = g(x)$, its local linearization at the point where $a = -1$ is given by $L(x) = -2 + 3(x + 1)$.

(a) Compute the values of $L(-1)$ and $L'(-1)$.

(b) What must be the values of $g(-1)$ and $g'(-1)$? Why?

(c) Do you expect the value of $g(-1.03)$ to be greater than or less than the value of $g(-1)$? Why?

(d) Use the local linearization to estimate the value of $g(-1.03)$.

(e) Suppose that you also know that $g''(-1) = 2$. What does this tell you about the graph of $y = g(x)$ at $a = -1$?

(f) For x near -1, sketch the graph of the local linearization $y = L(x)$ as well as a possible graph of $y = g(x)$ on the axes provided in Figure 3.39.

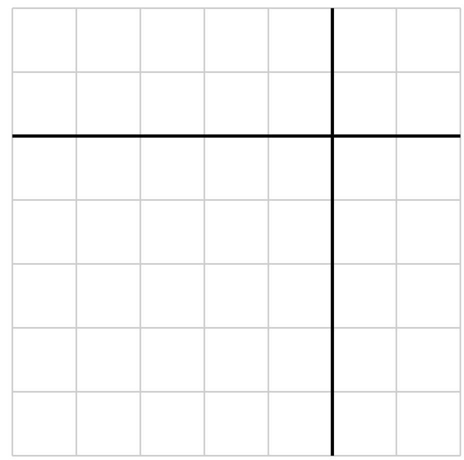


Figure 3.39: Axes for plotting $y = L(x)$ and $y = g(x)$.

As we saw in the example provided by Activity **??**, the local linearization $y = L(x)$ is a linear function that shares two important values with the function $y = f(x)$ that it is derived from. In particular, observe that since $L(x) = f(a) + f'(a)(x - a)$, it follows that $L(a) = f(a)$. In addition, since L is a linear function, its derivative is its slope. Hence, $L'(x) = f'(a)$ for every value of x, and specifically $L'(a) = f'(a)$. Therefore, we see that L is a linear function that has both the same value and the same slope as the function f at the point $(a, f(a))$.

In situations where we know the linear approximation $y = L(x)$, we therefore know the original function's value and slope at the point of tangency. What remains unknown, however, is the shape of the function f at the point of tangency. There are essentially four possibilities, as enumerated in Figure 3.40.

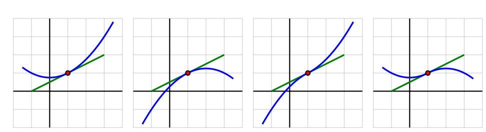


Figure 3.40: Four possible graphs for a nonlinear differentiable function and how it can be situated relative to its tangent line at a point.

These stem from the fact that there are three options for the value of the second derivative: either $f''(a) < 0$, $f''(a) = 0$, or $f''(a) > 0$. If $f''(a) > 0$, then we know the graph of f is concave up, and we see the first possibility on the left, where the tangent line lies entirely below the curve. If $f''(a) < 0$, then we find ourselves in the second situation (from left) where f is concave down and the tangent line lies above the curve. In the situation where $f''(a) = 0$ and f'' changes sign at $x = a$, the concavity of the graph will change, and we will see either the third or fourth option[5]. A fifth option (that is not very interesting) can occur, which is where the function f is linear, and so $f(x) = L(x)$ for all values of x.

[5] It is possible to have $f''(a) = 0$ and have f'' not change sign at $x = a$, in which case the graph will look like one of the first two options.

The plots in Figure 3.40 highlight yet another important thing that we can learn from the concavity of the graph near the point of tangency: whether the tangent line lies above or below the curve itself. This is key because it tells us whether or not the tangent line approximation's values will be too large or too small in comparison to the true value of f. For instance, in the first situation in the leftmost plot in Figure 3.40 where $f''(a) > 0$, since the tangent line falls below the curve, we know that $L(x) \leq f(x)$ for all values of x near a.

We explore these ideas further in the following activity.

Activity 3.6–2

This activity concerns a function $f(x)$ about which the following information is known:

- f is a differentiable function defined at every real number x
- $f(2) = -1$
- $y = f'(x)$ has its graph given in Figure 3.41

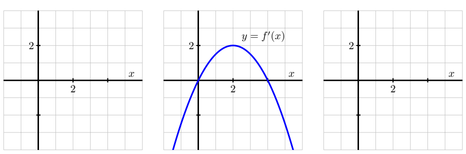


Figure 3.41: At center, a graph of $y = f'(x)$; at left, axes for plotting $y = f(x)$; at right, axes for plotting $y = f''(x)$.

Your task is to determine as much information as possible about f (especially near the value $a = 2$) by responding to the questions below.

(a) Find a formula for the tangent line approximation, $L(x)$, to f at the point $(2, -1)$.

(b) Use the tangent line approximation to estimate the value of $f(2.07)$. Show your work carefully and clearly.

(c) Sketch a graph of $y = f''(x)$ on the righthand grid in Figure 3.41; label it appropriately.

(d) Is the slope of the tangent line to $y = f(x)$ increasing, decreasing, or neither when $x = 2$? Explain.

(e) Sketch a possible graph of $y = f(x)$ near $x = 2$ on the lefthand grid in Figure 3.41. Include a sketch of $y = L(x)$ (found in part (a)). Explain how you know the graph of $y = f(x)$ looks like you have drawn it.

(f) Does your estimate in (b) over- or under-estimate the true value of $f(2)$? Why?

The idea that a differentiable function looks linear and can be well-approximated by a linear function is an important one that finds wide application in calculus. For example, by approximating a function with its local linearization, it is possible to develop an effective algorithm to estimate the zeroes of a function. Local linearity also helps us to make further sense of cer-

tain challenging limits. For instance, we have seen that a limit such as
$$\lim_{x\to 0}\frac{\sin(x)}{x}$$
is indeterminate because both its numerator and denominator tend to 0. While there is no algebra that we can do to simplify $\frac{\sin(x)}{x}$, it is straightforward to show that the linearization of $f(x) = \sin(x)$ at the point $(0,0)$ is given by $L(x) = x$. Hence, for values of x near 0, $\sin(x) \approx x$. As such, for values of x near 0,
$$\frac{\sin(x)}{x} \approx \frac{x}{x} = 1,$$
which makes plausible the fact that
$$\lim_{x\to 0}\frac{\sin(x)}{x} = 1.$$
These ideas and other applications of local linearity will be explored later on in our work.

Differentials

Recall that the derivative of a function f can be used to find the slopes of lines tangent to the graph of f. At $x = a$, the tangent line to the graph of f has equation
$$y = f(a) + f'(a)(x-a).$$
The tangent line can be used to find good approximations of $f(x)$ for values of x near a.

We now generalize this concept. Given $f(x)$ and an x–value a, the tangent line is $L(x) = f(a) + f'(a)(x-a)$. Clearly, $L(a) = f(a)$. Let Δx be a small number, representing a small change in x value. We assert that:
$$f(a+\Delta x) \approx L(a+\Delta x),$$
since the tangent line to a function approximates well the values of that function near $x = a$.

As the x value changes from a to $a+\Delta x$, the y value of f changes from $f(a)$ to $f(a+\Delta x)$. We call this change of y value Δy. That is:
$$\Delta y = f(a+\Delta x) - f(a).$$
Replacing $f(a+\Delta x)$ with its tangent line approximation, we have
$$\begin{aligned}\Delta y &\approx L(a+\Delta x) - f(a)\\ &= f'(a)\big((a+\Delta x)-a\big) + f(a) - f(a)\\ &= f'(a)\Delta x \qquad (3.4)\end{aligned}$$

This final equation is important; we'll come back to it in a moment.

We introduce two new variables, dx and dy in the context of a formal definition.

Differentials of x and y

Let $y = f(x)$ be differentiable. The **differential of** x, denoted dx, is any nonzero real number (usually taken to be a small number). The **differential of** y, denoted dy, is

$$dy = f'(x)dx.$$

It is helpful to organize our new concepts and notations in one place.

Differential Notation

Let $y = f(x)$ be a differentiable function.

(a) Δx represents a small, nonzero change in x value.

(b) dx represents a small, nonzero change in x value (i.e., $\Delta x = dx$).

(c) Δy is the change in y value as x changes by Δx; hence

$$\Delta y = f(x + \Delta x) - f(x).$$

(d) $dy = f'(x)dx$ which, by Equation (3.4), is an *approximation* of the change in y value as x changes by Δx; $dy \approx \Delta y$.

What is the value of differentials? Like many mathematical concepts, differentials provide both practical and theoretical benefits. We explore both here.

Example 1

Consider $f(x) = x^2$. Knowing $f(3) = 9$, approximate $f(3.1)$.

Solution. The x value is changing from $x = 3$ to $x = 3.1$; therefore, we see that $dx = 0.1$. If we know how much the y value changes from $f(3)$

to $f(3.1)$ (i.e., if we know Δy), we will know exactly what $f(3.1)$ is (since we already know $f(3)$). We can approximate Δy with dy.

$$\begin{aligned}\Delta y &\approx dy \\ &= f'(3)dx \\ &= 2 \cdot 3 \cdot 0.1 = 0.6.\end{aligned}$$

We expect the y value to change by about 0.6, so we approximate $f(3.1) \approx 9.6$.

We leave it to the reader to verify this, but the preceding discussion links the differential to the tangent line of $f(x)$ at $x = 3$. One can verify that the tangent line, evaluated at $x = 3.1$, also gives $y = 9.6$.

Of course, it is easy to compute the actual answer (by hand or with a calculator): $3.1^2 = 9.61$. (Before we get too cynical and say "Then why bother?", note our approximation is *really* good!)

So why bother?

In "most" real life situations, we do not know the function that describes a particular behavior. Instead, we can only take measurements of how things change – measurements of the derivative.

Imagine water flowing down a winding channel. It is easy to measure the speed and direction (i.e., the *velocity*) of water at any location. It is very hard to create a function that describes the overall flow, hence it is hard to predict where a floating object placed at the beginning of the channel will end up. However, we can *approximate* the path of an object using differentials. Over small intervals, the path taken by a floating object is essentially linear. Differentials allow us to approximate the true path by piecing together lots of short, linear paths. This technique is called Euler's Method, studied in introductory Differential Equations courses.

We use differentials once more to approximate the value of a function. Even though calculators are very accessible, it is neat to see how these techniques can sometimes be used to easily compute something that looks rather hard.

Example 2

Approximate $\sqrt{4.5}$.

Solution. We expect $\sqrt{4.5} \approx 2$, yet we can do better. Let $f(x) = \sqrt{x}$, and let $c = 4$. Thus $f(4) = 2$. We can compute $f'(x) = 1/(2\sqrt{x})$, so $f'(4) = 1/4$.

We approximate the difference between $f(4.5)$ and $f(4)$ using differentials, with $dx = 0.5$:

$$f(4.5) - f(4) = \Delta y \approx dy = f'(4) \cdot dx = 1/4 \cdot 1/2 = 1/8 = 0.125.$$

The approximate change in f from $x = 4$ to $x = 4.5$ is 0.125, so we approximate $\sqrt{4.5} \approx 2.125$.

Differentials are important when we discuss *integration*. When we study that topic, we will use notation such as

$$\int f(x)\,dx$$

quite often. While we don't discuss here what all of that notation means, note the existence of the differential dx. Proper handling of *integrals* comes with proper handling of differentials.

In light of that, we practice finding differentials in general.

Example 3

In each of the following, find the differential dy.

1) $y = \sin x$ 2) $y = e^x(x^2+2)$ 3) $y = \sqrt{x^2+3x-1}$

Solution.

1) $y = \sin x$: As $f(x) = \sin x$, $f'(x) = \cos x$. Thus

$$dy = \cos(x)dx.$$

2) $y = e^x(x^2+2)$: Let $f(x) = e^x(x^2+2)$. We need $f'(x)$, requiring the Product Rule.
We have $f'(x) = e^x(x^2+2) + 2xe^x$, so

$$dy = (e^x(x^2+2) + 2xe^x)dx.$$

3) $y = \sqrt{x^2+3x-1}$: Let $f(x) = \sqrt{x^2+3x-1}$; we need $f'(x)$, requiring the Chain Rule.
We have $f'(x) = \dfrac{1}{2}(x^2+3x-1)^{-\frac{1}{2}}(2x+3) = \dfrac{2x+3}{2\sqrt{x^2+3x-1}}$. Thus

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

Finding the differential dy of $y = f(x)$ is really no harder than finding the derivative of f; we just *multiply* $f'(x)$ by dx. It is important to remember that we are not simply adding the symbol "dx" at the end.

We have seen a practical use of differentials as they offer a good method of making certain approximations. Another use is *error propagation*. Suppose a length is measured to be x, although the actual value is $x + \Delta x$ (where we hope Δx is small). This measurement of x may be used to compute some other value; we can think of this as $f(x)$ for some function f. As the true length is $x + \Delta x$, one really should have computed $f(x + \Delta x)$.

The difference between $f(x)$ and $f(x+\Delta x)$ is the propagated error.

How close are $f(x)$ and $f(x+\Delta x)$? This is a difference in "y" values;

$$f(x+\Delta x) - f(x) = \Delta y \approx dy.$$

We can approximate the propagated error using differentials.

Example 4

A steel ball bearing is to be manufactured with a diameter of 2cm. The manufacturing process has a tolerance of ± 0.1mm in the diameter. Given that the density of steel is about 7.85g/cm^3, estimate the propagated error in the mass of the ball bearing.

Solution. The mass of a ball bearing is found using the equation mass = volume $\times$ density. In this situation the mass function is a product of the radius of the ball bearing, hence it is $m = 7.85\frac{4}{3}\pi r^3$. The differential of the mass is

$$dm = 31.4\pi r^2 dr.$$

The radius is to be 1cm; the manufacturing tolerance in the radius is ± 0.05mm, or ± 0.005cm. The propagated error is approximately:

$$\begin{aligned}\Delta m &\approx dm \\ &= 31.4\pi(1)^2(\pm 0.005) \\ &= \pm 0.493\text{g}\end{aligned}$$

Is this error significant? It certainly depends on the application, but we can get an idea by computing the *relative error*. The ratio between amount of error to the total mass is

$$\begin{aligned}\frac{dm}{m} &= \pm\frac{0.493}{7.85\frac{4}{3}\pi} \\ &= \pm\frac{0.493}{32.88} \\ &= \pm 0.015,\end{aligned}$$

or $\pm 1.5\%$.

We leave it to the reader to confirm this, but if the diameter of the ball was supposed to be 10cm, the same manufacturing tolerance would give a propagated error in mass of ± 12.33g, which corresponds to a *percent error* of $\pm 0.188\%$. While the amount of error is much greater ($12.33 > 0.493$), the percent error is much lower.

Summary

In this section, we encountered the following important ideas:

- The tangent line to a differentiable function $y = f(x)$ at the point $(a, f(a))$ is given in point-slope form by the equation
$$y - f(a) = f'(a)(x - a).$$
- The principle of local linearity tells us that if we zoom in on a point where a function $y = f(x)$ is differentiable, the function should become indistinguishable from its tangent line. That is, a differentiable function looks linear when viewed up close. We rename the tangent line to be the function $y = L(x)$ where $L(x) = f(a) + f'(a)(x - a)$ and note that $f(x) \approx L(x)$ for all x near $x = a$.
- If we know the tangent line approximation $L(x) = f(a) + f'(a)(x - a)$, then because $L(a) = f(a)$ and $L'(a) = f'(a)$, we also know both the value and the derivative of the function $y = f(x)$ at the point where $x = a$. In other words, the linear approximation tells us the height and slope of the original function. If, in addition, we know the value of $f''(a)$, we then know whether the tangent line lies above or below the graph of $y = f(x)$ depending on the concavity of f.

Exercises

Terms and Concepts

1) T/F: Given a differentiable function $y = f(x)$, we are generally free to choose a value for dx, which then determines the value of dy.

2) T/F: The symbols "dx" and "Δx" represent the same concept.

3) T/F: The symbols "dy" and "Δy" represent the same concept.

4) T/F: Differentials are important in the study of integration.

5) Fill in the blanks: The Quotient Rule is applied to $\dfrac{f(x)}{g(x)}$ when taking ________; l'Hôpital's Rule is applied when taking certain ________.

6) How are differentials and tangent lines related?

Problems

In exercises 7–11, use differentials to approximate the given value by hand.

7) 2.05^2

8) 5.1^3

9) $\sqrt{16.5}$

10) $\sqrt[3]{63}$

11) $\sin 3$

12) A certain function $y = p(x)$ has its local linearization at $a = 3$ given by $L(x) = -2x + 5$.

   (a) What are the values of $p(3)$ and $p'(3)$? Why?

   (b) Estimate the value of $p(2.79)$.

   (c) Suppose that $p''(3) = 0$ and you know that $p''(x) < 0$ for $x < 3$. Is your estimate in (b) too large or too small?

   (d) Suppose that $p''(x) > 0$ for $x > 3$. Use this fact and the additional information above to sketch an accurate graph of $y = p(x)$ near $x = 3$. Include a sketch of $y = L(x)$ in your work.

13) A potato is placed in an oven, and the potato's temperature F (in degrees Fahrenheit) at various points in time is taken and recorded in the following table. Time t is measured in minutes.

t	0	15	30	45	60	75	90
$F(t)$	70	180.5	251	296	324.5	342.8	354.5

   (a) Use a central difference to estimate $F'(60)$. Use this estimate as needed in subsequent questions.

   (b) Find the local linearization $y = L(t)$ to the function $y = F(t)$ at the point where $a = 60$.

   (c) Determine an estimate for $F(63)$ by employing the local linearization.

   (d) Do you think your estimate in (c) is too large or too small? Why?

14) An object moving along a straight line path has a differentiable position function $y = s(t)$. It is known that at time $t = 9$ seconds, the object's position is $s = 4$ feet (measured from its starting point at $t = 0$). Furthermore, the object's instantaneous velocity at $t = 9$ is -1.2 feet per second, and its acceleration at the same instant is 0.08 feet per second per second.

   (a) Use local linearity to estimate the position of the object at $t = 9.34$.

   (b) Is your estimate likely too large or too small? Why?

   (c) In everyday language, describe the behavior of the moving object at $t = 9$. Is it moving toward its starting point or away from it? Is its velocity increasing or decreasing?

15) For a certain function f, its derivative is known to be $f'(x) = (x-1)e^{-x^2}$. Note that you do not know a formula for $y = f(x)$.

   (a) At what x-value(s) is $f'(x) = 0$? Justify your answer algebraically, but include a graph of f' to support your conclusion.

   (b) Reasoning graphically, for what intervals of x-values is $f''(x) > 0$? What does this tell you about the behavior of the original function f? Explain.

   (c) Assuming that $f(2) = -3$, estimate the value of $f(1.88)$ by finding and using the tangent line approximation to f at $x = 2$. Is your estimate larger or smaller than the true value of $f(1.88)$? Justify your answer.

In exercises 16–25, compute the differential.

16) $y = x^2 + 3x - 5$

17) $y = x^7 - x^5$

18) $y = \dfrac{1}{4x^2}$

19) $y = (2x + \sin x)^2$

20) $y = x^2 e^{3x}$

21) $y = \dfrac{4}{x^4}$

22) $y = \dfrac{2x}{\tan x + 1}$

23) $y = \ln(5x)$

24) $y = e^x \sin x$

25) $y = \cos(\sin x)$

26) A set of plastic spheres are to be made with a diameter of 1cm. If the manufacturing process is accurate to 1mm, what is the propagated error in volume of the spheres?

27) The distance, in feet, a stone drops in t seconds is given by $d(t) = 16t^2$. The depth of a hole is to be approximated by dropping a rock and listening for it to hit the bottom. What is the propagated error if the time measurement is accurate to $2/10^{\text{ths}}$ of a second and the measured time is:

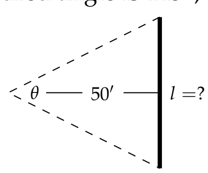
(a) 2 seconds?

(b) 5 seconds?

28) What is the propagated error in the measurement of the cross sectional area of a circular log if the diameter is measured at $15''$, accurate to $1/4''$?

29) A wall is to be painted that is $8'$ high and is measured to be $10'$, $7''$ long. Find the propagated error in the measurement of the wall's surface area if the measurement is accurate to $1/2''$.

30) The length l of a long wall is to be calculated by measuring the angle θ shown in the diagram (not to scale). Assume the formed triangle is an isosceles triangle. The measured angle is $143°$, accurate to $1°$.

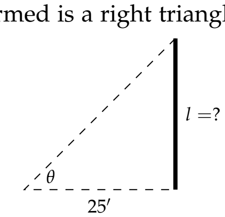


(a) What is the measured length of the wall?

(b) What is the propagated error?

(c) What is the percent error?

31) The length l of a long wall is to be approximated. The angle θ, as shown in the diagram (not to scale), is measured to be $85.2°$, accurate to $1°$. Assume that the triangle formed is a right triangle.



(a) What is the measured length l of the wall?

(b) What is the propagated error?

(c) What is the percent error?

32) Answer the questions of Exercise 31), but with a measured angle of $71.5°$, accurate to $1°$, measured from a point $100'$ from the wall.

33) The length of the walls in Exercises 31) – 30) are essentially the same. Which setup gives the most accurate result?

34) Consider the setup in Exercises 30). This time, assume the angle measurement of $143°$ is exact but the measured $50'$ from the wall is accurate to $6''$. What is the approximate percent error?

3.7 *Using Derivatives to Evaluate Limits*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- How can derivatives be used to help us evaluate indeterminate limits of the form $\frac{0}{0}$?
- What does it mean to say that $\lim_{x\to\infty} f(x) = L$ and $\lim_{x\to a} f(x) = \infty$?
- How can derivatives assist us in evaluating indeterminate limits of the form $\frac{\infty}{\infty}$?

Introduction

Because differential calculus is based on the definition of the derivative, and the definition of the derivative involves a limit, there is a sense in which all of calculus rests on limits. In addition, the limit involved in the limit definition of the derivative is one that always generates an indeterminate form of $\frac{0}{0}$. If f is a differentiable function for which $f'(x)$ exists, then when we consider

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

it follows that not only does $h \to 0$ in the denominator, but also $(f(x+h) - f(x)) \to 0$ in the numerator, since f is continuous. Thus, the fundamental form of the limit involved in the definition of $f'(x)$ is $\frac{0}{0}$. Remember, saying a limit has an indeterminate form only means that we don't yet know its value and have more work to do: indeed, limits of the form $\frac{0}{0}$ can take on any value, as is evidenced by evaluating $f'(x)$ for varying values of x for a function such as $f'(x) = x^2$.

Of course, we have learned many different techniques for evaluating the limits that result from the derivative definition, and including a large number of shortcut rules that enable us to evaluate these limits quickly and easily. In this section, we turn the situation upside-down: rather than using limits to evaluate derivatives, we explore how to use derivatives to evaluate certain limits. This topic will combine several different ideas, including limits, derivative shortcuts, local linearity, and the tangent line approximation.

Preview Activity 3.7

Let h be the function given by $h(x) = \dfrac{x^5 + x - 2}{x^2 - 1}$.

(a) What is the domain of h?

(b) Explain why $\lim_{x \to 1} \dfrac{x^5 + x - 2}{x^2 - 1}$ results in an indeterminate form.

(c) Next we will investigate the behavior of both the numerator and denominator of h near the point where $x = 1$. Let $f(x) = x^5 + x - 2$ and $g(x) = x^2 - 1$. Find the local linearizations of f and g at $a = 1$, and call these functions $L_f(x)$ and $L_g(x)$, respectively.

(d) Explain why $h(x) \approx \dfrac{L_f(x)}{L_g(x)}$ for x near $a = 1$.

(e) Using your work from (c) and (d), evaluate

$$\lim_{x \to 1} \frac{L_f(x)}{L_g(x)}.$$

What do you think your result tells us about $\lim_{x \to 1} h(x)$?

(f) Investigate the function $h(x)$ graphically and numerically near $x = 1$. What do you think is the value of $\lim_{x \to 1} h(x)$?

Using derivatives to evaluate indeterminate limits of the form $\frac{0}{0}$.

The fundamental idea of Preview Activity 2.8 – that we can evaluate an indeterminate limit of the form $\frac{0}{0}$ by replacing each of the numerator and denominator with their local linearizations at the point of interest – can be generalized in a way that enables us to easily evaluate a wide range of limits. We begin by assuming that we have a function $h(x)$ that can be written in the form $h(x) = \frac{f(x)}{g(x)}$ where f and g are both differentiable at $x = a$ and for which $f(a) = g(a) = 0$. We are interested in finding a way to evaluate the indeterminate limit given by $\lim_{x \to a} h(x)$.

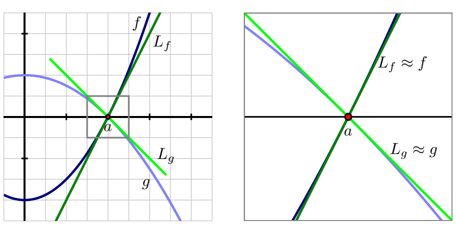


Figure 3.42: At left, the graphs of f and g near the value a, along with their tangent line approximations L_f and L_g at $x = a$. At right, zooming in on the point a and the four graphs.

In Figure 3.42, we see a visual representation of the situation involving such functions f and g. In particular, we see that both f and g have an x-intercept at the point where $x = a$. In addition, since each function is differentiable, each is locally linear, and we can find their respective tangent line approximations L_f and L_g at $x = a$, which are also shown in the figure. Since we are interested in the limit of $\frac{f(x)}{g(x)}$ as $x \to a$, the individual behaviors of $f(x)$ and $g(x)$ as $x \to a$ are key to understand. Here, we take advantage of the fact that each function and its tangent line approximation become indistinguishable as $x \to a$.

First, let's reall that $L_f(x) = f'(a)(x - a) + f(a)$ and $L_g(x) = g'(a)(x - a) + g(a)$. The critical observation we make is that when taking the limit, because x is getting arbitrarily close to a, we can replace f with L_f and replace g with L_g, and thus we observe that

$$\begin{aligned}\lim_{x\to a}\frac{f(x)}{g(x)} &= \lim_{x\to a}\frac{L_f(x)}{L_g(x)} \\ &= \lim_{x\to a}\frac{f'(a)(x-a)+f(a)}{g'(a)(x-a)+g(a)}.\end{aligned}$$

Next, we remember a key fundamental assumption: that both $f(a) = 0$ and $g(a) = 0$, as this is precisely what makes the original limit indeterminate. Substituting these values for $f(a)$ and $g(a)$ in the limit above, we now have

$$\begin{aligned}\lim_{x\to a}\frac{f(x)}{g(x)} &= \lim_{x\to a}\frac{f'(a)(x-a)}{g'(a)(x-a)} \\ &= \lim_{x\to a}\frac{f'(a)}{g'(a)},\end{aligned}$$

where the latter equality holds since x is approaching (but not equal to) a, so $\frac{x-a}{x-a} = 1$. Finally, we note that $\frac{f'(a)}{g'(a)}$ is constant with respect to x, and thus

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

We have, of course, implicitly made the assumption that $g'(a) \neq 0$, which is essential to the overall limit having the value $\frac{f'(a)}{g'(a)}$. We summarize our work above with the statement of L'Hopital's Rule, which is the formal name of the result we have shown.

L'Hopital's Rule

Let f and g be differentiable at $x = a$, and suppose that $f(a) = g(a) = 0$ and that $g'(a) \neq 0$. Then

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

In practice, we typically work with a slightly more general version of L'Hopital's Rule, which states that (under the identical assumptions as the boxed rule above)

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \lim_{x\to a}\frac{f'(x)}{g'(x)},$$

provided the righthand limit exists. This form reflects the fundamental benefit of L'Hopital's Rule: if $\frac{f(x)}{g(x)}$ produces an indeterminate limit of form $\frac{0}{0}$ as $x \to a$, it is equivalent to consider the limit of the quotient of the two functions' derivatives, $\frac{f'(x)}{g'(x)}$. For example, if we consider the limit from Preview Activity 2.8,

$$\lim_{x\to 1} \frac{x^5 + x - 2}{x^2 - 1},$$

by L'Hopital's Rule we have that

$$\lim_{x\to 1} \frac{x^5 + x - 2}{x^2 - 1} = \lim_{x\to 1} \frac{5x^4 + 1}{2x} = \frac{6}{2} = 3.$$

By being able to replace the numerator and denominator with their respective derivatives, we often move from an indeterminate limit to one whose value we can easily determine.

Activity 3.7–1

Evaluate each of the following limits. If you use L'Hopital's Rule, indicate where it was used, and be certain its hypotheses are met before you apply it.

(a) $\lim_{x\to 0} \frac{\ln(1+x)}{x}$

(b) $\lim_{x\to \pi} \frac{\cos(x)}{x}$

(c) $\lim_{x\to 1} \frac{2\ln(x)}{1 - e^{x-1}}$

(d) $\lim_{x\to 0} \frac{\sin(x) - x}{\cos(2x) - 1}$

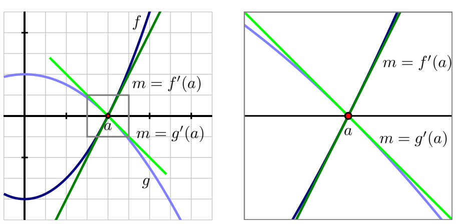


Figure 3.43: Two functions f and g that satisfy L'Hopital's Rule.

While L'Hopital's Rule can be applied in an entirely algebraic way, it is important to remember that the genesis of the rule is graphical: the main idea is that the slopes of the tangent lines to f and g at $x = a$ determine the value of the limit of $\frac{f(x)}{g(x)}$ as $x \to a$. We see this in Figure 3.43, which is a modified version of Figure 3.42, where we can see from the grid that $f'(a) = 2$ and $g'(a) = -1$, hence by L'Hopital's Rule,

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)} = \frac{2}{-1} = -2.$$

Indeed, what we observe is that it's not the fact that f and g both approach zero that matters most, but rather the *rate* at which each approaches zero that determines the value of the limit. This is a good way to remember what L'Hopital's Rule says: if $f(a) = g(a) = 0$, the the limit of $\frac{f(x)}{g(x)}$ as $x \to a$ is given by the ratio of the slopes of f and g at $x = a$.

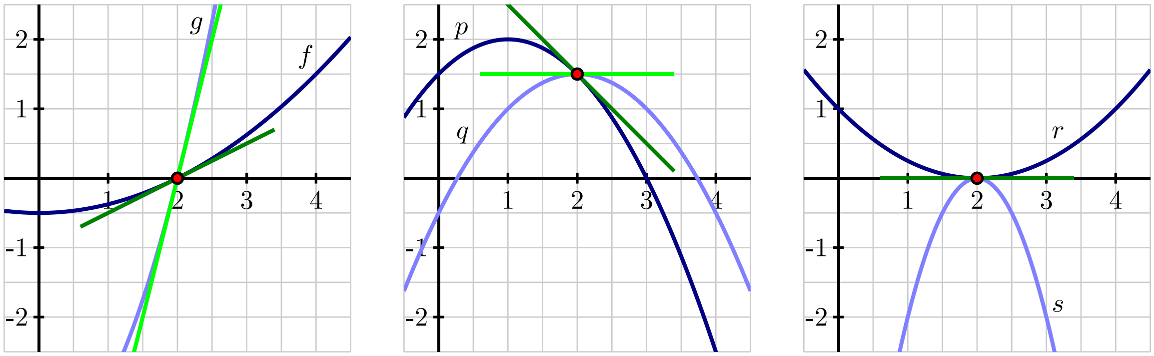


Figure 3.44: Three graphs referenced in the questions of Activity 3.7–2.

Activity 3.7–2

In this activity, we reason graphically to evaluate limits of ratios of functions about which some information is known.

(a) Use the left-hand graph to determine the values of $f(2)$, $f'(2)$, $g(2)$, and $g'(2)$. Then, evaluate

$$\lim_{x \to 2} \frac{f(x)}{g(x)}.$$

(b) Use the middle graph to find $p(2)$, $p'(2)$, $q(2)$, and $q'(2)$. Then, determine the value of

$$\lim_{x \to 2} \frac{p(x)}{q(x)}.$$

(c) Use the right-hand graph to compute $r(2)$, $r'(2)$, $s(2)$, $s'(2)$. Explain why you cannot determine the exact value of

$$\lim_{x \to 2} \frac{r(x)}{s(x)}$$

without further information being provided, but that you can determine the sign of $\lim_{x \to 2} \frac{r(x)}{s(x)}$. In addition, state what the sign of the limit will be, with justification.

Limits involving ∞

The concept of infinity, denoted ∞, arises naturally in calculus, like it does in much of mathematics. It is important to note from the outset that ∞ is a concept, but not a number itself. Indeed, the notion of ∞ naturally invokes the idea of limits. Consider, for example, the function $f(x) = \frac{1}{x}$, whose graph is pictured in Figure 3.45.

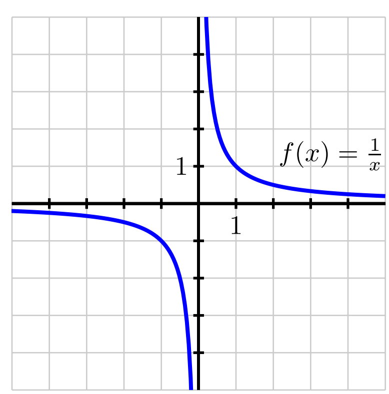


Figure 3.45: The graph of $f(x) = \frac{1}{x}$.

We note that $x = 0$ is not in the domain of f, so we may naturally wonder what happens as $x \to 0$. As $x \to 0^+$, we observe that $f(x)$ *increases without bound*. That is, we can make the value of $f(x)$ as large as we like by taking x closer and closer (but not equal) to 0, while keeping $x > 0$. This is a good way to think about what infinity represents: a quantity is tending to infinity if there is no single number that the quantity is always less than.

Recall that when we write $\lim_{x \to a} f(x) = L$, this means that can make $f(x)$ as close to L as we'd like by taking x sufficiently close (but not equal) to a. We thus expand this notation and language to include the possibility that either L or a can be ∞. For instance, for $f(x) = \frac{1}{x}$, we now write

$$\lim_{x \to 0^+} \frac{1}{x} = \infty,$$

by which we mean that we can make $\frac{1}{x}$ as large as we like by taking x sufficiently close (but not equal) to 0. In a similar way, we naturally write

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

since we can make $\frac{1}{x}$ as close to 0 as we'd like by taking x sufficiently large (i.e., by letting x increase without bound).

In general, we understand the notation $\lim_{x \to a} f(x) = \infty$ to mean that we can make $f(x)$ as large as we'd like by taking x sufficiently close (but not equal) to a, and the notation $\lim_{x \to \infty} f(x) = L$ to mean that we can make $f(x)$ as close to L as we'd like by taking x sufficiently large. This notation applies to left- and right-hand limits, plus we can also use limits involving $-\infty$. For example, returning to Figure 3.45 and $f(x) = \frac{1}{x}$, we can say that

$$\lim_{x \to 0^-} \frac{1}{x} = -\infty \quad \text{and} \quad \lim_{x \to -\infty} \frac{1}{x} = 0.$$

Finally, we write

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

when we can make the value of $f(x)$ as large as we'd like by taking x sufficiently large. For example,

$$\lim_{x \to \infty} x^2 = \infty.$$

Note particularly that limits involving infinity identify *vertical* and *horizontal asymptotes* of a function. If $\lim_{x \to a} f(x) = \infty$, then $x = a$ is a vertical asymptote of f, while if $\lim_{x \to \infty} f(x) = L$, then $y = L$ is a horizontal asymptote of f. Similar statements can be made using $-\infty$, as well as with left- and right-hand limits as $x \to a^-$ or $x \to a^+$.

In precalculus classes, it is common to study the *end behavior* of certain families of functions, by which we mean the behavior of a function as $x \to \infty$ and as $x \to -\infty$. Here we briefly examine a library of some familiar functions and note the values of several limits involving ∞.

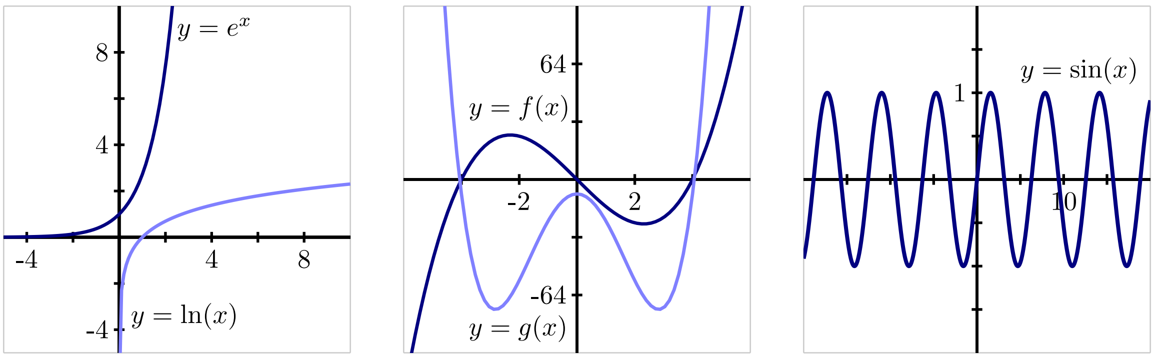


Figure 3.46: Graphs of some familiar functions whose end behavior as $x \to \pm\infty$ is known. In the middle graph, $f(x) = x^3 - 16x$ and $g(x) = x^4 - 16x^2 - 8$.

For the natural exponential function e^x, we note that $\lim_{x\to\infty} e^x = \infty$ and $\lim_{x\to-\infty} e^x = 0$, while for the related exponential decay function e^{-x}, observe that these limits are reversed, with $\lim_{x\to\infty} e^{-x} = 0$ and $\lim_{x\to-\infty} e^{-x} = \infty$. Turning to the natural logarithm function, we have $\lim_{x\to0^+} \ln(x) = -\infty$ and $\lim_{x\to\infty} \ln(x) = \infty$. While both e^x and $\ln(x)$ grow without bound as $x \to \infty$, the exponential function does so much more quickly than the logarithm function does. We'll soon use limits to quantify what we mean by "quickly."

For polynomial functions of the form $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots a_1 x + a_0$, the end behavior depends on the sign of a_n and whether the highest power n is even or odd. If n is even and a_n is positive, then $\lim_{x\to\infty} p(x) = \infty$ and $\lim_{x\to-\infty} p(x) = \infty$, as in the plot of g in Figure 3.46. If instead a_n is negative, then $\lim_{x\to\infty} p(x) = -\infty$ and $\lim_{x\to-\infty} p(x) = -\infty$. In the situation where n is odd, then either $\lim_{x\to\infty} p(x) = \infty$ and $\lim_{x\to-\infty} p(x) = \infty$ (which occurs when a_n is positive, as in the graph of f in Figure 3.46), or $\lim_{x\to\infty} p(x) = \infty$ and $\lim_{x\to-\infty} p(x) = \infty$ (when a_n is negative).

A function can fail to have a limit as $x \to \infty$. For example, consider the plot of the sine function at right in Figure 3.46. Because the function continues oscillating between -1 and 1 as $x \to \infty$, we say that $\lim_{x\to\infty} \sin(x)$ does not exist.

Finally, it is straightforward to analyze the behavior of any rational function as $x \to \infty$. Consider, for example, the function

$$q(x) = \frac{3x^2 - 4x + 5}{7x^2 + 9x - 10}.$$

Note that both $(3x^2 - 4x + 5) \to \infty$ as $x \to \infty$ and $(7x^2 + 9x - 10) \to \infty$ as $x \to \infty$. Here we say that $\lim_{x\to\infty} q(x)$ has indeterminate form $\frac{\infty}{\infty}$, much like we did when we encountered limits of the form $\frac{0}{0}$. We can determine the value of this limit through a standard algebraic approach. Multiplying the numerator and denominator each by $\frac{1}{x^2}$, we find that

$$\begin{aligned} \lim_{x\to\infty} q(x) &= \lim_{x\to\infty} \frac{3x^2 - 4x + 5}{7x^2 + 9x - 10} \cdot \frac{\frac{1}{x^2}}{\frac{1}{x^2}} \\ &= \lim_{x\to\infty} \frac{3 - 4\frac{1}{x} + 5\frac{1}{x^2}}{7 + 9\frac{1}{x} - 10\frac{1}{x^2}} \\ &= \frac{3}{7} \end{aligned}$$

since $\frac{1}{x^2} \to 0$ and $\frac{1}{x} \to 0$ as $x \to \infty$. This shows that the rational function q has a horizontal asymptote at $y = \frac{3}{7}$. A similar approach can be used to determine the limit of any rational function as $x \to \infty$.

But how should we handle a limit such as

$$\lim_{x\to\infty} \frac{x^2}{e^x}?$$

Here, both $x^2 \to \infty$ and $e^x \to \infty$, but there is not an obvious algebraic approach that enables us to find the limit's value. Fortunately, it turns out that L'Hopital's Rule extends to cases involving infinity.

L'Hopital's Rule (∞)

If f and g are differentiable and both approach zero or both approach $\pm\infty$ as $x \to a$ (where a is allowed to be ∞), then

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

To evaluate $\lim_{x\to\infty} \frac{x^2}{e^x}$, we observe that we can apply L'Hopital's Rule, since both $x^2 \to \infty$ and $e^x \to \infty$. Doing so, it follows that

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

This updated limit is still indeterminate and of the form $\frac{\infty}{\infty}$, but it is simpler since $2x$ has replaced x^2. Hence, we can apply

L'Hopital's Rule again, by which we find that

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

Now, since 2 is constant and $e^x \to \infty$ as $x \to \infty$, it follows that $\frac{2}{e^x} \to 0$ as $x \to \infty$, which shows that

$$\lim_{x\to\infty} \frac{x^2}{e^x} = 0.$$

Activity 3.7–3

Evaluate each of the following limits. If you use L'Hopital's Rule, indicate where it was used, and be certain its hypotheses are met before you apply it.

(a) $\displaystyle\lim_{x\to\infty} \frac{x}{\ln(x)}$

(b) $\displaystyle\lim_{x\to\infty} \frac{e^x + x}{2e^x + x^2}$

(c) $\displaystyle\lim_{x\to 0^+} \frac{\ln(x)}{\frac{1}{x}}$

(d) $\displaystyle\lim_{x\to \frac{\pi}{2}^-} \frac{\tan(x)}{x - \frac{\pi}{2}}$

(e) $\displaystyle\lim_{x\to\infty} xe^{-x}$

When we are considering the limit of a quotient of two functions $\frac{f(x)}{g(x)}$ that results in an indeterminate form of $\frac{\infty}{\infty}$, in essence we are asking which function is growing faster without bound. We say that the function g *dominates* the function f as $x \to \infty$ provided that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0,$$

whereas f dominates g provided that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = \infty$. Finally, if the value of $\lim_{x\to\infty} \frac{f(x)}{g(x)}$ is finite and nozero, we say that f and g *grow at the same rate*. For example, from earlier work we know that $\lim_{x\to\infty} \frac{x^2}{e^x} = 0$, so e^x dominates x^2, while $\lim_{x\to\infty} \frac{3x^2-4x+5}{7x^2+9x-10} = \frac{3}{7}$, so $f(x) = 3x^2 - 4x + 5$ and $g(x) = 7x^2 + 9x - 10$ grow at the same rate.

Indeterminate Forms $0 \cdot \infty$ and $\infty - \infty$

L'Hôpital's Rule can only be applied to ratios of functions. When faced with an indeterminate form such as $0 \cdot \infty$ or $\infty - \infty$, we can sometimes apply algebra to rewrite the limit so that l'Hôpital's Rule can be applied. We demonstrate the general idea in the next example.

Example 1

Evaluate the following limits.

1) $\lim_{x\to 0^+} x \cdot e^{1/x}$

2) $\lim_{x\to 0^-} x \cdot e^{1/x}$

3) $\lim_{x\to\infty} \ln(x+1) - \ln x$

4) $\lim_{x\to\infty} x^2 - e^x$

Solution.

1) As $x \to 0^+$, $x \to 0$ and $e^{1/x} \to \infty$. Thus we have the indeterminate form $0 \cdot \infty$. We rewrite the expression $x \cdot e^{1/x}$ as $\dfrac{e^{1/x}}{1/x}$; now, as $x \to 0^+$, we get the indeterminate form ∞/∞ to which l'Hôpital's Rule can be applied.

$$\lim_{x\to 0^+} x \cdot e^{1/x} = \lim_{x\to 0^+} \frac{e^{1/x}}{1/x} \overset{\text{by LHR}}{=} \lim_{x\to 0^+} \frac{(-1/x^2)e^{1/x}}{-1/x^2} = \lim_{x\to 0^+} e^{1/x} = \infty.$$

Interpretation: $e^{1/x}$ grows "faster" than x shrinks to zero, meaning their product grows without bound.

2) As $x \to 0^-$, $x \to 0$ and $e^{1/x} \to e^{-\infty} \to 0$. The the limit evaluates to $0 \cdot 0$ which is not an indeterminate form. We conclude then that

$$\lim_{x\to 0^-} x \cdot e^{1/x} = 0.$$

3) This limit initially evaluates to the indeterminate form $\infty - \infty$. By applying a logarithmic rule, we can rewrite the limit as

$$\lim_{x\to\infty} \ln(x+1) - \ln x = \lim_{x\to\infty} \ln\left(\frac{x+1}{x}\right).$$

As $x \to \infty$, the argument of the ln term approaches ∞/∞, to which we can apply l'Hôpital's Rule.

$$\lim_{x\to\infty} \frac{x+1}{x} \overset{\text{by LHR}}{=} \frac{1}{1} = 1.$$

Since $x \to \infty$ implies $\dfrac{x+1}{x} \to 1$, it follows that

$$x \to \infty \quad \text{implies} \quad \ln\left(\frac{x+1}{x}\right) \to \ln 1 = 0.$$

Thus

$$\lim_{x\to\infty} \ln(x+1) - \ln x = \lim_{x\to\infty} \ln\left(\frac{x+1}{x}\right) = 0.$$

Interpretation: since this limit evaluates to 0, it means that for large x, there is essentially no difference between $\ln(x+1)$ and $\ln x$; their difference is essentially 0.

4) The limit $\lim\limits_{x\to\infty} x^2 - e^x$ initially returns the indeterminate form $\infty - \infty$. We can rewrite the expression by factoring out x^2; $x^2 - e^x = x^2\left(1 - \dfrac{e^x}{x^2}\right)$. We need to evaluate how e^x/x^2 behaves as $x \to \infty$:

$$\lim_{x\to\infty} \frac{e^x}{x^2} \overset{\text{by LHR}}{=} \lim_{x\to\infty} \frac{e^x}{2x} \overset{\text{by LHR}}{=} \lim_{x\to\infty} \frac{e^x}{2} = \infty.$$

Thus $\lim_{x\to\infty} x^2(1 - e^x/x^2)$ evaluates to $\infty \cdot (-\infty)$, which is not an indeterminate form; rather, $\infty \cdot (-\infty)$ evaluates to $-\infty$. We conclude that $\lim\limits_{x\to\infty} x^2 - e^x = -\infty$.

Interpretation: as x gets large, the difference between x^2 and e^x grows very large.

Indeterminate Forms 0^0, 1^∞ and ∞^0

When faced with an indeterminate form that involves a power, it often helps to employ the natural logarithmic function.

Evaluating Limits Involving Indeterminate Forms 0^0, 1^∞ and ∞^0

If $\lim\limits_{x\to c} \ln\big(f(x)\big) = L$, then $\lim\limits_{x\to c} f(x) = \lim\limits_{x\to c} e^{\ln(f(x))} = e^L$.

Example 2

Evaluate the following limits.

1) $\lim\limits_{x\to\infty} \left(1 + \dfrac{1}{x}\right)^x$

2) $\lim\limits_{x\to 0^+} x^x$.

Solution.

1) This limit is a special limit, and it has important applications within mathematics and finance. Note that the exponent approaches ∞ while the base approaches 1, leading to the indeterminate form 1^∞. Let $f(x) = (1+1/x)^x$; the problem asks to evaluate $\lim\limits_{x\to\infty} f(x)$. Let's first evaluate $\lim\limits_{x\to\infty} \ln\big(f(x)\big)$.

$$\begin{aligned}
\lim_{x\to\infty} \ln\big(f(x)\big) &= \lim_{x\to\infty} \ln\left(1 + \frac{1}{x}\right)^x \\
&= \lim_{x\to\infty} x \ln\left(1 + \frac{1}{x}\right) \\
&= \lim_{x\to\infty} \frac{\ln\left(1 + \frac{1}{x}\right)}{1/x}
\end{aligned}$$

This produces the indeterminate form 0/0, so we apply l'Hôpital's Rule.

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

Thus $\lim\limits_{x\to\infty} \ln\big(f(x)\big) = 1$, and

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x = \lim_{x\to\infty} f(x) = \lim_{x\to\infty} e^{\ln(f(x))} = e^1 = e.$$

2) This limit leads to the indeterminate form 0^0. Let $f(x) = x^x$ and consider first $\lim\limits_{x\to 0^+} \ln\big(f(x)\big)$.

$$\lim_{x\to 0^+} \ln\big(f(x)\big) = \lim_{x\to 0^+} \ln\left(x^x\right)$$
$$= \lim_{x\to 0^+} x\ln x$$
$$= \lim_{x\to 0^+} \frac{\ln x}{1/x}.$$

This produces the indeterminate form $-\infty/\infty$ so we apply l'Hôpital's Rule.

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

Thus $\lim\limits_{x\to 0^+} \ln\big(f(x)\big) = 0$, and

$$\lim_{x\to 0^+} x^x = \lim_{x\to 0^+} f(x) = \lim_{x\to 0^+} e^{\ln(f(x))} = e^0 = 1.$$

This result is supported by the graph of $f(x) = x^x$ given in Figure 3.47.

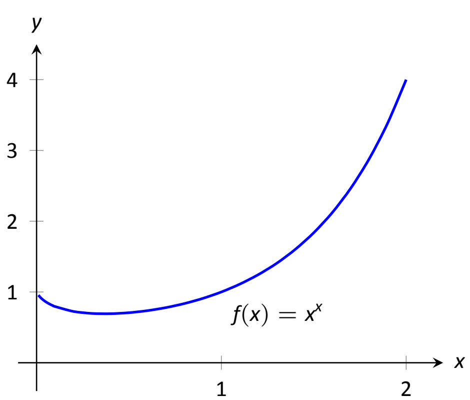


Figure 3.47: A graph of $f(x) = x^x$ supporting the fact that as $x \to 0^+$, $f(x) \to 1$.

Our brief revisit of limits will be especially rewarded when we consider *improper integration* in Section **??**.

Summary

In this section, we encountered the following important ideas:

- Derivatives be used to help us evaluate indeterminate limits of the form $\frac{0}{0}$ through L'Hopital's Rule, which is developed by replacing the functions in the numerator and denominator with their tangent line approximations. In particular, if $f(a) = g(a) = 0$ and f and g are differentiable at a, L'Hopital's Rule tells us that
$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}.$$
- When we write $x \to \infty$, this means that x is increasing without bound. We thus use ∞ along with limit notation to write $\lim_{x\to\infty} f(x) = L$, which means we can make $f(x)$ as close to L as we like by choosing x to be sufficiently large, and similarly $\lim_{x\to a} f(x) = \infty$, which means we can make $f(x)$ as large as we like by choosing x sufficiently close to a.
- A version of L'Hopital's Rule also allows us to use derivatives to assist us in evaluating indeterminate limits of the form $\frac{\infty}{\infty}$. In particular, If f and g are differentiable and both approach zero or both approach $\pm\infty$ as $x \to a$ (where a is allowed to be ∞), then
$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}.$$

Exercises

Terms and Concepts

1) List the different indeterminate forms described in this section.
2) T/F: l'Hôpital's Rule provides a faster method of computing derivatives.
3) T/F: l'Hôpital's Rule states that $\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right]=\frac{f'(x)}{g'(x)}$.
4) Explain what the indeterminate form "1^∞" means.
5) Fill in the blanks: The Quotient Rule is applied to $\frac{f(x)}{g(x)}$ when taking ________; l'Hôpital's Rule is applied when taking certain ________.
6) Create (but do not evaluate!) a limit that returns "∞^0".
7) Create a function $f(x)$ such that $\lim\limits_{x\to 1} f(x)$ returns "0^0".

Problems

In exercises 8–52, evaluate the given limit.

8) $\lim\limits_{x\to 1}\frac{x^2+x-2}{x-1}$
9) $\lim\limits_{x\to 2}\frac{x^2+x-6}{x^2-7x+10}$
10) $\lim\limits_{x\to \pi}\frac{\sin x}{x-\pi}$
11) $\lim\limits_{x\to \pi/4}\frac{\sin x-\cos x}{\cos(2x)}$
12) $\lim\limits_{x\to 0}\frac{\sin(5x)}{x}$
13) $\lim\limits_{x\to 0}\frac{\sin(2x)}{x+2}$
14) $\lim\limits_{x\to 0}\frac{\sin(2x)}{\sin(3x)}$
15) $\lim\limits_{x\to 0}\frac{\sin(ax)}{\sin(bx)}$
16) $\lim\limits_{x\to 0^+}\frac{e^x-1}{x^2}$
17) $\lim\limits_{x\to 0^+}\frac{e^x-x-1}{x^2}$
18) $\lim\limits_{x\to 0^+}\frac{x-\sin x}{x^3-x^2}$
19) $\lim\limits_{x\to \infty}\frac{x^4}{e^x}$
20) $\lim\limits_{x\to \infty}\frac{\sqrt{x}}{e^x}$
21) $\lim\limits_{x\to \infty}\frac{e^x}{\sqrt{x}}$
22) $\lim\limits_{x\to \infty}\frac{e^x}{2^x}$
23) $\lim\limits_{x\to \infty}\frac{e^x}{3^x}$
24) $\lim\limits_{x\to 3}\frac{x^3-5x^2+3x+9}{x^3-7x^2+15x-9}$
25) $\lim\limits_{x\to -2}\frac{x^3+4x^2+4x}{x^3+7x^2+16x+12}$
26) $\lim\limits_{x\to \infty}\frac{\ln x}{x}$
27) $\lim\limits_{x\to \infty}\frac{\ln(x^2)}{x}$
28) $\lim\limits_{x\to \infty}\frac{\left(\ln x\right)^2}{x}$
29) $\lim\limits_{x\to 0^+} x\cdot\ln x$
30) $\lim\limits_{x\to 0^+} \sqrt{x}\cdot\ln x$
31) $\lim\limits_{x\to \infty} x^3-x^2$
32) $\lim\limits_{x\to \infty} \sqrt{x}-\ln x$
33) $\lim\limits_{x\to -\infty} xe^x$
34) $\lim\limits_{x\to 0^+} \frac{1}{x^2}e^{-1/x}$
35) $\lim\limits_{x\to 0^+} (1+x)^{1/x}$
36) $\lim\limits_{x\to 0^+} (2x)^x$
37) $\lim\limits_{x\to 0^+} (2/x)^x$
38) $\lim\limits_{x\to 0^+} (\sin x)^x$ Hint: use the Squeeze Theorem.
39) $\lim\limits_{x\to 1^+} (1-x)^{1-x}$
40) $\lim\limits_{x\to \infty} (x)^{1/x}$
41) $\lim\limits_{x\to \infty} (1/x)^x$
42) $\lim\limits_{x\to 1^+} (\ln x)^{1-x}$
43) $\lim\limits_{x\to \infty} (1+x)^{1/x}$
44) $\lim\limits_{x\to \infty} (1+x^2)^{1/x}$
45) $\lim\limits_{x\to \pi/2} \tan x\cos x$
46) $\lim\limits_{x\to \pi/2} \tan x\sin(2x)$
47) $\lim\limits_{x\to 1^+} \frac{1}{\ln x}-\frac{1}{x-1}$
48) $\lim\limits_{x\to 3^+} \frac{5}{x^2-9}-\frac{x}{x-3}$
49) $\lim\limits_{x\to \infty} x\tan(1/x)$
50) $\lim\limits_{x\to \infty} \frac{(\ln x)^3}{x}$
51) $\lim\limits_{x\to 1} \frac{x^2+x-2}{\ln x}$
52) $\lim\limits_{x\to 0^+} xe^{1/x}$

3.8 *Newton's Method*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- If we are unable to solve an equation algebraically, how can we approximate the solutions of the equation using calculus?

Introduction

Solving equations is one of the most important things we do in mathematics, yet we are surprisingly limited in what we can solve analytically. For instance, equations as simple as $x^5 + x + 1 = 0$ or $\cos(x) = x$ cannot be solved by algebraic methods in terms of familiar functions. Fortunately, there are methods that can give us *approximate* solutions to equations like these. These methods can usually give an approximation correct to as many decimal places as we like. In Section 1.3 we learned about the Bisection Method. This section focuses on another technique (which generally works faster), called Newton's Method.

Newton's Method

Newton's Method is built around tangent lines. The main idea is that if x is sufficiently close to a root of $f(x)$, then the tangent line to the graph at $(x, f(x))$ will cross the x-axis at a point closer to the root than x.

We start Newton's Method with an initial guess about roughly where the root is. Call this x_0. (See Figure 3.48-(a).) Draw the tangent line to the graph at $(x_0, f(x_0))$ and see where it meets the x-axis. Call this point x_1. Then repeat the process – draw the tangent line to the graph at $(x_1, f(x_1))$ and see where it meets the x-axis. (See Figure 3.48-(b).) Call this point x_2. Repeat the process again to get x_3, x_4, etc. This sequence of points will often converge rather quickly to a root of f.

We can use this *geometric* process to create an *algebraic* process. Let's look at how we found x_1. We started with the tangent line to the graph at $(x_0, f(x_0))$. The slope of this tangent line is $f'(x_0)$ and the equation of the line is

$$y = f'(x_0)(x - x_0) + f(x_0).$$

This line crosses the x-axis when $y = 0$, and the x–value where it crosses is what we called x_1. So let $y = 0$ and replace x with x_1, giving the equation:

$$0 = f'(x_0)(x_1 - x_0) + f(x_0).$$

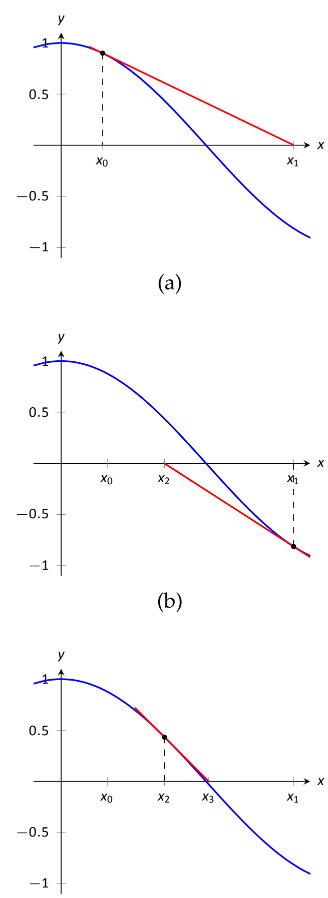


Figure 3.48: Demonstrating the geometric concept behind Newton's Method.

Now solve for x_1:

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

Since we repeat the same geometric process to find x_2 from x_1, we have

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

In general, given an approximation x_n, we can find the next approximation, x_{n+1} as follows:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

We summarize this process as follows.

Newton's Method

Let f be a differentiable function on an interval I with a root in I. To approximate the value of the root, accurate to d decimal places:

1. Choose a value x_0 as an initial approximation of the root. (This is often done by looking at a graph of f.)
2. Create successive approximations iteratively; given an approximation x_n, compute the next approximation x_{n+1} as
$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$
3. Stop the iterations when successive approximations do not differ in the first d places after the decimal point.

Newton's Method is not infallible. The sequence of approximate values may not converge, or it may converge so slowly that one is "tricked" into thinking a certain approximation is better than it actually is. These issues will be discussed at the end of the section.

Let's practice Newton's Method with a concrete example.

Example 1

Approximate the real root of $x^3 - x^2 - 1 = 0$, accurate to the first 3 places after the decimal, using Newton's Method and an initial approximation of $x_0 = 1$.

Solution. To begin, we compute $f'(x) = 3x^2 - 2x$. Then we apply the

Newton's Method algorithm.

$$x_1 = 1 - \frac{f(1)}{f'(1)} = 1 - \frac{1^3 - 1^2 - 1}{3 \cdot 1^2 - 2 \cdot 1} = 2,$$

$$x_2 = 2 - \frac{f(2)}{f'(2)} = 2 - \frac{2^3 - 2^2 - 1}{3 \cdot 2^2 - 2 \cdot 2} = 1.625,$$

$$x_3 = 1.625 - \frac{f(1.625)}{f'(1.625)} = 1.625 - \frac{1.625^3 - 1.625^2 - 1}{3 \cdot 1.625^2 - 2 \cdot 1.625} \approx 1.48579.$$

$$x_4 = 1.48579 - \frac{f(1.48579)}{f'(1.48579)} \approx 1.46596$$

$$x_5 = 1.46596 - \frac{f(1.46596)}{f'(1.46596)} \approx 1.46557$$

We performed 5 iterations of Newton's Method to find a root accurate to the first 3 places after the decimal; our final approximation is 1.465. The exact value of the root, to six decimal places, is 1.465571; It turns out that our x_5 is accurate to more than just 3 decimal places.

A graph of $f(x)$ is given in Figure!3.49. We can see from the graph that our initial approximation of $x_0 = 1$ was not particularly accurate; a closer guess would have been $x_0 = 1.5$. Our choice was based on ease of initial calculation, and shows that Newton's Method can be robust enough that we do not have to make a very accurate initial approximation.

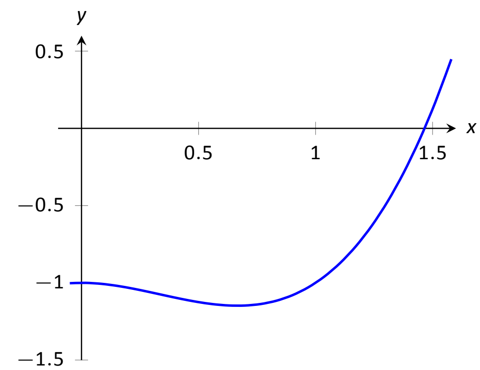


Figure 3.49: A graph of $f(x) = x^3 - x^2 - 1$ in Example 1.

We can automate this process on a calculator that has an `Ans` key that returns the result of the previous calculation. Start by pressing `1` and then `Enter`. (We have just entered our initial guess, $x_0 = 1$.) Now compute

$$\texttt{Ans} - \frac{f(\texttt{Ans})}{f'(\texttt{Ans})}$$

by entering the following and repeatedly press the `Enter` key:

```
Ans-(Ans^3-Ans^2-1)/(3*Ans^2-2*Ans)
```

Each time we press the `Enter` key, we are finding the successive approximations, $x_1, x_2, \ldots,$ and each one is getting closer to the root. In fact, once we get past around x_7 or so, the approximations don't appear to be changing. They actually are changing, but the change is far enough to the right of the decimal point that it doesn't show up on the calculator's display. When this happens, we can be pretty confident that we have found an accurate approximation.

Using a calculator in this manner makes the calculations simple; many iterations can be computed very quickly. In general, one would usually run Newton's Method in this way, finding approximations until the difference between two successive approximations is less than some prescribed tolerance, like maybe 10^{-10}, whatever is necessary for the problem at hand.

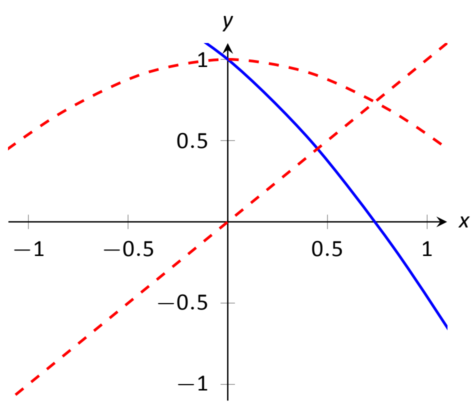


Figure 3.50: A graph of $f(x) = \cos x - x$ used to find an initial approximation of its root.

Example 2

Use Newton's Method to approximate a solution to $\cos(x) = x$, accurate to 5 places after the decimal.

Solution. Newton's Method provides a method of solving $f(x) = 0$; it is not (directly) a method for solving equations like $f(x) = g(x)$. However, this is not a problem; we can rewrite the latter equation as $f(x) - g(x) = 0$ and then use Newton's Method.

So we rewrite $\cos(x) = x$ as $\cos(x) - x = 0$. Written this way, we are finding a root of $f(x) = \cos(x) - x$. We compute $f'(x) = -\sin x - 1$. Next we need a starting value, x_0. Consider Figure 3.50, where $f(x) = \cos(x) - x$ is graphed. It seems that $x_0 = 0.75$ is pretty close to the root, so we will use that as our x_0. (The figure also shows the graphs of $y = \cos(x)$ and $y = x$, drawn in red. Note how they intersect at the same x value as when $f(x) = 0$.)

We now compute x_1, x_2, etc. The formula for x_1 is

$$x_1 = 0.75 - \frac{\cos(0.75) - 0.75}{-\sin(0.75) - 1} \approx 0.7391111388.$$

To 11 decimal places, this gives .7391111388. We then compute Apply Newton's Method again to find x_2:

$$x_2 = 0.7391111388 - \frac{\cos(0.7391111388) - 0.7391111388}{-\sin(0.7391111388) - 1} \approx 0.7390851334.$$

We can continue this way, but it is really best to automate this process. On a calculator with an Ans key, we would start by pressing 0.75, then `Enter`, inputting our initial approximation. We then enter:

`\begin{center}\texttt{Ans - (cos(Ans)-Ans)/(-sin(Ans)-1)}.\end{center}`

Repeatedly pressing the `Enter` key gives successive approximations. We quickly find:

$$x_3 = 0.7390851332$$
$$x_4 = 0.7390851332.$$

Our approximations x_2 and x_3 did not differ for at least the first 5 places after the decimal, so we could have stopped. However, using our calculator in the manner described is easy, so finding x_4 was not hard. It is interesting to see how we found an approximation, accurate to as many decimal places as our calculator displays, in just 4 iterations.

If you know how to program, you can translate the following pseudocode into your favorite language to perform the computation in this problem.

```
x = .75
while true
    oldx = x
    x = x - (cos(x)-x)/(-sin(x)-1)
    print x
    if abs(x-oldx) < .0000000001
```

```
        break
```

This code calculates x_1, x_2, etc., storing each result in the variable `x`. The previous approximation is stored in the variable `oldx`. We continue looping until the difference between two successive approximations, `abs(x-oldx)`, is less than some small tolerance, in this case, `.0000000001`.

Convergence of Newton's Method

What should one use for the initial guess, x_0? Generally, the closer to the actual root the initial guess is, the better. However, some initial guesses should be avoided. For instance, consider Example 1 where we sought the root to $f(x) = x^3 - x^2 - 1$. Choosing $x_0 = 0$ would have been a particularly poor choice. Consider Figure 3.51, where $f(x)$ is graphed along with its tangent line at $x = 0$. Since $f'(0) = 0$, the tangent line is horizontal and does not intersect the x–axis. Graphically, we see that Newton's Method fails.

We can also see analytically that it fails. Since

$$x_1 = 0 - \frac{f(0)}{f'(0)}$$

and $f'(0) = 0$, we see that x_1 is not well defined.

This problem can also occur if, for instance, it turns out that $f'(x_5) = 0$. Adjusting the initial approximation x_0 will likely ameliorate the problem.

It is also possible for Newton's Method to not converge while each successive approximation is well defined. Consider $f(x) = x^{1/3}$, as shown in Figure 3.52. It is clear that the root is $x = 0$, but let's approximate this with $x_0 = 0.1$. Figure 3.52-(a) shows graphically the calculation of x_1; notice how it is farther from the root than x_0. Figures 3.52-(b) and -(c) show the calculation of x_2 and x_3, which are even farther away; our successive approximations are getting worse. (It turns out that in this particular example, each successive approximation is twice as far from the true answer as the previous approximation.)

There is no "fix" to this problem; Newton's Method simply will not work and another method must be used.

While Newton's Method does not always work, it does work "most of the time," and it is generally very fast. Once the approximations get close to the root, Newton's Method can as much as double the number of correct decimal places with each successive approximation. A course in Numerical Analysis will introduce the reader to more iterative root finding methods, as well as give greater detail about the strengths and weaknesses of Newton's Method.

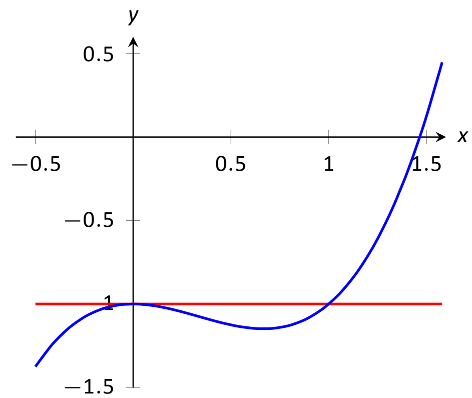
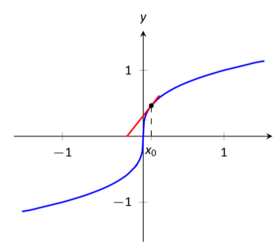
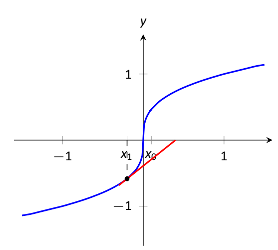


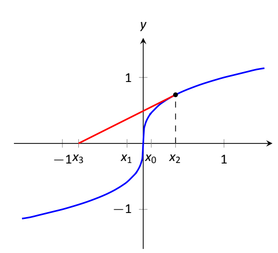
Figure 3.51: A graph of $f(x) = x^3 - x^2 - 1$, showing why an initial approximation of $x_0 = 0$ with Newton's Method fails.



(a)



(b)



(c)

Figure 3.52: Newton's Method fails to find a root of $f(x) = x^{1/3}$, regardless of the choice of x_0.

Summary

In this section, we encountered the following important ideas:

- Derivatives and tangent lines can be used to approximate the solutions of an equation. Create successive approximations iteratively; given an approximation x_n, compute the next approximation x_{n+1} as

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Exercises

Terms and Concepts

1) T/F: Given a function $f(x)$, Newton's Method produces an exact solution to $f(x) = 0$.

2) T/F: In order to get a solution to $f(x) = 0$ accurate to d places after the decimal, at least $d + 1$ iterations of Newton's Method must be used.

Problems

In exercises 3–7, the roots of $f(x)$ are known or are easily found. Use 5 iterations of Newton's Method with the given initial approximation to approximate the root. Compare it to the known value of the root.

3) $f(x) = \cos x$, $x_0 = 1.5$

4) $f(x) = \sin x$, $x_0 = 1$

5) $f(x) = x^2 + x - 2$, $x_0 = 0$

6) $f(x) = x^2 - 2$, $x_0 = 1.5$

7) $f(x) = \ln x$, $x_0 = 2$

In exercises 8–11, use Newton's Method to approximate all roots of the given functions accurate to 3 places after the decimal. If an interval is given, find only the roots that lie in that interval.

8) $f(x) = x^3 + 5x^2 - x - 1$

9) $f(x) = x^4 + 2x^3 - 7x^2 - x + 5$

10) $f(x) = x^{17} - 2x^{13} - 10x^8 + 10$ on $(-2, 2)$

11) $f(x) = x^2 \cos x + (x - 1)\sin x$ on $(-3, 3)$

In exercises 12–15, use Newton's Method to approximate when the given functions are equal, accurate to 3 places after the decimal.

12) $f(x) = x^2$, $g(x) = \cos x$

13) $f(x) = x^2 - 1$, $g(x) = \sin x$

14) $f(x) = e^{x^2}$, $g(x) = \cos x$

15) $f(x) = x$, $g(x) = \tan x$ on $[-6, 6]$

16) Why does Newton's Method fail in finding a root of $f(x) = x^3 - 3x^2 + x + 3$ when $x_0 = 1$?

17) Why does Newton's Method fail in finding a root of $f(x) = -17x^4 + 130x^3 - 301x^2 + 156x + 156$ when $x_0 = 1$?

Chapter 4
Integration

4.1 Determining distance traveled from velocity

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- If we know the velocity of a moving body at every point in a given interval, can we determine the distance the object has traveled on the time interval?
- How is the problem of finding distance traveled related to finding the area under a certain curve?
- If velocity is negative, how does this impact the problem of finding distance traveled?

Introduction

In the first section of the text, we considered a situation where a moving object had a known position at time t. In particular, we stipulated that a tennis ball tossed into the air had its height s (in feet) at time t (in seconds) given by $s(t) = 64 - 16(t-1)^2$. From this starting point, we investigated the average velocity of the ball on a given interval $[a, b]$, computed by the difference quotient $\frac{s(b)-s(a)}{b-a}$, and eventually found that we could determine the exact instantaneous velocity of the ball at time t by taking the derivative of the position function,

$$s'(t) = \lim_{h \to 0} \frac{s(t+h) - s(t)}{h}.$$

Thus, given a differentiable position function, we are able to know the exact velocity of the moving object at any point in time.

Moreover, from this foundational problem involving position and velocity we have learned a great deal. Given a differentiable function f, we are now able to find its derivative and use this derivative to determine the function's instantaneous rate of change at any point in the domain, as well as to find where the function is increasing or decreasing, is concave up or concave down, and has relative extremes. The vast majority of the problems and applications we have considered have involved the situation where a particular function is known and we seek information that relies on knowing the function's instantaneous rate of change. That is, we have typically proceeded from a function f to its derivative, f', and then used the meaning of the derivative to help us answer important questions.

In a much smaller number of situations so far, we have encountered the reverse situation where we instead know the derivative, f', and have tried to deduce information about f. It is this particular problem that will be the focus of our attention in most of Chapter 4: if we know the instantaneous rate of change of a function, are we able to determine the function itself? To begin, we start with a more focused question: if we know the instantaneous velocity of an object moving along a straight line path, can we determine its corresponding position function?

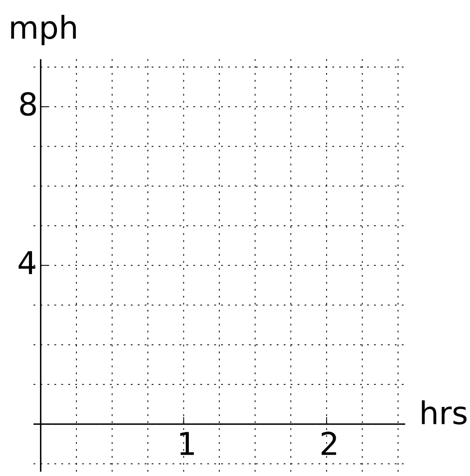


Figure 4.1: Axes for plotting $y = v(t)$.

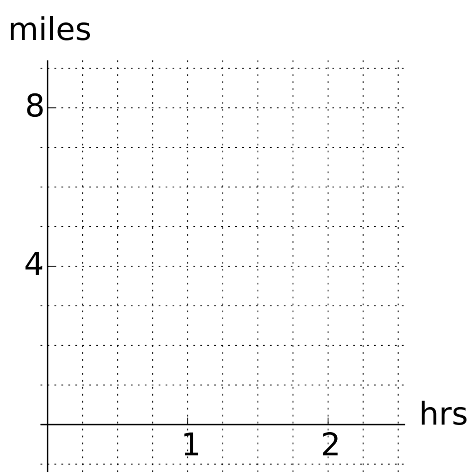


Figure 4.2: Axes for plotting $y = s(t)$.

Preview Activity 4.1

Suppose that a person is taking a walk along a long straight path and walks at a constant rate of 3 miles per hour.

(a) On the axes provided in Figure 4.1, sketch a labeled graph of the velocity function $v(t) = 3$.

Note that while the scale on the two sets of axes is the same, the units on the axes provided in Figure 4.1 differ from those in Figure 4.2.

(b) How far did the person travel during the two hours? How is this distance related to the area of a certain region under the graph of $y = v(t)$?

(c) Find an algebraic formula, $s(t)$, for the position of the person at time t, assuming that $s(0) = 0$. Explain your thinking.

(d) On the axes provided in Figure 4.2, sketch a labeled graph of the position function $y = s(t)$.

(e) For what values of t is the position function s increasing? Explain why this is the case using relevant information about the velocity function v.

Area under the graph of the velocity function

In Preview Activity 4.1, we encountered a fundamental fact: when a moving object's velocity is constant (and positive), the area under the velocity curve over a given interval tells us the distance the object traveled. As seen in Figure 4.3, if we consider an object moving at 2 miles per hour over the time interval $[1, 1.5]$, then the area A_1 of the shaded region under $y = v(t)$ on

$[1, 1.5]$ is

$$A_1 = 2\,\frac{\text{miles}}{\text{hour}} \cdot \frac{1}{2}\,\text{hours} = 1\,\text{mile}.$$

This principle holds in general simply due to the fact that distance equals rate times time, provided the rate is constant. Thus, if $v(t)$ is constant on the interval $[a, b]$, then the distance traveled on $[a, b]$ is the area A that is given by

$$A = v(a)(b - a) = v(a)\triangle t,$$

where $\triangle t$ is the change in t over the interval. Note, too, that we could use any value of $v(t)$ on the interval $[a, b]$, since the velocity is constant; we simply chose $v(a)$, the value at the interval's left endpoint. For several examples where the velocity function is piecewise constant, see `http://gvsu.edu/s/9T`.

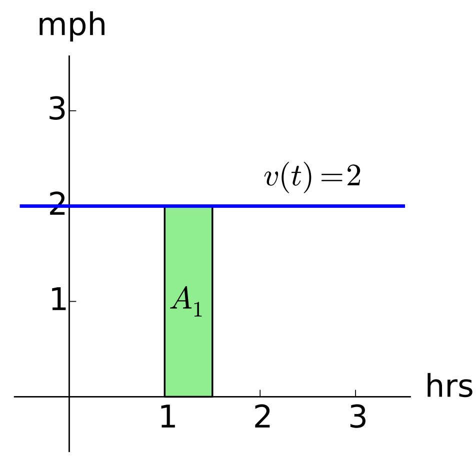


Figure 4.3: A constant velocity function.

The situation is obviously more complicated when the velocity function is not constant. At the same time, on relatively small intervals on which $v(t)$ does not vary much, the area principle allows us to estimate the distance the moving object travels on that time interval. For instance, for the non-constant velocity function shown in Figure 4.4, we see that on the interval $[1, 1.5]$, velocity varies from $v(1) = 2.5$ down to $v(1.5) \approx 2.1$. Hence, one estimate for distance traveled is the area of the pictured rectangle,

$$A_2 = v(1)\triangle t = 2.5\,\frac{\text{miles}}{\text{hour}} \cdot \frac{1}{2}\,\text{hours} = 1.25\,\text{miles}.$$

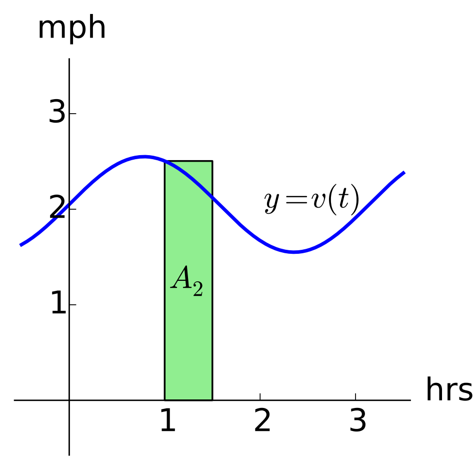


Figure 4.4: A non-constant velocity function.

Because v is decreasing on $[1, 1.5]$ and the rectangle lies above the curve, clearly $A_2 = 1.25$ is an over-estimate of the actual distance traveled.

If we want to estimate the area under the non-constant velocity function on a wider interval, say $[0, 3]$, it becomes apparent that one rectangle probably will not give a good approximation. Instead, we could use the six rectangles pictured in Figure 4.5, find the area of each rectangle, and add up the total. Obviously there are choices to make and issues to understand: how many rectangles should we use? Where should we evaluate the function to decide the rectangle's height? What happens if velocity is sometimes negative? Can we attain the exact area under any non-constant curve? These questions and more are ones we will study in what follows; for now it suffices to realize that the simple idea of the area of a rectangle gives us a powerful tool for estimating both distance traveled from a velocity function as well as the area under an arbitrary curve. To explore the setting of multiple rectangles to approximate area under a non-constant velocity function, see the applet found at `http://gvsu.edu/s/9U`.

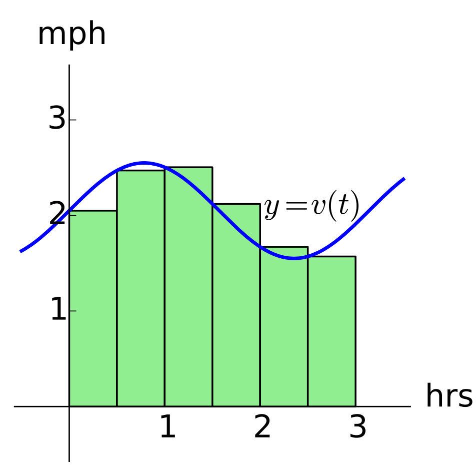


Figure 4.5: Using six rectangles to estimate the area under $y = v(t)$ on $[0, 3]$.

t	$v(t)$
0.00	1.5000
0.25	1.7891
0.50	1.9375
0.75	1.9922
1.00	2.0000
1.25	2.0078
1.50	2.0625
1.75	2.2109
2.00	2.5000

Table 4.1: A table of velocities.

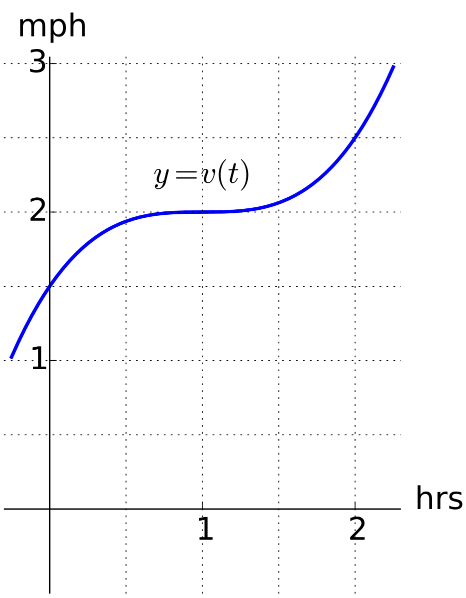


Figure 4.6: The graph of $y = v(t)$.

Activity 4.1–1

Suppose that a person is walking in such a way that her velocity varies slightly according to the information given in Table 4.1 and Figure 4.6.

(a) Using the grid, graph, and given data appropriately, estimate the distance traveled by the walker during the two hour interval from $t = 0$ to $t = 2$. You should use time intervals of width $\triangle t = 0.5$, choosing a way to use the function consistently to determine the height of each rectangle in order to approximate distance traveled.

(b) How could you get a better approximation of the distance traveled on $[0, 2]$? Explain, and then find this new estimate.

(c) Now suppose that you know that v is given by $v(t) = 0.5t^3 - 1.5t^2 + 1.5t + 1.5$. Remember that v is the derivative of the walker's position function, s. Find a formula for s so that $s' = v$.

(d) Based on your work in (c), what is the value of $s(2) - s(0)$? What is the meaning of this quantity?

Two approaches: area and antidifferentiation

When the velocity of a moving object is positive, the object's position is always increasing. While we will soon consider situations where velocity is negative and think about the ramifications of this condition on distance traveled, for now we continue to assume that we are working with a positive velocity function. In that setting, we have established that whenever v is actually constant, the exact distance traveled on an interval is the area under the velocity curve; furthermore, we have observed that when v is not constant, we can estimate the total distance traveled by finding the areas of rectangles that help to approximate the area under the velocity curve on the given interval. Hence, we see the importance of the problem of finding the area between a curve and the horizontal axis: besides being an interesting geometric question, in the setting of the curve being the (positive) velocity of a moving object, the area under the curve over an interval tells us the exact distance traveled on the interval. We can estimate this area any time we have a graph of the velocity function or a table of data that tells us some relevant values of the function.

In Activity 4.1–1, we also encountered an alternate approach to finding the distance traveled. In particular, if we know a formula for the instantaneous velocity, $y = v(t)$, of the moving body at time t, then we realize that v must be the derivative of some corresponding position function s. If we can find a formula for s from the formula for v, it follows that we know the position of the object at time t. In addition, under the assumption that velocity is positive, the change in position over a given interval then tells us the distance traveled on that interval.

For a simple example, consider the situation from Preview Activity 4.1, where a person is walking along a straight line and has velocity function $v(t) = 3$ mph. As pictured in Figure 4.7 and Figure 4.8, we see the already noted relationship between area and distance traveled on the velocity function. In addition, because the velocity is constant at 3, we know that if[1] $s(t) = 3t$, then $s'(t) = 3$, so $s(t) = 3t$ is a function whose derivative is $v(t)$. Furthermore, we now observe that $s(1.5) = 4.5$ and $s(0.25) = 0.75$, which are the respective locations of the person at times $t = 0.25$ and $t = 1.5$, and therefore

$$s(1.5) - s(0.25) = 4.5 - 0.75 = 3.75 \text{ miles.}$$

This is not only the change in position on $[0.25, 1.5]$, but also precisely the distance traveled on $[0.25, 1.5]$, which can also be computed by finding the area under the velocity curve over the same interval. There are profound ideas and connections present in this example that we will spend much of the remainder of Chapter 4 studying and exploring.

For now, it is most important to observe that if we are given a formula for a velocity function v, it can be very helpful to find a function s that satisfies $s' = v$. In this context, we say that s is an *antiderivative* of v. More generally, just as we say that f' is the derivative of f for a given function f, if we are given a function g and G is a function such that $G' = g$, we say that G is an *antiderivative* of g. For example, if $g(x) = 3x^2 + 2x$, an antiderivative of g is $G(x) = x^3 + x^2$, since $G'(x) = g(x)$. Note that we say "an" antiderivative of g rather than "the" antiderivative of g because $H(x) = x^3 + x^2 + 5$ is also a function whose derivative is g, and thus H is another antiderivative of g.

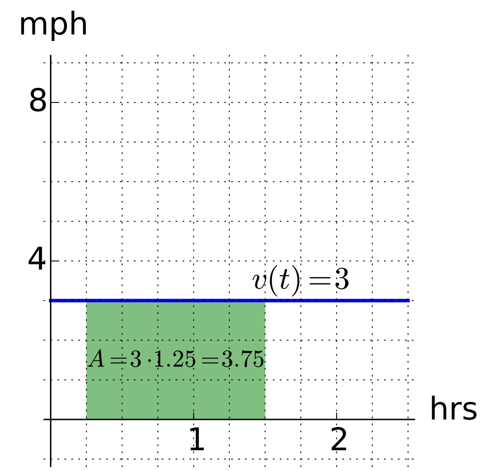


Figure 4.7: The velocity function $v(t) = 3$.

[1] Here we are making the implicit assumption that $s(0) = 0$; we will further discuss the different possibilities for values of $s(0)$ in subsequent study.

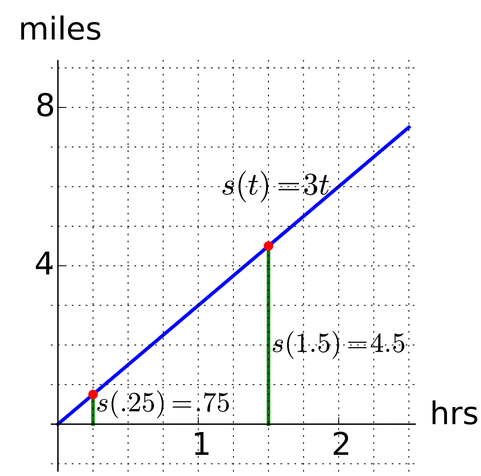


Figure 4.8: The position function $s(t) = 3t$.

Activity 4.1–2

A ball is tossed vertically in such a way that its velocity function is given by $v(t) = 32 - 32t$, where t is measured in seconds and v in feet per second. Assume that this function is valid for $0 \le t \le 2$.

(a) For what values of t is the velocity of the ball positive? What does this tell you about the motion of the ball on this interval of time values?

(b) Find an antiderivative, s, of v that satisfies $s(0) = 0$.

(c) Compute the value of $s(1) - s(\frac{1}{2})$. What is the meaning of the value you find?

(d) Using the graph of $y = v(t)$ provided in Figure 4.9, find the exact area of the region under the velocity curve between $t = \frac{1}{2}$ and $t = 1$. What is the meaning of the value you find?

(e) Answer the same questions as in (c) and (d) but instead using the

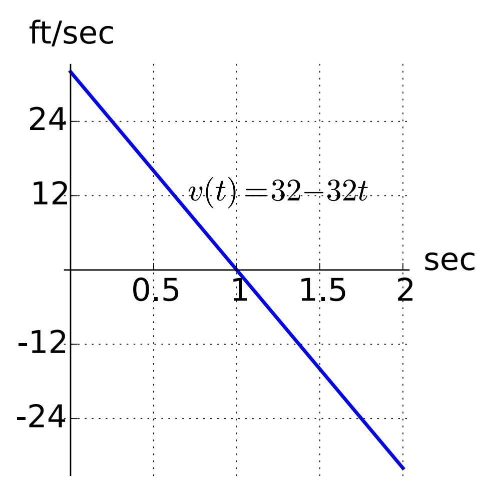


Figure 4.9: The graph of $y = v(t)$.

interval $[0, 1]$.

(f) What is the value of $s(2) - s(0)$? What does this result tell you about the flight of the ball? How is this value connected to the provided graph of $y = v(t)$? Explain.

When velocity is negative

Most of our work in this section has occurred under the assumption that velocity is positive. This hypothesis guarantees that the movement of the object under consideration is always in a single direction, and hence ensures that the moving body's change in position is the same as the distance it travels on a given interval. As we saw in Activity 4.1–2, there are natural settings in which a moving object's velocity is negative; we would like to understand this scenario fully as well.

Consider a simple example where a person goes for a walk on a beach along a stretch of very straight shoreline that runs east-west. We can naturally assume that their initial position is $s(0) = 0$, and further stipulate that their position function increases as they move east from their starting location. For instance, a position of $s = 1$ mile represents being one mile east of the start location, while $s = -1$ tells us the person is one mile west of where they began walking on the beach. Now suppose the person walks in the following manner. From the outset at $t = 0$, the person walks due east at a constant rate of 3 mph for 1.5 hours. After 1.5 hours, the person stops abruptly and begins walking due west at the constant rate of 4 mph and does so for 0.5 hours. Then, after another abrupt stop and start, the person resumes walking at a constant rate of 3 mph to the east for one more hour. What is the total distance the person traveled on the time interval $t = 0$ to $t = 3$? What is the person's total change in position over that time?

On one hand, these are elementary questions to answer because the velocity involved is constant on each interval. From $t = 0$ to $t = 1.5$, the person traveled

$$D_{[0,1.5]} = 3 \text{ miles per hour} \times 1.5 \text{ hours} = 4.5 \text{ miles}.$$

Similarly, on $t = 1.5$ to $t = 2$, having a different rate, the distance traveled is

$$D_{[1.5,2]} = 4 \text{ miles per hour} \times 0.5 \text{ hours} = 2 \text{ miles}.$$

Finally, similar calculations reveal that in the final hour, the person walked

$$D_{[2,3]} = 3 \text{ miles per hour} \times 1 \text{ hours} = 3 \text{ miles},$$

so the total distance traveled is

$$D = D_{[0,1.5]} + D_{[1.5,2]} + D_{[2,3]} = 4.5 + 2 + 3 = 9.5 \text{ miles.}$$

Since the velocity on $1.5 < t < 2$ is actually $v = -4$, being negative to indication motion in the westward direction, this tells us that the person first walked 4.5 miles east, then 2 miles west, followed by 3 more miles east. Thus, the walker's total change in position is

$$\text{change in position} = 4.5 - 2 + 3 = 5.5 \text{ miles.}$$

While we have been able to answer these questions fairly easily, it is also important to think about this problem graphically in order that we can generalize our solution to the more complicated setting when velocity is not constant, as well as to note the particular impact that negative velocity has.

In Figure 4.10 and Figure 4.11, we see how the distances we computed above can be viewed as areas: $A_1 = 4.5$ comes from taking rate times time $(3 \cdot 1.5)$, as do A_2 and A_3 for the second and third rectangles. The big new issue is that while A_2 is an area (and is therefore positive), because this area involves an interval on which the velocity function is negative, its area has a negative sign associated with it. This helps us to distinguish between distance traveled and change in position.

The distance traveled is the sum of the areas,

$$D = A_1 + A_2 + A_3 = 4.5 + 2 + 3 = 9.5 \text{ miles.}$$

But the change in position has to account for the sign associated with the area, where those above the t-axis are considered positive while those below the t-axis are viewed as negative, so that

$$s(3) - s(0) = (+4.5) + (-2) + (+3) = 5.5 \text{ miles,}$$

assigning the "-2" to the area in the interval $[1.5, 2]$ because there velocity is negative and the person is walking in the "negative" direction. In other words, the person walks 4.5 miles in the positive direction, followed by two miles in the negative direction, and then 3 more miles in the positive direction. This affect of velocity being negative is also seen in the graph of the function $y = s(t)$, which has a negative slope (specifically, its slope is -4) on the interval $1.5 < t < 2$ since the velocity is -4 on that interval, which shows the person's position function is decreasing due to the fact that she is walking west, rather than

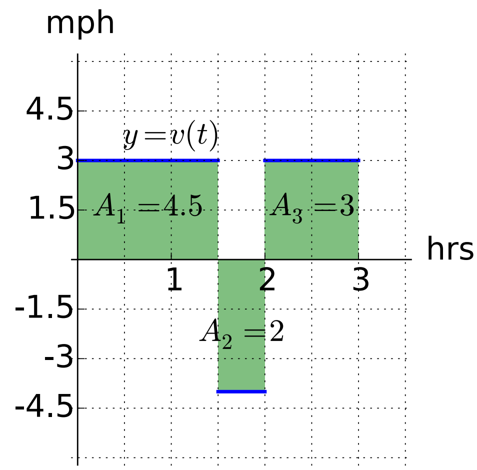


Figure 4.10: The velocity function of the person walking.

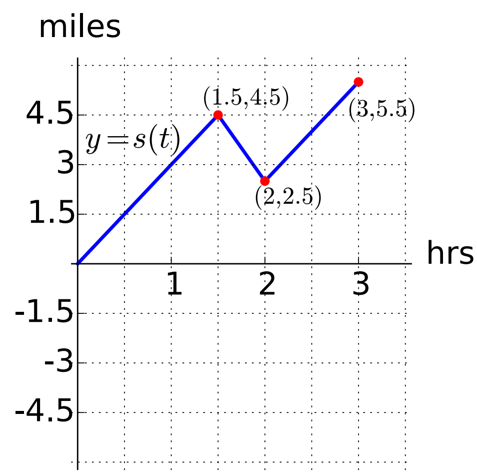


Figure 4.11: The position function of the person walking.

east. On the intervals where she is walking east, the velocity function is positive and the slope of the position function s is therefore also positive.

To summarize, we see that if velocity is sometimes negative, this makes the moving object's change in position different from its distance traveled. By viewing the intervals on which velocity is positive and negative separately, we may compute the distance traveled on each such interval, and then depending on whether we desire total distance traveled or total change in position, we may account for negative velocities that account for negative change in position, while still contributing positively to total distance traveled. We close this section with one additional activity that further explores the effects of negative velocity on the problem of finding change in position and total distance traveled.

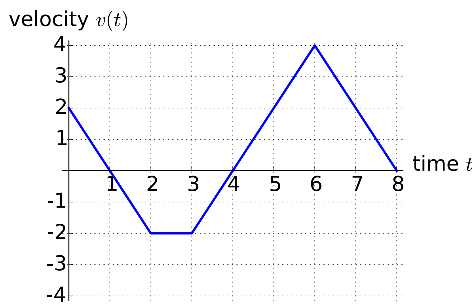


Figure 4.12: The velocity function of a moving object.

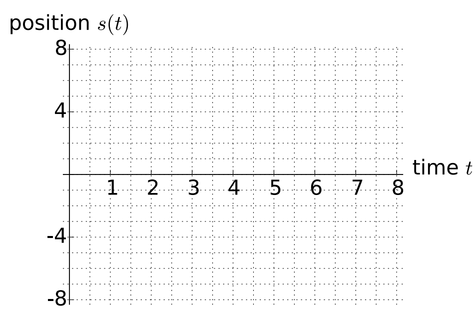


Figure 4.13: The position function of a moving object.

Activity 4.1–3

Suppose that an object moving along a straight line path has its velocity v (in meters per second) at time t (in seconds) given by the piecewise linear function whose graph is pictured in Figure 4.12. We view movement to the right as being in the positive direction (with positive velocity), while movement to the left is in the negative direction.

Suppose further that the object's initial position at time $t = 0$ is $s(0) = 1$.

(a) Determine the total distance traveled and the total change in position on the time interval $0 \le t \le 2$. What is the object's position at $t = 2$?

(b) On what time intervals is the moving object's position function increasing? Why? On what intervals is the object's position decreasing? Why?

(c) What is the object's position at $t = 8$? How many total meters has it traveled to get to this point (including distance in both directions)? Is this different from the object's total change in position on $t = 0$ to $t = 8$?

(d) Find the exact position of the object at $t = 1, 2, 3, \ldots, 8$ and use this data to sketch an accurate graph of $y = s(t)$ on the axes provided. How can you use the provided information about $y = v(t)$ to determine the concavity of s on each relevant interval?

Summary

In this section, we encountered the following important ideas:

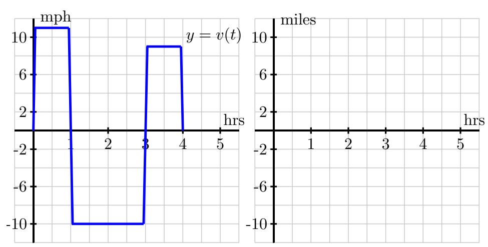
- If we know the velocity of a moving body at every point in a given interval and the velocity is positive throughout, we can estimate the object's distance traveled and in some circumstances determine this value exactly.
- In particular, when velocity is positive on an interval, we can find the total distance traveled by finding the area under the velocity curve and above the t-axis on the given time interval. We may only be able to estimate this area, depending on the shape of the velocity curve.
- The antiderivative of a function f is a new function F whose derivative is f. That is, F is an antiderivative of f provided that $F' = f$. In the context of velocity and position, if we know a velocity function v, an antiderivative of v is a position function s that satisfies $s' = v$. If v is positive on a given interval, say $[a, b]$, then the change in position, $s(b) - s(a)$, measures the distance the moving object traveled on $[a, b]$.
- In the setting where velocity is sometimes negative, this means that the object is sometimes traveling in the opposite direction (depending on whether velocity is positive or negative), and thus involves the object backtracking. To determine distance traveled, we have to think about the problem separately on intervals where velocity is positive and negative and account for the change in position on each such interval.

Exercises

Problems

1) Along the eastern shore of Lake Michigan from Lake Macatawa (near Holland) to Grand Haven, there is a bike bath that runs almost directly north-south. For the purposes of this problem, assume the road is completely straight, and that the function $s(t)$ tracks the position of the biker along this path in miles north of Pigeon Lake, which lies roughly halfway between the ends of the bike path.

   Suppose that the biker's velocity function is given by the graph below on the time interval $0 \le t \le 4$ (where t is measured in hours), and that $s(0) = 1$.

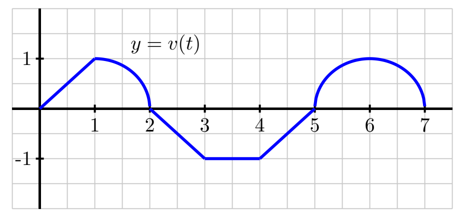
   

   (a) Approximately how far north of Pigeon Lake was the cyclist when she was the greatest distance away from Pigeon Lake? At what time did this occur?
   (b) What is the cyclist's total change in position on the time interval $0 \le t \le 2$? At $t = 2$, was she north or south of Pigeon Lake?
   (c) What is the total distance the biker traveled on $0 \le t \le 4$? At the end of the ride, how close was she to the point at which she started?
   (d) Sketch an approximate graph of $y = s(t)$, the position function of the cyclist, on the interval $0 \le t \le 4$. Label at least four important points on the graph of s.

2) A toy rocket is launched vertically from the ground on a day with no wind. The rocket's vertical velocity at time t (in seconds) is given by $v(t) = 500 - 32t$ feet/sec.

   (a) At what time after the rocket is launched does the rocket's velocity equal zero? Call this time value a. What happens to the rocket at $t = a$?
   (b) Find the value of the total area enclosed by $y = v(t)$ and the t-axis on the interval $0 \le t \le a$. What does this area represent in terms of the physical setting of the problem?
   (c) Find a function s such that $s'(t) = v(t)$, or an antiderivative of $v(t)$.
   (d) Compute the value of $s(a) - s(0)$. What does this number represent in terms of the physical setting of the problem?
   (e) Compute $s(5) - s(1)$. What does this number tell you about the rocket's flight?

3) An object moving along a horizontal axis has its instantaneous velocity at time t in seconds given by the function v pictured below, where v is measured in feet/sec.

   Assume that the curves that make up the parts of the graph of $y = v(t)$ are either portions of straight lines or portions of circles.

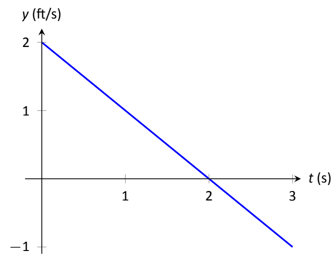
   (a) Determine the exact total distance the object traveled on $0 \le t \le 2$.
   (b) What is the value and meaning of $s(5) - s(2)$, where $y = s(t)$ is the position function of the moving object?
   (c) On which time interval did the object travel the greatest distance: $[0, 2]$, $[2, 4]$, or $[5, 7]$?
   (d) On which time interval(s) is the position function s increasing? At which point(s) does s achieve a relative maximum?

4) Filters at a water treatment plant become dirtier over time and thus become less effective; they are replaced every 30 days. During one 30-day period, the rate, $p(t)$, at which pollution passes through the filters into a nearby lake (in units of particulate matter per day) is measured every 6 days and is given in the following table. The time t is measured in days since the filters were replaced.

Day, t	0	6	12	18	24	30
$p(t)$	7	8	10	13	18	35

   (a) Plot the given data on a set of axes with time on the horizontal axis and the rate of pollution on the vertical axis.
   (b) Explain why the amount of pollution that entered the lake during this 30-day period would be given exactly by the area bounded by $y = p(t)$ and the t-axis on the time interval $[0, 30]$.
   (c) Estimate the total amount of pollution entering the lake during this 30-day period. Carefully explain how you determined your estimate.

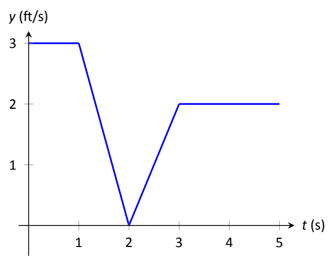
In Exercises 5–6, a graph of the velocity function of an object moving in a straight line is given. Answer the questions based on the graph.

5) 

(a) What is the object's maximum velocity?

(b) What is the object's maximum displacement?

(c) What is the object's total displacement on $[0, 3]$?

6) 

(a) What is the object's maximum velocity?

(b) What is the object's maximum displacement?

(c) What is the object's total displacement on $[0, 5]$?

7) An object is thrown straight up with a velocity, in ft/s, given by $v(t) = -32t + 64$, where t is in seconds, from a height of 48 feet.

(a) What is the object's maximum velocity?

(b) What is the object's maximum displacement?

(c) When does the maximum displacement occur?

(d) When will the object reach a height of 0? (Hint: find when the displacement is -48ft.)

8) An object is thrown straight up with a velocity, in ft/s, given by $v(t) = -32t + 96$, where t is in seconds, from a height of 64 feet.

(a) What is the object's initial velocity?

(b) When is the object's displacement 0?

(c) How long does it take for the object to return to its initial height?

(d) When will the object reach a height of 210 feet?

4.2 *Riemann Sums*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- How can we use a Riemann sum to estimate the area between a given curve and the horizontal axis over a particular interval?
- What are the differences among left, right, middle, and random Riemann sums?
- What is sigma notation and how does this enable us to write Riemann sums in an abbreviated form?

Introduction

In Section 4.1, we learned that if we have a moving object with velocity function v, whenever $v(t)$ is positive, the area between $y = v(t)$ and the t-axis over a given time interval tells us the distance traveled by the object over that time period; in addition, if $v(t)$ is sometimes negative and we view the area of any region below the t-axis as having an associated negative sign, then the sum of these signed areas over a given interval tells us the moving object's change in position over the time interval.

For instance, for the velocity function given in Figure 4.14, if the areas of shaded regions are A_1, A_2, and A_3 as labeled, then the total distance D traveled by the moving object on $[a, b]$ is

$$D = A_1 + A_2 + A_3,$$

while the total change in the object's position on $[a, b]$ is

$$s(b) - s(a) = A_1 - A_2 + A_3.$$

Because the motion is in the negative direction on the interval where $v(t) < 0$, we subtract A_2 when determining the object's total change in position.

Of course, finding D and $s(b) - s(a)$ for the situation given in Figure 4.14 presumes that we can actually find the areas represented by A_1, A_2, and A_3. In most of our work in Section 4.1, such as in Activities 4.1–2 and 4.1–3, we worked with velocity functions that were either constant or linear, so that by finding the areas of rectangles and triangles, we could find the area bounded by the velocity function and the horizontal axis exactly. But when the curve that bounds a region is not one for which we have a known formula for area, we are unable to find this area exactly. Indeed, this is one of our biggest goals in Chapter 4: to learn how to find the exact area bounded between a curve and

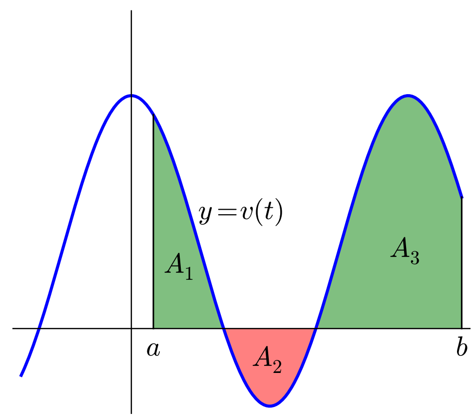


Figure 4.14: A velocity function that is sometimes negative.

the horizontal axis for as many different types of functions as possible.

To begin, we expand on the ideas in Activity 4.1–1, where we encountered a nonlinear velocity function and approximated the area under the curve using four and eight rectangles, respectively. In the following preview activity, we focus on three different options for deciding how to find the heights of the rectangles we will use.

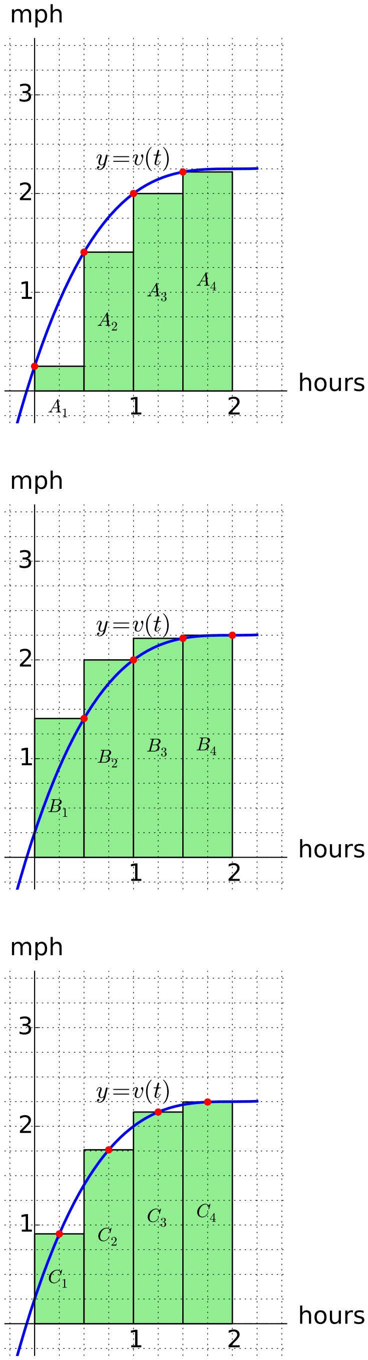


Figure 4.15: Estimating the area under $y = v(t)$.

Preview Activity 4.2

A person walking along a straight path has her velocity in miles per hour at time t given by the function $v(t) = 0.25t^3 - 1.5t^2 + 3t + 0.25$, for times in the interval $0 \leq t \leq 2$. The graph of this function is also given in each of the three diagrams. Note that in each diagram, we use four rectangles to estimate the area under $y = v(t)$ on the interval $[0, 2]$, but the method by which the four rectangles' respective heights are decided varies among the three individual graphs.

(a) How are the heights of rectangles in the left-most diagram being chosen? Explain, and hence determine the value of

$$S = A_1 + A_2 + A_3 + A_4$$

by evaluating the function $y = v(t)$ at appropriately chosen values and observing the width of each rectangle. Note, for example, that

$$A_3 = v(1) \cdot \frac{1}{2} = 2 \cdot \frac{1}{2} = 1.$$

(b) Explain how the heights of rectangles are being chosen in the middle diagram and find the value of

$$T = B_1 + B_2 + B_3 + B_4.$$

(c) Likewise, determine the pattern of how heights of rectangles are chosen in the right-most diagram and determine

$$U = C_1 + C_2 + C_3 + C_4.$$

(d) Of the estimates S, T, and U, which do you think is the best approximation of D, the total distance the person traveled on $[0, 2]$? Why?

Example 1

Approximate the area of the region under $y = 4x - x^2$ on the interval $[0, 4]$ using L_4, R_4, and M_4.

Solution. We partition the interval $[0, 4]$ into the four subintervals $[0, 1]$, $[1, 2]$, $[2, 3]$, and $[3, 4]$. In Figure 4.16-(a) we see 4 rectangles drawn on $f(x) = 4x - x^2$ using the L_4. Note how in the first subinterval, $[0, 1]$, the rectangle has height $f(0) = 0$. We add up the areas of each rectangle, which is width × height, for our L_4 approximation:

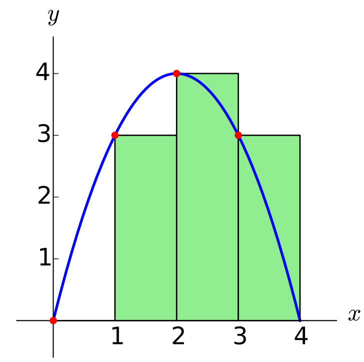
$$\begin{aligned} L_4 &= 1\cdot f(0)+1\cdot f(1)+1\cdot f(2)+1\cdot f(3) \\ &= 0+3+4+3 \\ &= 10. \end{aligned}$$

Figure 4.16-(b) shows 4 rectangles drawn under f using R_4; note how the rectangle on the subinterval $[3,4]$ has height 0. Our approximation gives the same answer as before, though calculated a different way:

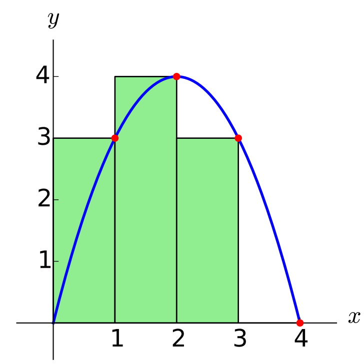
$$\begin{aligned} R_4 &= 1\cdot f(1)+1\cdot f(2)+1\cdot f(3)+1\cdot f(4) \\ &= 3+4+3+0 \\ &= 10. \end{aligned}$$

Figure 4.16-(c) shows 4 rectangles drawn under f using M_4. Notice that the height of each of the rectangles is determined by the point in the middle of each subinterval, which are 0.5, 1.5, 2.5, and 3.5. So

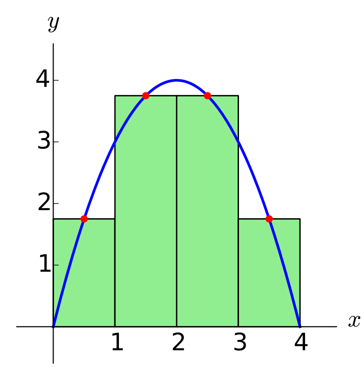
$$\begin{aligned} M_4 &= 1\cdot f(0.5)+1\cdot f(1.5)+1\cdot f(2.5)+1\cdot f(3.5) \\ &= 1.75+3.75+3.75+1.75 \\ &= 11. \end{aligned}$$



(a) L_n



(b) R_n



(c) M_n

Figure 4.16: Approximating the area under $y = 4x - x^2$ on $[0,4]$ using L_n, R_n, and M_n

Sigma Notation

It is apparent from several different problems we have considered that sums of areas of rectangles is one of the main ways to approximate the area under a curve over a given interval. Intuitively, we expect that using a larger number of thinner rectangles will provide a way to improve the estimates we are computing. As such, we anticipate dealing with sums with a large number of terms. To do so, we introduce the use of so-called *sigma notation*, named for the Greek letter Σ, which is the capital letter S in the Greek alphabet.

For example, say we are interested in the sum

$$1+2+3+\cdots+100,$$

which is the sum of the first 100 natural numbers. Sigma notation provides a shorthand notation that recognizes the general pattern in the terms of the sum. It is equivalent to write

$$\sum_{k=1}^{100} k = 1+2+3+\cdots+100.$$

We read the symbol $\sum\limits_{k=1}^{100} k$ as "the sum from k equals 1 to 100 of k." The variable k is usually called the index of summation, and the letter that is used for this variable is immaterial.

Suppose we wish to add up a list of numbers $a_1, a_2, a_3, \ldots, a_9$. Instead of writing $a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8 + a_9$, we use sigma notation and write

$$\sum_{i=1}^{9} a_i.$$

Again, the capital sigma represents the term "sum." The index of summation in this example is i. By convention, the index takes on only the integer values between (and including) the lower and upper bounds.

Each summand in sigma notation involves a function of the index; for example,

$$\begin{aligned}\sum_{k=1}^{10}(k^2+2k) &= (1^2+2\cdot 1)+(2^2+2\cdot 2)+(3^2+2\cdot 3)\\ &\quad +\cdots+(10^2+2\cdot 10),\end{aligned}$$

and more generally,

$$\sum_{k=1}^{n} f(k) = f(1) + f(2) + \cdots + f(n).$$

Sigma notation allows us the flexibility to easily vary the function being used to track the pattern in the sum, as well as to adjust the number of terms in the sum simply by changing the value of n. We test our understanding of this new notation in the following activity.

Activity 4.2–1

For each sum written in sigma notation, write the sum long-hand and evaluate the sum to find its value. For each sum written in expanded form, write the sum in sigma notation.

(a) $\displaystyle\sum_{k=1}^{5}(k^2+2)$

(b) $\displaystyle\sum_{i=3}^{6}(2i-1)$

(c) $3+7+11+15+\cdots+27$

(d) $4+8+16+32\cdots+256$

(e) $\displaystyle\sum_{i=1}^{6}\frac{1}{2^i}$

It might seem odd to stress a new, concise way of writing summations only to write each term out as we add them up. It is. Therefore, we give some properties of summations that allow us to work with them without writing individual terms.

Properties of Summations

1) $\sum_{i=1}^{n} c = c \cdot n$, where c is a constant.

2) $\sum_{i=1}^{n} c \cdot a_i = c \cdot \sum_{i=1}^{n} a_i$

3) $\sum_{i=m}^{n} (a_i \pm b_i) = \sum_{i=m}^{n} a_i \pm \sum_{i=m}^{n} b_i$

4) $\sum_{i=m}^{j} a_i + \sum_{i=j+1}^{n} a_i = \sum_{i=m}^{n} a_i$

5) $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$

6) $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$

7) $\sum_{i=1}^{n} i^3 = \left(\frac{n(n+1)}{2}\right)^2$

Example 2

Evaluate $\sum_{i=1}^{6}(2i-1)$.

Solution.

$$\begin{aligned}
\sum_{i=1}^{6}(2i-1) &= \sum_{i=1}^{6} 2i - \sum_{i=1}^{6} 1 \\
&= 2\sum_{i=1}^{6} i - \sum_{i=1}^{6} 1 \\
&= 2\left(\frac{6(6+1)}{2}\right) - 6 \\
&= 42 - 6 \\
&= 36
\end{aligned}$$

We obtained the same answer without writing out all six terms. When dealing with small sizes of n, it may be faster to write the terms out by hand. However, the Properties of Summations are incredibly important when dealing with large sums as we'll soon see.

Riemann Sums

When a moving body has a positive velocity function $y = v(t)$ on a given interval $[a, b]$, we know that the area under the curve over the interval is the total distance the body travels on $[a, b]$. While this is the fundamental motivating force behind our interest in the area bounded by a function, we are also interested more generally in being able to find the exact area bounded by $y = f(x)$ on an interval $[a, b]$, regardless of the meaning or context of the function f. For now, we continue to focus on determining an accurate estimate of this area through the use of a sum of the areas of rectangles, doing so in the setting where $f(x) \geq 0$ on $[a, b]$. Throughout, unless otherwise indicated, we also assume that f is continuous on $[a, b]$.

The first choice we make in any such approximation is the number of rectangles. If we say that the total number of rectangles is n, and we desire n rectangles of equal width to subdivide the interval $[a, b]$, then each rectangle must have width $\triangle x = \frac{b-a}{n}$. We observe further that $x_1 = x_0 + \triangle x$, $x_2 = x_0 + 2\triangle x$, and thus in general $x_i = a + i\triangle x$, as pictured in Figure 4.17.

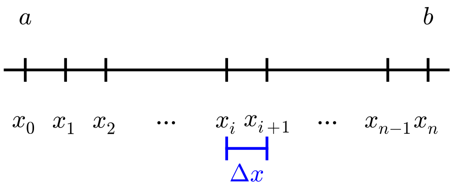


Figure 4.17: Subdividing the interval $[a, b]$ into n subintervals of equal length Δx.

We use each subinterval $[x_i, x_{i+1}]$ as the base of a rectangle, and next must choose how to decide the height of the rectangle that will be used to approximate the area under $y = f(x)$ on the subinterval. There are three standard choices: use the left endpoint of each subinterval, the right endpoint of each subinterval, or the midpoint of each. These are precisely the options encountered in Preview Activity 4.2 and seen in Figure 4.14. We next explore how these choices can be reflected in sigma notation.

If we now consider an arbitrary positive function f on $[a, b]$ with the interval subdivided as shown in Figure 4.17, and choose to use left endpoints, then on each interval of the form $[x_i, x_{i+1}]$, the area of the rectangle formed is given by

$$A_{i+1} = f(x_i) \cdot \triangle x,$$

as seen in Figure 4.18.

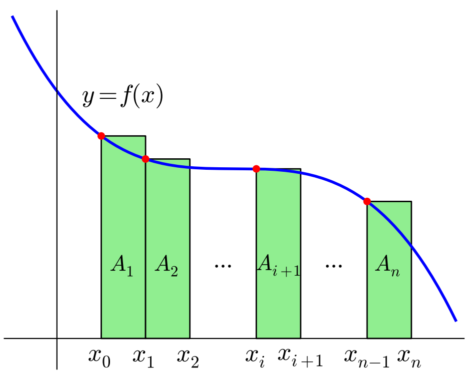


Figure 4.18: Subdividing the interval $[a, b]$ into n subintervals of equal length $\triangle x$ and approximating the area under $y = f(x)$ over $[a, b]$ using left rectangles.

If we let L_n denote the sum of the areas of rectangles whose heights are given by the function value at each respective left endpoint, then we see that

$$\begin{aligned} L_n &= A_1 + A_2 + \cdots + A_{i+1} + \cdots + A_n \\ &= f(x_0) \cdot \triangle x + f(x_1) \cdot \triangle x + \cdots + f(x_i) \cdot \triangle x \\ &\quad + \cdots + f(x_{n-1}) \cdot \triangle x. \end{aligned}$$

In the more compact sigma notation, we have

$$L_n = \sum_{i=0}^{n-1} f(x_i) \triangle x.$$

Note particularly that since the index of summation begins at 0 and ends at $n - 1$, there are indeed n terms in this sum. We call L_n the *left Riemann sum* for the function f on the interval $[a, b]$.

There are now two fundamental issues to explore: the number of rectangles we choose to use and the selection of the pattern by which we identify the height of each rectangle. It is best to explore these choices dynamically, and the applet found at `http://gvsu.edu/s/a9` is a particularly useful one. There we see the image shown in Figure 4.19, but with the opportunity to adjust the slider bars for the left endpoint and the number of subintervals. By moving the sliders, we can see how the heights of the rectangles change as we consider left endpoints, midpoints, and right endpoints, as well as the impact that a larger number of narrower rectangles has on the approximation of the exact area bounded by the function and the horizontal axis.

To see how the Riemann sums for right endpoints and midpoints are constructed, we consider Figure 4.20 and Figure 4.21. For the sum with right endpoints, we see that the area of the rectangle on an arbitrary interval $[x_i, x_{i+1}]$ is given by

$$B_{i+1} = f(x_{i+1}) \cdot \triangle x,$$

so that the sum of all such areas of rectangles is given by

$$\begin{aligned} R_n &= B_1 + B_2 + \cdots + B_{i+1} + \cdots + B_n \\ &= f(x_1) \cdot \triangle x + f(x_2) \cdot \triangle x + \cdots + f(x_{i+1}) \cdot \triangle x \\ &\quad + \cdots + f(x_n) \cdot \triangle x \\ &= \sum_{i=1}^{n} f(x_i) \triangle x. \end{aligned}$$

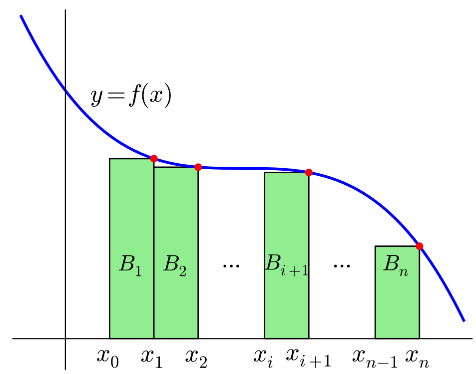


Figure 4.20: Riemann sums using right endpoints.

We call R_n the *right Riemann sum* for the function f on the interval $[a, b]$.

For the sum that uses midpoints, we introduce the notation

$$\overline{x}_{i+1} = \frac{x_i + x_{i+1}}{2}$$

so that $\overline{x}_{i+1}$ is the midpoint of the interval $[x_i, x_{i+1}]$. For instance, for the rectangle with area C_1 in Figure 4.21, we now have

$$C_1 = f(\overline{x}_1) \cdot \triangle x.$$

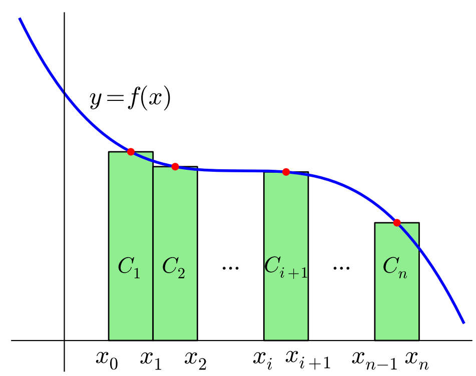


Figure 4.21: Riemann sums using midpoints.

Hence, the sum of all the areas of rectangles that use midpoints is

$$\begin{aligned} M_n &= C_1 + C_2 + \cdots + C_{i+1} + \cdots + C_n \\ &= f(\overline{x_1}) \cdot \triangle x + f(\overline{x_2}) \cdot \triangle x + \cdots + f(\overline{x}_{i+1}) \cdot \triangle x \\ &\quad + \cdots + f(\overline{x}_n) \cdot \triangle x \\ &= \sum_{i=1}^{n} f(\overline{x}_i) \triangle x, \end{aligned}$$

and we say that M_n is the *middle Riemann sum* for f on $[a,b]$.

When $f(x) \geq 0$ on $[a,b]$, each of the Riemann sums L_n, R_n, and M_n provides an estimate of the area under the curve $y = f(x)$ over the interval $[a,b]$; momentarily, we will discuss the meaning of Riemann sums in the setting when f is sometimes negative. We also recall that in the context of a nonnegative velocity function $y = v(t)$, the corresponding Riemann sums are approximating the distance traveled on $[a,b]$ by the moving object with velocity function v.

There is a more general way to think of Riemann sums, and that is to not restrict the choice of where the function is evaluated to determine the respective rectangle heights. That is, rather than saying we'll always choose left endpoints, or always choose midpoints, we simply say that a point x^*_{i+1} will be selected at random in the interval $[x_i, x_{i+1}]$ (so that $x_i \leq x^*_{i+1} \leq x_{i+1}$), which makes the Riemann sum given by

$$\begin{aligned}\sum_{i=1}^{n} f(x_i^*)\triangle x &= f(x_1^*)\cdot\triangle x + f(x_2^*)\cdot\triangle x + \cdots + f(x_{i+1}^*)\cdot\triangle x \\ &\quad + \cdots + f(x_n^*)\cdot\triangle x.\end{aligned}$$

At `http://gvsu.edu/s/a9`, the applet noted earlier and referenced in Figure 4.19, by unchecking the "relative" box at the top left, and instead checking "random," we can easily explore the effect of using random point locations in subintervals on a given Riemann sum. In computational practice, we most often use L_n, R_n, or M_n, while the random Riemann sum is useful in theoretical discussions. In the following activity, we investigate several different Riemann sums for a particular velocity function.

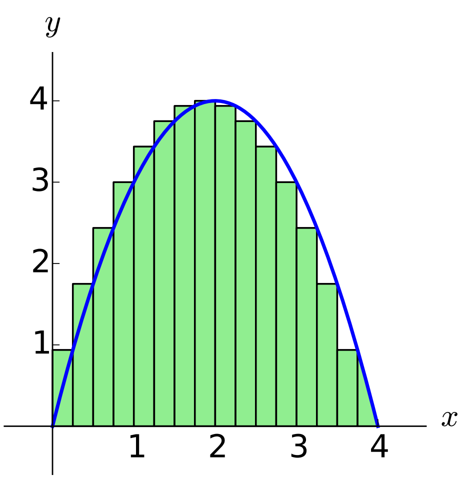


Figure 4.22: Approximating the area under $y = 4x - x^2$ on $[0,4]$ using R_n with $n = 16$.

Example 3

Approximate the area under $y = 4x - x^2$ on $[0,4]$ using R_n and summation formulas with 16 and 1000 equally spaced intervals.

Solution. Using 16 rectangles, we can approximate the area as

$$\sum_{i=1}^{16} f(x_i)\Delta x$$

where $\Delta x = 4/16 = 1/4$ and $x_i = 0 + i\Delta x = i\Delta x$. Using the properties of summations, we have

$$
\begin{aligned}
R_{16} &\approx \sum_{i=1}^{16} f(x_i)\Delta x \\
&= \sum_{i=1}^{16} f(i\Delta x)\Delta x \\
&= \sum_{i=1}^{16} \left(4i\Delta x - (i\Delta x)^2\right)\Delta x \\
&= \sum_{i=1}^{16} \left(4i\Delta x^2 - i^2\Delta x^3\right) \\
&= \left(4\Delta x^2\right)\sum_{i=1}^{16} i - \left(\Delta x^3\right)\sum_{i=1}^{16} i^2 \qquad (4.1) \\
&= (4\Delta x^2)\frac{16\cdot 17}{2} - \left(\Delta x^3\right)\frac{16(17)(33)}{6} \\
&= 4\cdot\left(\frac{1}{4}\right)^2\cdot\frac{16\cdot 17}{2} - \left(\frac{1}{4}\right)^3\cdot\frac{16(17)(33)}{6} \\
&= 2\cdot 17 - \frac{17\cdot 11}{8} \\
&= \frac{85}{8} = 10.625
\end{aligned}
$$

We were able to sum up the areas of 16 rectangles with very little computation. Notice Equation (4.1); by changing the 16's to 1,000's (and appropriately changing the value of Δx to $4/1000$), we can use that equation to sum up 1000 rectangles!

$$
\begin{aligned}
R_{1000} &\approx \sum_{i=1}^{1000} f(x_i)\Delta x \\
&= \left(4\Delta x^2\right)\sum_{i=1}^{1000} i - \left(\Delta x^3\right)\sum_{i=1}^{1000} i^2 \\
&= \left(4\Delta x^2\right)\frac{1000\cdot 1001}{2} - \left(\Delta x^3\right)\frac{1000(1001)(2001)}{6} \\
&= 4\cdot(0.004)^2\cdot 500500 - (0.004)^3\cdot 333,833,500 \\
&= 10.666656
\end{aligned}
$$

Using this many rectangles, it's likely that we have a good approximation of the area under $y = 4x - x^2$ on $[0,4]$. That is, the area is approximately 10.666656.

Activity 4.2–2

Suppose that an object moving along a straight line path has its velocity in feet per second at time t in seconds given by $v(t) = \frac{2}{9}(t-3)^2 + 2$.

(a) Carefully sketch the region whose exact area will tell you the value of the distance the object traveled on the time interval $2 \le t \le 5$.

(b) Estimate the distance traveled on $[2,5]$ by computing L_4, R_4, and M_4.

(c) Does averaging L_4 and R_4 result in the same value as M_4? If not, what do you think the average of L_4 and R_4 measures?

(d) For this question, think about an arbitrary function f, rather than the particular function v given above. If f is positive and increasing on $[a,b]$, will L_n over-estimate or under-estimate the exact area under f on $[a,b]$? Will R_n over- or under-estimate the exact area under f on $[a,b]$? Explain.

When the function is sometimes negative

For a Riemann sum such as

$$L_n = \sum_{i=0}^{n-1} f(x_i)\triangle x,$$

we can of course compute the sum even when f takes on negative values. We know that when f is positive on $[a,b]$, the corresponding left Riemann sum L_n estimates the area bounded by f and the horizontal axis over the interval.

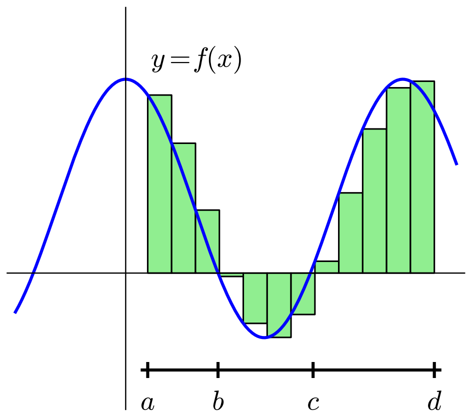


Figure 4.23: A left Riemann sum for a function f that is sometimes negative.

For a function such as the one pictured in Figure 4.23, where in the figure a left Riemann sum is being taken with 12 subintervals over $[a,d]$, we observe that the function is negative on the interval $b \le x \le c$, and so for the four left endpoints that fall in $[b,c]$, the terms $f(x_i)\triangle x$ have negative function values. This means that those four terms in the Riemann sum produce an estimate of the *opposite* of the area bounded by $y = f(x)$ and the x-axis on $[b,c]$.

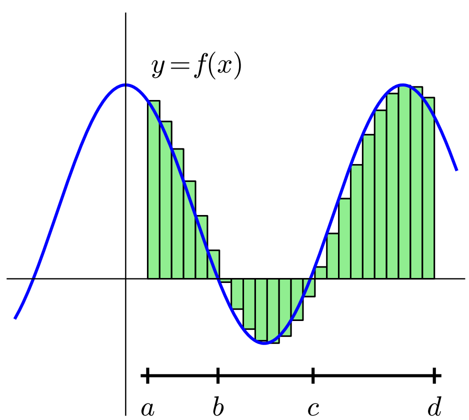


Figure 4.24: A left Riemann sum using more rectangles for a function f that is sometimes negative.

In Figure 4.24, we also see evidence that by increasing the number of rectangles used in a Riemann sum, it appears that the approximation of the area (or the opposite of the area) bounded by a curve appears to improve. For instance, we use 24 left rectangles, and from the shaded areas, it appears that we have decreased the error from the approximation that uses 12. When we proceed to Section 4.3, we will discuss the natural idea of letting the number of rectangles in the sum increase without bound.

For now, it is most important for us to observe that, in general, any Riemann sum of a continuous function f on an interval $[a,b]$ approximates the difference between the area that lies above the horizontal axis on $[a,b]$ and under f and the area that lies below the horizontal axis on $[a,b]$ and above f. In the notation of Figure 4.25, we may say that

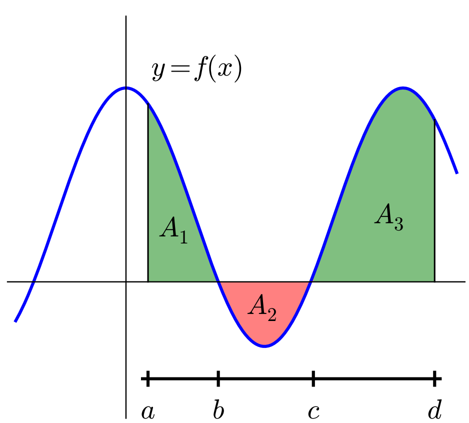


Figure 4.25: The areas bounded by f on the interval $[a,d]$.

$$L_{24} \approx A_1 - A_2 + A_3,$$

where L_{24} is the left Riemann sum using 24 subintervals shown in the middle graph, and A_1 and A_3 are the areas of the regions where f is positive on the interval of interest, while A_2 is the area of the region where f is negative. We will also call the quantity $A_1 - A_2 + A_3$ the *net signed area* bounded by f over the interval $[a, d]$, where by the phrase "signed area" we indicate that we are attaching a minus sign to the areas of regions that fall below the horizontal axis.

Finally, we recall from the introduction to this present section that in the context where the function f represents the velocity of a moving object, the total sum of the areas bounded by the curve tells us the total distance traveled over the relevant time interval, while the total net signed area bounded by the curve computes the object's change in position on the interval.

Example 4

Approximate the net area between $f(x) = 5x + 2$ and the x-axis on $[-2, 3]$ using the Midpoint Rule and 10 equally spaced intervals.

Solution. Beginning with Δx and x_i, we have

$$\Delta x = \frac{3-(-2)}{10} = \frac{1}{2} \quad \text{and} \quad x_i = -2 + \frac{1}{2} \cdot i = \frac{i}{2} - 2.$$

As we are using the Midpoint Rule, we will also need x_{i-1} and $\dfrac{x_{i-1} + x_i}{2}$. Since $x_i = \dfrac{i}{2} - 2$, $x_{i-1} = \dfrac{i-1}{2} - 2 = \dfrac{i}{2} - \dfrac{5}{2}$. Then

$$\frac{x_{i-1} + x_i}{2} = \frac{\frac{i}{2} - \frac{5}{2} + \frac{i}{2} - 2}{2} = \frac{i - \frac{9}{2}}{2} = \frac{i}{2} - \frac{9}{4}.$$

We now construct the Riemann sum and compute its value using summation formulas.

$$\begin{aligned} M_{10} &\approx \sum_{i=1}^{10} f\left(\frac{x_{i-1} + x_i}{2}\right) \Delta x \\ &= \sum_{i=1}^{10} f\left(\frac{i}{2} - \frac{9}{4}\right) \Delta x \\ &= \sum_{i=1}^{10} \left(5\left(\frac{i}{2} - \frac{9}{4}\right) + 2\right) \Delta x \\ &= \Delta x \sum_{i=1}^{10} \left[\left(\frac{5}{2}\right) i - \frac{37}{4}\right] \\ &= \Delta x \left(\frac{5}{2} \sum_{i=1}^{10} (i) - \sum_{i=1}^{10} \left(\frac{37}{4}\right)\right) \\ &= \frac{1}{2}\left(\frac{5}{2} \cdot \frac{10(11)}{2} - 10 \cdot \frac{37}{4}\right) \\ &= \frac{45}{2} = 22.5 \end{aligned}$$

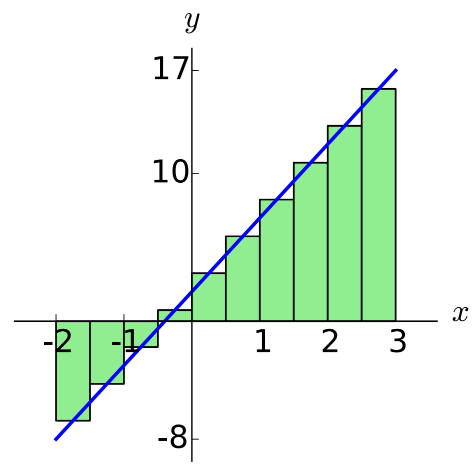


Figure 4.26: Approximating net signed area of $f(x) = 5x + 2$ using the Midpoint Rule and 10 evenly spaced subintervals.

Activity 4.2–3

Suppose that an object moving along a straight line path has its velocity v (in feet per second) at time t (in seconds) given by

$$v(t) = \frac{1}{2}t^2 - 3t + \frac{7}{2}.$$

(a) Compute M_5, the middle Riemann sum, for v on the time interval $[1,5]$. Be sure to clearly identify the value of $\triangle t$ as well as the locations of t_0, t_1, $\cdots$, t_5. In addition, provide a careful sketch of the function and the corresponding rectangles that are being used in the sum.

(b) Building on your work in (a), estimate the total change in position of the object on the interval $[1,5]$.

(c) Building on your work in (a) and (b), estimate the total distance traveled by the object on $[1,5]$.

(d) Use appropriate computing technology (For instance, consider the applet at `http://gvsu.edu/s/a9` and change the function and adjust the locations of the blue points that represent the interval endpoints a and b.) to compute M_{10} and M_{20}. What exact value do you think the middle sum eventually approaches as n increases without bound? What does that number represent in the physical context of the overall problem?

Summary

In this section, we encountered the following important ideas:

- A Riemann sum is simply a sum of products of the form $f(x_i^*)\triangle x$ that estimates the area between a positive function and the horizontal axis over a given interval. If the function is sometimes negative on the interval, the Riemann sum estimates the difference between the areas that lie above the horizontal axis and those that lie below the axis.
- The three most common types of Riemann sums are left, right, and middle sums, plus we can also work with a more general, random Riemann sum. The only difference among these sums is the location of the point at which the function is evaluated to determine the height of the rectangle whose area is being computed in the sum. For a left Riemann sum, we evaluate the function at the left endpoint of each subinterval, while for right and middle sums, we use right endpoints and midpoints, respectively.
- The left, right, and middle Riemann sums are denoted L_n, R_n, and M_n, with formulas

$$L_n = f(x_0)\triangle x + f(x_1)\triangle x + \cdots + f(x_{n-1})\triangle x = \sum_{i=0}^{n-1} f(x_i)\triangle x,$$

$$R_n = f(x_1)\triangle x + f(x_2)\triangle x + \cdots + f(x_n)\triangle x = \sum_{i=1}^{n} f(x_i)\triangle x,$$

$$M_n = f(\overline{x}_1)\triangle x + f(\overline{x}_2)\triangle x + \cdots + f(\overline{x}_n)\triangle x = \sum_{i=1}^{n} f(\overline{x}_i)\triangle x,$$

where $x_0 = a$, $x_i = a + i\triangle x$, and $x_n = b$, using $\triangle x = \frac{b-a}{n}$. For the midpoint sum, $\overline{x}_i = (x_{i-1} + x_i)/2$.

Exercises

Terms and Concepts

1) A fundamental calculus technique is to use ________ to refine approximations to get an exact answer.

2) What is the upper bound in the summation $\sum_{i=7}^{14}(48i - 201)$?

3) This section approximates using what geometric shape?

4) T/F: A sum using R_n is an example of a Riemann Sum.

Problems

In exercises 5–11, write out each term of the summation have compute the sum.

5) $\sum_{i=2}^{4} i^2$

6) $\sum_{i=-1}^{3} (4i-2)$

7) $\sum_{i=-2}^{2} \sin(\pi i/2)$

8) $\sum_{i=1}^{5} \frac{1}{i}$

9) $\sum_{i=1}^{6} (-1)^i i$

10) $\sum_{i=1}^{4} \left(\frac{1}{i} - \frac{1}{i+1}\right)$

11) $\sum_{i=0}^{5} (-1)^i \cos(\pi i)$

In exercises 12–15, write each sum in summation notation.

12) $3 + 6 + 9 + 12 + 15$

13) $-1 + 0 + 3 + 8 + 15 + 24 + 35 + 48 + 63$

14) $\frac{1}{2} + \frac{2}{3} + \frac{3}{4} + \frac{4}{5}$

15) $1 - e + e^2 - e^3 + e^4$

In exercises 16–25, evaluate the summation.

16) $\sum_{i=1}^{25} i$

17) $\sum_{i=1}^{10} (3i^2 - 2i)$

18) $\sum_{i=1}^{15} (2i^3 - 10)$

19) $\sum_{i=1}^{10} (i^3 - 3i^2 + 2i + 7)$

20) $\sum_{i=1}^{10} (-4i^3 + 10i^2 - 7i + 11)$

21) $1 + 4 + 9 + \ldots + 361 + 400$

22) $\sum_{i=11}^{20} i$

23) $\sum_{i=16}^{25} i^3$

24) $\sum_{i=7}^{12} 4$

25) $\sum_{i=5}^{10} 4i^3$

In exercises 26–31, approximate the area under the given function using the given conditions.

26) $f(x) = x^2$ on $[-3, 3]$ with 6 rectangles using L_n.

27) $f(x) = 5 - x^2$ on $[0, 2]$ with 4 rectangles using M_n.

28) $f(x) = \sin(x)$ on $[0, \pi]$ with 6 rectangles using R_n.

29) $f(x) = 2^x$ on $[0, 3]$ with 5 rectangles using L_n.

30) $f(x) = \ln(x)$ on $[1, 2]$ with 3 rectangles using M_n.

31) $f(x) = \frac{1}{x}$ on $[1, 9]$ with 4 rectangles using R_n.

In exercises 32–37,

(a) find a formula to approximate the area under the given function using n subintervals and the provided rule;

(b) evaluate the formula using $n = 10$, 100, and 1000; and

(c) find the limit of the formula as $n \to \infty$ to compute the exact value of the area under the given function.

32) $f(x) = x^3$ on $[0, 1]$ using R_n.

33) $f(x) = 3x^2$ on $[-1, 1]$ using L_n.

34) $f(x) = 3x - 1$ on $[-1, 3]$ using M_n.

35) $f(x) = 2x^2 - 3$ on $[1, 4]$ using L_n.

36) $f(x) = 5 - x$ on $[-10, 10]$ using R_n.

37) $f(x) = x^3 - x^2$ on $[0, 1]$ using R_n.

38) Consider the function $f(x) = 3x + 4$.

(a) Compute M_4 for $y = f(x)$ on the interval $[2, 5]$. Be sure to clearly identify the value of $\triangle x$, as well as the locations of $x_0, x_1, \ldots, x_4$. Include a careful sketch of the function and the corresponding rectangles being used in the sum.

(b) Use a familiar geometric formula to determine the exact value of the area of the region bounded by $y = f(x)$ and the x-axis on $[2, 5]$.

(c) Explain why the values you computed in (a) and (b) turn out to be the same. Will this be true if we use a number different than $n = 4$ and compute M_n? Will L_4 or R_4 have the same value as the exact area of the region found in (b)?

(d) Describe the collection of functions g for which it will always be the case that M_n, regardless of the value of n, gives the exact net signed area bounded between the function g and the x-axis on the interval $[a, b]$.

39) Let S be the sum given by

$$\begin{aligned} S =& ((1.4)^2 + 1) \cdot 0.4 + ((1.8)^2 + 1) \cdot 0.4 \\ &+ ((2.2)^2 + 1) \cdot 0.4 + ((2.6)^2 + 1) \cdot 0.4 \\ &+ ((3.0)^2 + 1) \cdot 0.4 \end{aligned}$$

(a) Assume that S is a right Riemann sum. For what function f and what interval $[a, b]$ is S an approximation of the area under f and above the x-axis on $[a, b]$? Why?

(b) How does your answer to (a) change if S is a left Riemann sum? a middle Riemann sum?

(c) Suppose that S really is a right Riemann sum. What is geometric quantity does S approximate?

(d) Use sigma notation to write a new sum R that is the right Riemann sum for the same function, but that uses twice as many subintervals as S.

40) A car traveling along a straight road is braking and its velocity, in ft/sec, is measured at several different points in time, as given in the following table.

seconds, t	0	0.3	0.6	0.9	1.2	1.5	1.8
$v(t)$	100	88	74	59	40	19	0

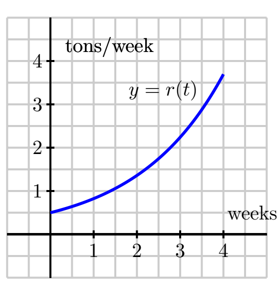
(a) Plot the given data on a set of axes with time on the horizontal axis and the velocity on the vertical axis.

(b) Estimate the total distance traveled during the car the time brakes using a middle Riemann sum with 3 subintervals.

(c) Estimate the total distance traveled on $[0, 1.8]$ by computing L_6, R_6, and $\frac{1}{2}(L_6 + R_6)$.

(d) Assuming that $v(t)$ is always decreasing on $[0, 1.8]$, what is the maximum possible distance the car traveled before it stopped? Why?

41) The rate at which pollution escapes a scrubbing process at a manufacturing plant increases over time as filters and other technologies become less effective. For this particular example, assume that the rate of pollution (in tons per week) is given by the function r that is pictured below.



(a) Use the graph to estimate the value of M_4 on the interval $[0, 4]$.

(b) What is the meaning of M_4 in terms of the pollution discharged by the plant?

(c) Suppose that $r(t) = 0.5e^{0.5t}$. Use this formula for r to compute L_5 on $[0, 4]$.

(d) Determine an upper bound on the total amount of pollution that can escape the plant during the pictured four week time period that is accurate within an error of at most one ton of pollution.

4.3 *The Definite Integral*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- How does increasing the number of subintervals affect the accuracy of the approximation generated by a Riemann sum?
- What is the definition of the definite integral of a function f over the interval $[a,b]$?
- What does the definite integral measure exactly, and what are some of the key properties of the definite integral?

Introduction

In this chapter, we have computed or approximated the net signed area that is bounded by a function on a given interval. We now give mathematical notation to signify the net signed area.

Net Signed Area as The Definite Integral

Let $y = f(x)$ be defined on a closed interval $[a,b]$. The net signed area from $x = a$ to $x = b$ under f is signified by the *definite integral* of f on $[a,b]$, which is denoted

$$\int_a^b f(x)\,dx.$$

The area definition of the definite integral allows us to compute the definite integral of some simple functions using geometry as we did in section 4.1.

For instance, if we wish to evaluate the definite integral $\int_1^4 (2x+1)\,dx$, we can observe that the region bounded by this function and the x-axis is the trapezoid shown in Figure 4.27, and by the known formula for the area of a trapezoid, its area is $A = \frac{1}{2}(3+9)\cdot 3 = 18$, so

$$\int_1^4 (2x+1)\,dx = 18.$$

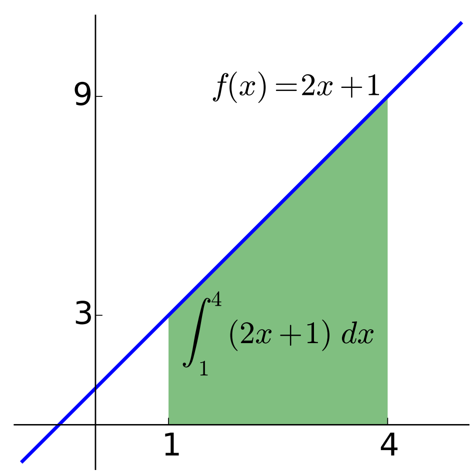


Figure 4.27: The area bounded by $f(x) = 2x+1$ and the x-axis on the interval $[1,4]$.

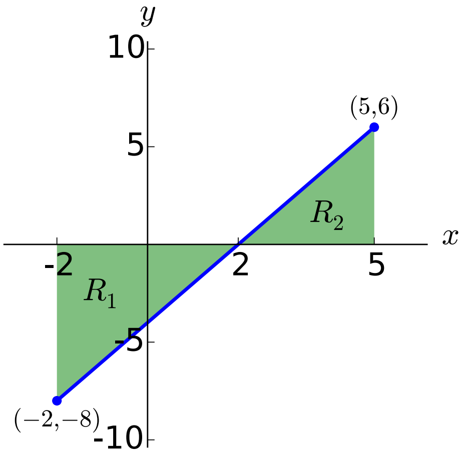


Figure 4.28: A graph of $f(x) = 2x - 4$.

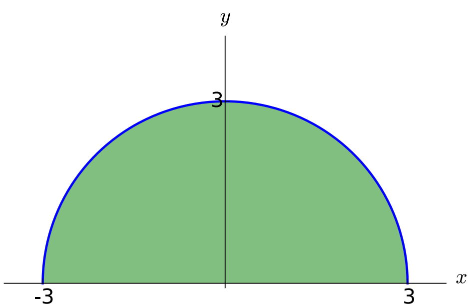


Figure 4.29: A graph of $f(x) = \sqrt{9 - x^2}$.

Example 1

Evaluate the following definite integrals using geometry.

1) $\int_{-2}^{5} (2x - 4)\, dx$ 2) $\int_{-3}^{3} \sqrt{9 - x^2}\, dx$

Solution.

1) It is useful to sketch the function in the integrand, as shown in Figure 4.28. We see we need to compute the areas of two regions, which we have labeled R_1 and R_2. Both are triangles, so the area computation is straightforward:

$$R_1 = \frac{1}{2}(4)(8) = 16 \qquad R_2 = \frac{1}{2}(3)6 = 9.$$

Region R_1 lies under the x-axis, hence it is counted as negative area (we can think of the height as being "-8"), so

$$\int_{-2}^{5} (2x - 4)\, dx = 9 - 16 = -7.$$

2) Recognize that the integrand of this definite integral is a half circle, as sketched in Figure 4.29, with radius 3. Thus the area is:

$$\int_{-3}^{3} \sqrt{9 - x^2}\, dx = \frac{1}{2}\pi r^2 = \frac{9}{2}\pi.$$

In section 4.2, we found evidence that if we increase the number of rectangles in a Riemann sum, our accuracy of the approximation of the net signed area. We thus explore the natural idea of allowing the number of rectangles to increase without bound in an effort to compute the exact net signed area bounded by a function on an interval. In addition, it is important to think about the differences among left, right, and middle Riemann sums and the different results they generate as the value of n increases. As we have done throughout our investigations with area, we begin with functions that are exclusively positive on the interval under consideration.

Preview Activity 4.3

Consider the applet found at `http://gvsu.edu/s/aw`. There, you will initially see the situation shown in Figure 4.30.

Observe that we can change the window in which the function is viewed, as well as the function itself. Set the minimum and maximum values of x and y so that we view the function on the window where $1 \leq x \leq 4$ and $-1 \leq y \leq 12$, where the function is $f(x) = 2x + 1$ (note that you need to enter "`2*x+1`" as the function's formula). You should see the updated figure shown in Figure 4.31.

Note that the value of the Riemann sum of our choice is displayed in the upper left corner of the window. Further, by updating the value in the "Intervals" window and/or the "Method", we can see the different value of the Riemann sum that arises by clicking the "Compute!" button.

(a) Update the applet so that the function being considered is $f(x) = 2x + 1$ on $[1, 4]$, as directed above. For this function on this interval, compute L_n, M_n, R_n for $n = 10$, $n = 100$, and $n = 1000$. What do you conjecture is the exact area bounded by $f(x) = 2x + 1$ and the x-axis on $[1, 4]$?

(b) Use basic geometry to determine the exact area bounded by $f(x) = 2x + 1$ and the x-axis on $[1, 4]$.

(c) Based on your work in (a) and (b), what do you observe occurs when we increase the number of subintervals used in the Riemann sum?

(d) Update the applet to consider the function $f(x) = x^2 + 1$ on the interval $[1, 4]$ (note that you will want to increase the maximum value of y to at least 17, and you need to enter "`x^2 + 1`" for the function formula). Use the applet to compute L_n, M_n, R_n for $n = 10$, $n = 100$, and $n = 1000$. What do you conjecture is the exact area bounded by $f(x) = x^2 + 1$ and the x-axis on $[1, 4]$?

(e) Why can we not compute the exact value of the area bounded by $f(x) = x^2 + 1$ and the x-axis on $[1, 4]$ using a formula like we did in (b)?

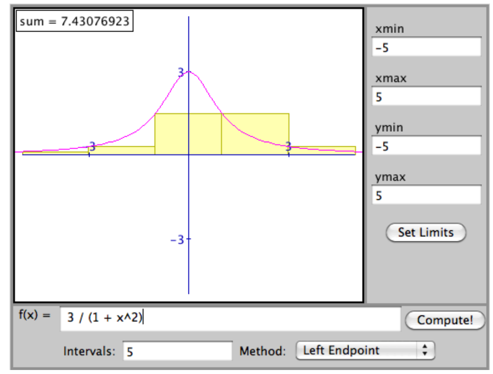


Figure 4.30: A left Riemann sum with 5 subintervals for the function $f(x) = \frac{3}{1+x^2}$ on the interval $[-5, 5]$. The value of the sum is $L_5 = 7.43076923$.

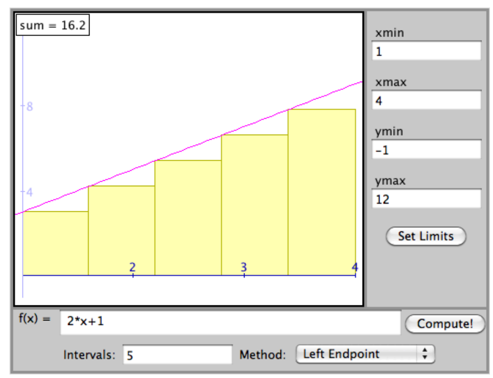


Figure 4.31: A left Riemann sum with 5 subintervals for the function $f(x) = 2x + 1$ on the interval $[1, 4]$. The value of the sum is $L_5 = 16.2$.

The formal definition of the definite integral

In both examples in Preview Activity 4.5, we saw that as the number of rectangles got larger and larger, the values of L_n, M_n, and R_n all grew closer and closer to the same value. It turns out that this occurs for any continuous function on an interval $[a, b]$, and even more generally for a Riemann sum using any point x_{i+1}^* in the interval $[x_i, x_{i+1}]$. Said differently, as we let $n \to \infty$, it doesn't really matter where we choose to evaluate the function within a given subinterval, because

$$\lim_{n\to\infty} L_n = \lim_{n\to\infty} R_n = \lim_{n\to\infty} M_n = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*) \triangle x.$$

That these limits always exist (and share the same value) for a continuous[2] function f allows us to make the following definition.

[2] It turns out that a function need not be continuous in order to have a definite integral. For our purposes, we assume that the functions we consider are continuous on the interval(s) of interest. It is straightforward to see that any function that is piecewise continuous on an interval of interest will also have a well-defined definite integral.

The Definite Integral

The *definite integral* of a continuous function f on the interval $[a, b]$, denoted $\displaystyle\int_a^b f(x)\,dx$, is the real number given by

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\triangle x,$$

where $\Delta x = \frac{b-a}{n}$, $x_i = a + i\triangle x$ (for $i = 0, \ldots, n$), and x_i^* satisfies $x_{i-1} \leq x_i^* \leq x_i$ (for $i = 1, \ldots, n$).

We call the symbol $\int$ the *integral sign*, the values a and b the *limits or bounds of integration*, and the function f the *integrand*. The process of determining the real number $\int_a^b f(x)\,dx$ is called *evaluating the definite integral*. While we will come to understand that there are several different interpretations of the value of the definite integral, for now the most important is that $\int_a^b f(x)\,dx$ measures the net signed area bounded by $y = f(x)$ and the x-axis on the interval $[a, b]$.

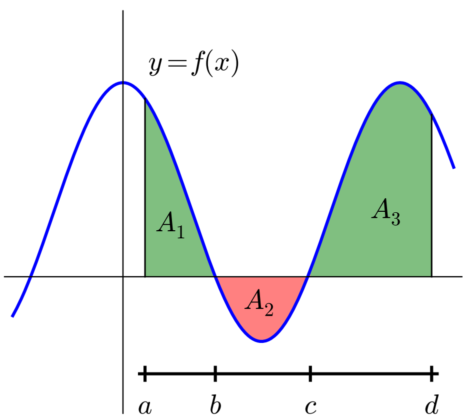


Figure 4.32: A continuous function f on the interval $[a, d]$.

For example, in the notation of the definite integral, if f is the function pictured in Figure 4.32 and A_1, A_2, and A_3 are the exact areas bounded by f and the x-axis on the respective intervals $[a, b]$, $[b, c]$, and $[c, d]$, then

$$\int_a^b f(x)\,dx = A_1,\ \int_b^c f(x)\,dx = -A_2,\ \int_c^d f(x)\,dx = A_3,$$

and

$$\int_a^d f(x)\,dx = A_1 - A_2 + A_3.$$

We can also use definite integrals to express the change in position and distance traveled by a moving object. In the setting of a velocity function v on an interval $[a, b]$, it follows from our work above and in preceding sections that the change in position, $s(b) - s(a)$, is given by

$$s(b) - s(a) = \int_a^b v(t)\,dt.$$

If the velocity function is nonnegative on $[a, b]$, then $\int_a^b v(t)\,dt$ tells us the distance the object traveled. When velocity is sometimes negative on $[a, b]$, the areas bounded by the function on intervals where v does not change sign can be found using integrals, and the sum of these values will tell us the distance the object traveled.

If we wish to compute the value of a definite integral using the definition, we have to take the limit of a sum. While this is possible to do in select circumstances, it is also tedious and time-consuming; therefore, we typically take the limit of only relatively simple sums.

Example 2

Use limits to compute the exact value of

$$A = \int_0^4 (4x - x^2)\, dx,$$

recalling that

$$A \approx \frac{32}{3}\left(1 - \frac{1}{n^2}\right).$$

Solution. Taking the limit as $n \to \infty$,

$$A = \lim_{n\to\infty} \frac{32}{3}\left(1 - \frac{1}{n^2}\right) = \frac{32}{3}.$$

Example 3

Find a formula that approximates $\int_{-2}^{3} (5x+2)\, dx$ using R_n; then take the limit as $n \to \infty$ to find the exact area.

Solution. We have $\Delta x = \dfrac{3-(-2)}{n} = \dfrac{5}{n}$ and $x_i = -2 + \dfrac{5}{n}i$, and we construct the Riemann sum and compute its value using summation formulas.

$$\begin{aligned}
\int_{-2}^{3} (5x+2)\, dx &\approx \sum_{i=1}^{n} f(x_i)\Delta x \\
&= \sum_{i=1}^{n} f(-2 + i\Delta x)\Delta x \\
&= \sum_{i=1}^{n} \left(5(-2 + i\Delta x) + 2\right)\Delta x \\
&= \Delta x \sum_{i=1}^{n} [(5\Delta x)\, i - 8] \\
&= \Delta x \left[5\Delta x \sum_{i=1}^{n} (i) - \sum_{i=1}^{n} 8\right] \\
&= \frac{5}{n}\left[\frac{25}{n} \cdot \frac{n(n+1)}{2} - 8 \cdot n\right] \\
&= \frac{125n(n+1)}{2n^2} - 40
\end{aligned}$$

Taking the limit we get

$$\int_{-2}^{3}(5x+2)\,dx = \lim_{n\to\infty}\frac{125n(n+1)}{2n^2} - 40 = \frac{45}{2}$$

Activity 4.3–1

Use the properties of summations and limits to find the exact net area of the region underneath the function x^3 on the interval $[-1,5]$.

(a) Sketch a graph of $f(x) = x^3$.

(b) Find formulas for Δx and x_i in terms of n.

(c) Generate an approximation for $\int_{-1}^{5} x^3\,dx$ in terms of n.

(d) Using the approximation you found in (c), take the limit as $n \to \infty$ to find the exact net area.

In Section 4.5, we will learn the Second Part of the Fundamental Theorem of Calculus, a theorem that provides an analytical process for evaluating a large class of definite integrals. This will enable us to determine the exact net signed area bounded by a continuous function and the x-axis in many circumstances, including examples such as $\displaystyle\int_1^4 (x^2+1)\,dx$, which we approximated by Riemann sums in Preview Activity 4.5.

For now, our goal is to understand the meaning and properties of the definite integral, rather than how to actually compute its value using ideas in calculus. Thus, we temporarily rely on the net signed area interpretation of the definite integral and observe that if a given curve produces regions whose areas we can compute exactly through known area formulas, we can thus compute the exact value of the integral.

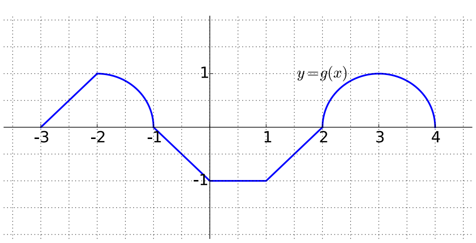


Figure 4.33: A function g that is piecewise defined; each piece of the function is part of a circle or part of a line.

Activity 4.3–2

Use known geometric formulas and the net signed area interpretation of the definite integral to evaluate each of the definite integrals below.

(a) $\displaystyle\int_0^1 3x\,dx$

(b) $\displaystyle\int_{-1}^4 (2-2x)\,dx$

(c) $\displaystyle\int_{-1}^1 \sqrt{1-x^2}\,dx$

(d) $\displaystyle\int_{-3}^4 g(x)\,dx$, where g is the function pictured in Figure 4.33. Assume that each portion of g is either part of a line or part of a circle.

Some properties of the definite integral

With the perspective that the definite integral of a function f over an interval $[a, b]$ measures the net signed area bounded by f and the x-axis over the interval, we naturally arrive at several different standard properties of the definite integral. In addition, it is helpful to remember that the definite integral is defined in terms of Riemann sums that fundamentally consist of the areas of rectangles.

Properties of the Definite Integral

Let f and g be defined on a closed interval I that contains the values a, b and c, and let k be a constant. The following hold:

1) $\displaystyle\int_a^a f(x)\,dx = 0$

2) $\displaystyle\int_a^b f(x)\,dx + \int_b^c f(x)\,dx = \int_a^c f(x)\,dx$

3) $\displaystyle\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$

4) $\displaystyle\int_a^b k\cdot f(x)\,dx = k\cdot\int_a^b f(x)\,dx$

5) $\displaystyle\int_a^b \big(f(x)\pm g(x)\big)\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$

We give a brief justification the properties here.

1) If we consider the definite integral $\int_a^a f(x)\,dx$ for any real number a, it is evident that no area is being bounded because the interval begins and ends with the same point. Hence, this definite integral is 0.

2) This states that total area is the sum of the areas of subregions. In Figure 4.34, we see that

$$\int_a^b f(x)\,dx = A_1,\ \int_b^c f(x)\,dx = A_2, \text{ and } \int_a^c f(x)\,dx = A_1 + A_2,$$

It is important to note that this still holds true even if $a < b < c$ is not true.

3) This result makes sense because if we integrate from a to b, then in the defining Riemann sum $\triangle x = \frac{b-a}{n}$, while if we integrate from b to a, $\triangle x = \frac{a-b}{n} = -\frac{b-a}{n}$, and this is the only change in the sum used to define the integral.

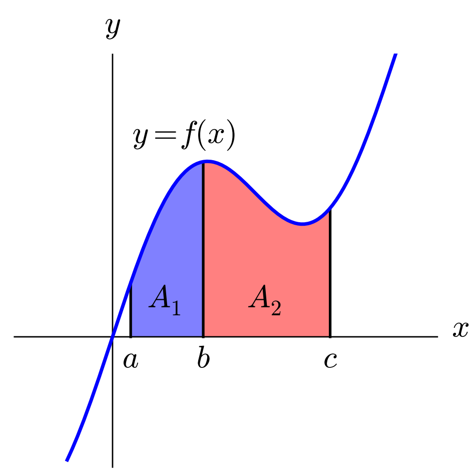


Figure 4.34: A continuous function f on the interval $[a, d]$.

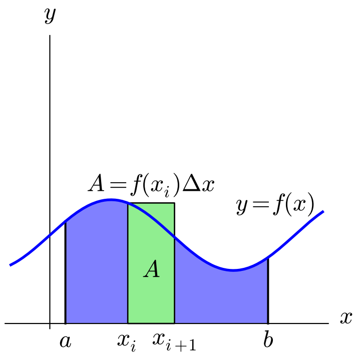


Figure 4.35: The area bounded by $y = f(x)$ on $[a, b]$.

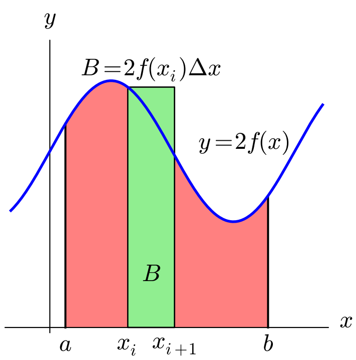


Figure 4.36: The area bounded by $y = 2f(x)$ on $[a, b]$.

4) First, let's consider the situation pictured in Figure 4.36, where we examine the effect of multiplying a function by a factor of 2 on the area it bounds with the x-axis. Because multiplying the function by 2 doubles its height at every x-value, we see that if we consider a typical rectangle from a Riemann sum, the difference in area comes from the changed height of the rectangle: $f(x_i)$ for the original function, versus $2f(x_i)$ in the doubled function, in the case of left sum. Hence, we see that for the pictured rectangles with areas A and B, it follows $B = 2A$. As this will happen in every such rectangle, regardless of the value of n and the type of sum we use, we see that in the limit, the area of the red region bounded by $y = 2f(x)$ will be twice that of the area of the blue region bounded by $y = f(x)$. As there is nothing special about the value 2 compared to an arbitrary constant k, it turns out that the general principle holds.

5) Finally, we see a similar situation geometrically with the sum of two functions f and g. In particular, as shown in Figure 4.37, if we take the sum of two functions f and g, at every point in the interval, the height of the function $f + g$ is given by $(f + g)(x_i) = f(x_i) + g(x_i)$, which is the sum of the individual function values of f and g (taken at left endpoints). Hence, for the pictured rectangles with areas A, B, and C, it follows that $C = A + B$, and because this will occur for every such rectangle, in the limit the area of the gray region will be the sum of the areas of the blue and red regions.

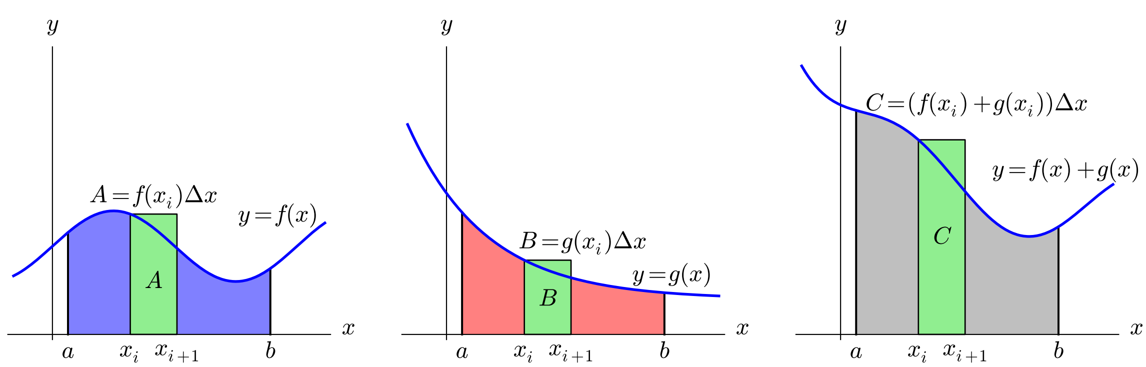


Figure 4.37: The areas bounded by $y = f(x)$ and $y = g(x)$ on $[a, b]$, as well as the area bounded by $y = f(x) + g(x)$.

Activity 4.3–3

Suppose that the following information is known about the functions f, g, x^2, and x^3:

- $\int_0^2 f(x)\,dx = -3$; $\int_2^5 f(x)\,dx = 2$
- $\int_0^2 g(x)\,dx = 4$; $\int_2^5 g(x)\,dx = -1$
- $\int_0^2 x^2\,dx = \frac{8}{3}$; $\int_2^5 x^2\,dx = \frac{117}{3}$
- $\int_0^2 x^3\,dx = 4$; $\int_2^5 x^3\,dx = \frac{609}{4}$

Use the provided information and the rules discussed in the preceding section to evaluate each of the following definite integrals.

(a) $\int_5^2 f(x)\,dx$

(b) $\int_0^5 g(x)\,dx$

(c) $\int_0^5 (f(x)+g(x))\,dx$

(d) $\int_2^5 (3x^2 - 4x^3)\,dx$

(e) $\int_5^0 (2x^3 - 7g(x))\,dx$

Summary

In this section, we encountered the following important ideas:

- Any Riemann sum of a continuous function f on an interval $[a,b]$ provides an estimate of the net signed area bounded by the function and the horizontal axis on the interval. Increasing the number of subintervals in the Riemann sum improves the accuracy of this estimate, and letting the number of subintervals increase without bound results in the values of the corresponding Riemann sums approaching the exact value of the enclosed net signed area.
- When we take the just described limit of Riemann sums, we arrive at what we call the definite integral of f over the interval $[a,b]$. In particular, the symbol $\int_a^b f(x)\,dx$ denotes the definite integral of f over $[a,b]$, and this quantity is defined by the equation

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\triangle x,$$

  where $\triangle x = \frac{b-a}{n}$, $x_i = a + i\triangle x$ (for $i = 0, \ldots, n$), and x_i^* satisfies $x_{i-1} \le x_i^* \le x_i$ (for $i = 1, \ldots, n$).
- The definite integral $\int_a^b f(x)\,dx$ measures the exact net signed area bounded by f and the horizontal axis on $[a,b]$. In the setting where we consider the integral of a velocity function v, $\int_a^b v(t)\,dt$ measures the exact change in position of the moving object on $[a,b]$; when v is nonnegative, $\int_a^b v(t)\,dt$ is the object's distance traveled on $[a,b]$.
- The definite integral is a sophisticated sum, and thus has some of the same natural properties that finite sums have. Perhaps most important of these is how the definite integral respects sums and constant multiples of functions, which can be summarized by the rule

$$\int_a^b [cf(x) \pm kg(x)]\,dx = c\int_a^b f(x)\,dx \pm k\int_a^b g(x)\,dx$$

  where f and g are continuous functions on $[a,b]$ and c and k are arbitrary constants.

Exercises

Terms and Concepts

1) What is "total signed area"?

2) What is "displacement"?

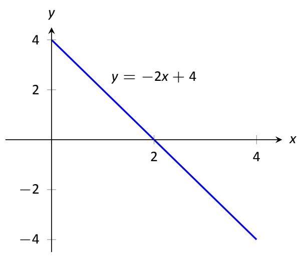
3) What is $\int_3^3 \sin x\, dx$?

4) Give a single definite integral that has the same value as
$$\int_0^1 (2x+3)\, dx + \int_1^2 (2x+3)\, dx.$$

Problems

In exercises 5–9, a graph of a function $f(x)$ is given. Using the geometry of the graph, evaluate the definite integral.

5)



(a) $\int_0^1 (-2x+4)\, dx$

(b) $\int_0^2 (-2x+4)\, dx$

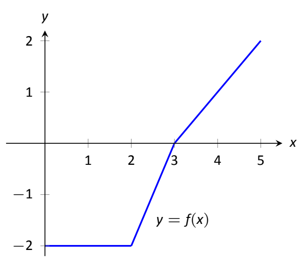
(c) $\int_0^3 (-2x+4)\, dx$

(d) $\int_1^3 (-2x+4)\, dx$

(e) $\int_2^4 (-2x+4)\, dx$

(f) $\int_0^1 (-6x+12)\, dx$

6)



(a) $\int_0^2 f(x)\, dx$

(b) $\int_0^3 f(x)\, dx$

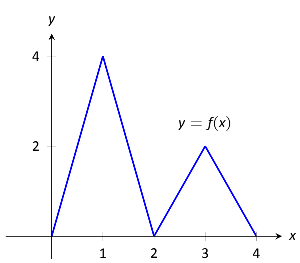
(c) $\int_0^5 f(x)\, dx$

(d) $\int_2^5 f(x)\, dx$

(e) $\int_5^3 f(x)\, dx$

(f) $\int_0^3 -2f(x)\, dx$

7)



(a) $\int_0^2 f(x)\, dx$

(b) $\int_2^4 f(x)\, dx$

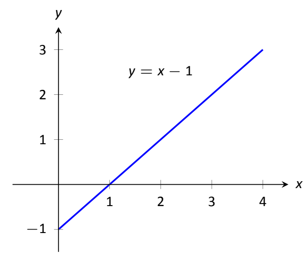
(c) $\int_2^4 2f(x)\, dx$

(d) $\int_0^1 4x\, dx$

(e) $\int_2^3 (2x-4)\, dx$

(f) $\int_2^3 (4x-8)\, dx$

8)



(a) $\int_0^1 (x-1)\, dx$

(b) $\int_0^2 (x-1)\, dx$

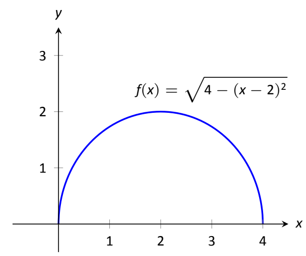
(c) $\int_0^3 (x-1)\, dx$

(d) $\int_2^3 (x-1)\, dx$

(e) $\int_1^4 (x-1)\, dx$

(f) $\int_1^4 \big((x-1)+1\big)\, dx$

9)



(a) $\int_0^2 f(x)\, dx$

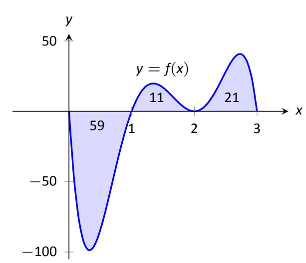
(b) $\int_2^4 f(x)\, dx$

(c) $\int_0^4 f(x)\, dx$

(d) $\int_0^4 5f(x)\, dx$

In exercises 10–13, a graph of a function $f(x)$ is given; the numbers inside the shaded regions give the area of that region. Evaluate the definite integrals using this area information.

10)

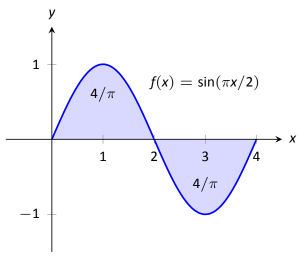


(a) $\int_0^1 f(x)\, dx$

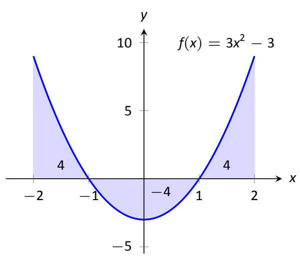
(b) $\int_0^2 f(x)\, dx$

(c) $\int_0^3 f(x)\, dx$

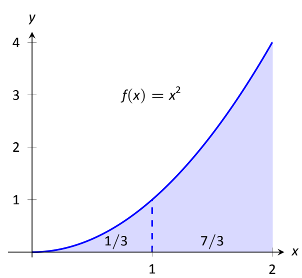
(d) $\int_1^2 -3f(x)\, dx$

11) 

(a) $\int_0^2 f(x)\,dx$
(b) $\int_2^4 f(x)\,dx$
(c) $\int_0^4 f(x)\,dx$
(d) $\int_0^1 f(x)\,dx$

12) 

(a) $\int_{-2}^{-1} f(x)\,dx$
(b) $\int_1^2 f(x)\,dx$
(c) $\int_{-1}^1 f(x)\,dx$
(d) $\int_0^1 f(x)\,dx$

13) 

(a) $\int_0^2 5x^2\,dx$
(b) $\int_0^2 (x^2+3)\,dx$
(c) $\int_1^3 (x-1)^2\,dx$
(d) $\int_2^4 \left((x-2)^2+5\right)\,dx$

In exercises 14–17, let

- $\int_0^2 f(x)\,dx = 5$,
- $\int_0^2 g(x)\,dx = -3$, **and**
- $\int_0^3 f(x)\,dx = 7$,
- $\int_2^3 g(x)\,dx = 5$.

Use these values to evaluate the given definite integrals.

14) $\int_0^2 \left(f(x)+g(x)\right)\,dx$

15) $\int_0^3 \left(f(x)-g(x)\right)\,dx$

16) $\int_2^3 \left(3f(x)+2g(x)\right)\,dx$

17) Find values for a and b such that
$$\int_0^3 \left(af(x)+bg(x)\right)\,dx = 0$$

In exercises 18–21, let

- $\int_0^3 s(t)\,dt = 10$,
- $\int_3^5 r(t)\,dx = -1$, **and**
- $\int_3^5 s(t)\,dx = 8$,
- $\int_0^5 r(t)\,dx = 11$.

Use these values to evaluate the given definite integrals.

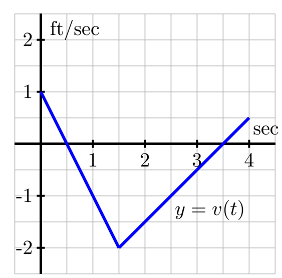
18) $\int_0^3 \left(s(t)+r(t)\right)\,dt$

19) $\int_5^0 \left(s(t)-r(t)\right)\,dt$

20) $\int_3^3 \left(\pi s(t)-7r(t)\right)\,dt$

21) Find values for a and b such that
$$\int_0^5 \left(ar(t)+bs(t)\right)\,dt = 0$$

22) The velocity of an object moving along an axis is given by the piecewise linear function v that is pictured below. Assume that the object is moving to the right when its velocity is positive, and moving to the left when its velocity is negative. Assume that the given velocity function is valid for $t = 0$ to $t = 4$.



(a) Write an expression involving definite integrals whose value is the total change in position of the object on the interval $[0,4]$.
(b) Use the provided graph of v to determine the value of the total change in position on $[0,4]$.
(c) Write an expression involving definite integrals whose value is the total distance traveled by the object on $[0,4]$. What is the exact value of the total distance traveled on $[0,4]$?
(d) What is the object's exact average velocity on $[0,4]$?
(e) Find an algebraic formula for the object's position function on $[0,1.5]$ that satisfies $s(0) = 0$.

4.4 *The Fundamental Theorem of Calculus, Part I*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- Given a function f, how does the rule $A(x) = \int_a^x f(t)\,dt$ define a new function A?
- What is the statement of the Fundamental Theorem of Calculus, and how do we apply the theorem?

Introduction

Given a continuous function f defined on $[a, b]$, we define the corresponding integral, or area, function A according to the rule

$$A(x) = \int_a^x f(t)\,dt. \tag{4.2}$$

Note particularly that because we are using the variable x as the independent variable in the function A, and x determines the other endpoint of the interval over which we integrate (starting from a), we need to use a variable other than x as the variable of integration. A standard choice is t, but any variable other than x is acceptable.

One way to think of the function A is as the "net-signed area from a up to x" function, where we consider the region bounded by $y = f(t)$ on the relevant interval. For example, in Figure 4.38, we see a given function f and its corresponding area function (choosing $a = 0$), $A(x) = \int_0^x f(t)\,dt$.

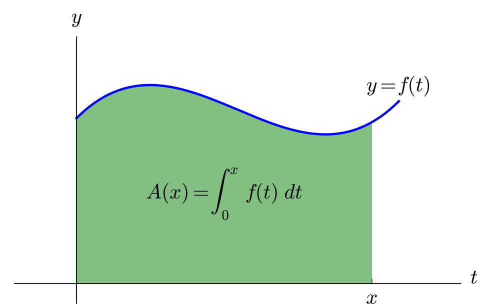


Figure 4.38: A function f and its corresponding area function $A(x) = \int_0^x f(t)\,dt$.

Note particularly that the function A measures the net-signed area from $t = 0$ to $t = x$ bounded by the curve $y = f(t)$; this value is then reported as the corresponding height on the graph of $y = A(x)$—see Figure 4.39. It is even more natural to think of this relationship between f and A dynamically. At `http://gvsu.edu/s/cz`, we find a java applet that brings this relationship to life. There, the user can move the red point on the function f and see how the corresponding height changes at the light blue point on the graph of A.

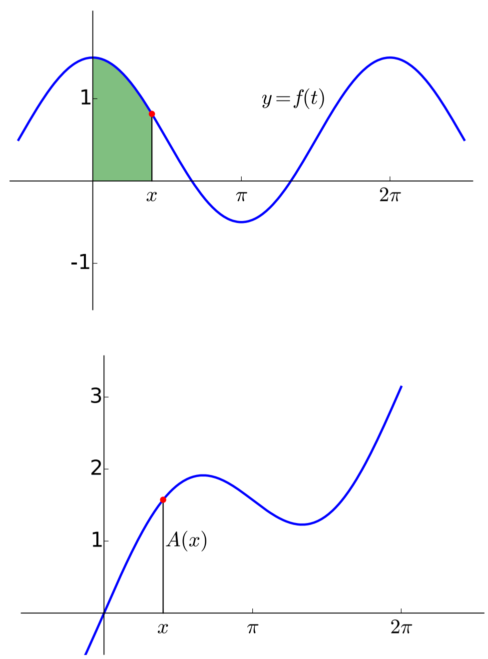


Figure 4.39: The area under f and the height of $A(x)$.

Preview Activity 4.4

Consider the function A defined by the rule

$$A(x) = \int_1^x f(t)\,dt,$$

where $f(t) = 4 - 2t$.

(a) Compute $A(1)$ and $A(2)$ exactly using geometry.

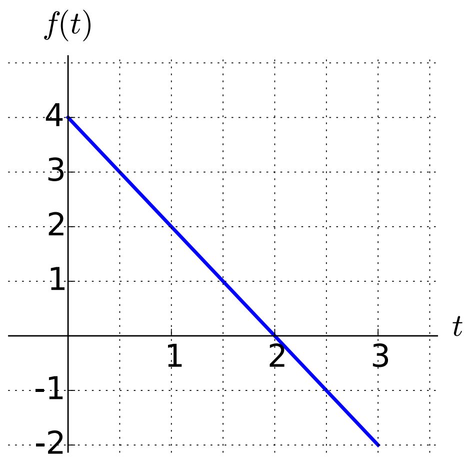


Figure 4.40: $f(t) = 4 - 2t$

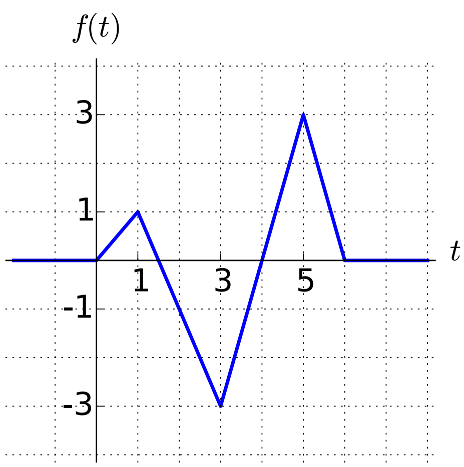


Figure 4.41: $y = g(t)$

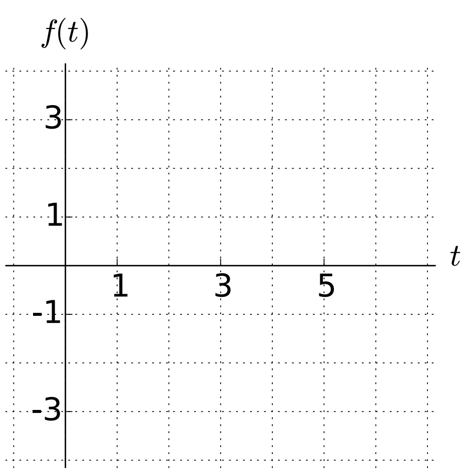


Figure 4.42: Sketch $A(x) = \int_1^x g(t)\,dt$

(b) Suppose that we can produce a formula for $A(x)$ that does not use integrals, and suppose that formula is $A(x) = -x^2 + 4x - 3$. Observe that f is a linear function; what kind of function is this additional formula for A?

(c) Using the additional formula for $A(x)$ found in (b) that does not involve integrals, compute $A(1)$, $A(2)$, and $A'(x)$.

(d) While we have defined f by the rule $f(t) = 4 - 2t$, it is equivalent to say that f is given by the rule $f(x) = 4 - 2x$. What do you observe about the relationship between A and f?

The choice of the lower bournd of integration a is somewhat arbitrary. In the activity that follows, we explore how the value of a affects the graph of the integral function, as well as some additional related issues.

Activity 4.4–1

Suppose that g is given by the graph at left in Figure 4.41 and that A is the corresponding integral function defined by $A(x) = \int_1^x g(t)\,dt$.

(a) On what interval(s) is A an increasing function? On what intervals is A decreasing? Why?

(b) On what interval(s) do you think A is concave up? concave down? Why?

(c) At what point(s) does A have a relative minimum? a relative maximum?

(d) Use the given information to determine the exact values of $A(0)$, $A(1)$, $A(2)$, $A(3)$, $A(4)$, $A(5)$, and $A(6)$.

(e) Based on your responses to all of the preceding questions, sketch a complete and accurate graph of $y = A(x)$ on the axes provided, being sure to indicate the behavior of A for $x < 0$ and $x > 6$.

(f) How does the graph of B compare to A if B is instead defined by $B(x) = \int_0^x g(t)\,dt$?

The Fundamental Theorem of Calculus

We can also apply calculus ideas to $A(x)$; in particular, we can compute its derivative. While this may seem like an innocuous thing to do, it has far–reaching implications, as demonstrated by the fact that the result is given as an important theorem.

Fundamental Theorem of Calculus, Part 1

Let f be continuous on $[a, b]$ and let $A(x) = \int_a^x f(t)\,dt$.

Then A is a differentiable function on (a, b), and

$$A'(x) = f(x).$$

Proof: In general, if f is any continuous function, and we define the function A by the rule

$$A(x) = \int_a^x f(t)\, dt,$$

where a is an arbitrary constant, then we can show that A is an antiderivative of f. To see why, let's demonstrate that $A'(x) = f(x)$ by using the limit definition of the derivative. Doing so, we observe that

$$\begin{aligned} A'(x) &= \lim_{h\to 0} \frac{A(x+h) - A(x)}{h} \\ &= \lim_{h\to 0} \frac{\int_a^{x+h} f(t)\, dt - \int_a^x f(t)\, dt}{h} \\ &= \lim_{h\to 0} \frac{\int_x^{x+h} f(t)\, dt}{h}, \qquad (4.3) \end{aligned}$$

where Equation (4.3) in the preceding chain follows from the fact that $\int_a^x f(t)\, dt + \int_x^{x+h} f(t)\, dt = \int_a^{x+h} f(t)\, dt$. Now, observe that for small values of h,

$$\int_x^{x+h} f(t)\, dt \approx f(x) \cdot h,$$

by a simple left-hand approximation of the integral. Thus, as we take the limit in Equation (4.3) as h approaches 0, it follows that

$$A'(x) = \lim_{h\to 0} \frac{\int_x^{x+h} f(t)\, dt}{h} = \lim_{h\to 0} \frac{f(x)\cdot h}{h} = f(x).$$

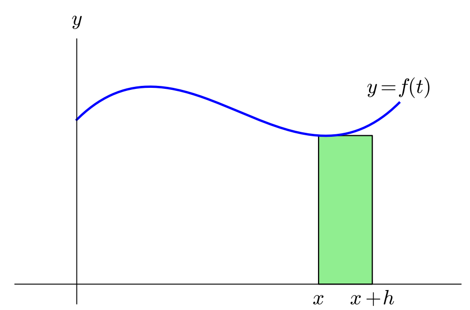


Figure 4.43: Approximating the area under f from x to $x+h$.

Example 1

Let $A(x) = \int_{-5}^{x} \left(t^2 + \sin(t)\right)\, dt$. What is $A'(x)$?

Solution. Using the Fundamental Theorem of Calculus, we have

$$A'(x) = x^2 + \sin(x).$$

Example 2

Let $A(x) = \int_{\pi}^{x} \sin(t^2)\, dt$. What is $A'(x)$?

Solution. Using the Fundamental Theorem of Calculus, we have

$$A'(x) = \sin(x^2).$$

Activity 4.4–2

Compute each of the following derivatives using the Fundamental Theorem of Calculus.

(a) $\dfrac{d}{dx}\left[\displaystyle\int_4^x e^{t^2}\, dt\right]$

(b) $\dfrac{d}{dx}\left[\displaystyle\int_{-2}^x \frac{t^4}{1+t^4}\, dt\right]$

(c) $\dfrac{d}{dx}\left[\displaystyle\int_x^1 \cos(t^3)\, dt\right]$

(d) $\dfrac{d}{dx}\left[\displaystyle\int_3^x \ln(1+t^2)\, dt\right]$

The Fundamental Theorem and the Chain Rule

The Fundamental Theorem of Calculus states that given $A(x) = \int_a^x f(t)\, dt$, $A'(x) = f(x)$, or we can use other notation to write $\frac{d}{dx}[A(x)] = f(x)$. It may be of further use to compose such a function with another. As an example, we may compose $A(x)$ with $g(x)$ to get

$$A \circ g(x) = \int_a^{g(x)} f(t)\, dt.$$

What is the derivative of such a function? Since it is a composition, the Chain Rule can be employed. We decompose the composition as

$$y = \int_a^u f(t)\, dt \quad \text{where} \quad u = g(x).$$

Using the Leibniz version of the Chain Rule, we have

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx},$$

where

$$\frac{dy}{du} = f(u) \quad \text{and} \quad \frac{du}{dx} = g'(x).$$

Therefore,

$$\frac{dy}{dx} = f(u) \cdot g'(x) = f\left(g(x)\right) \cdot g'(x).$$

Example 3

Using the FTC and the Chain Rule, compute the derivative of $A(x) = \int_2^{x^2} \ln(t)\, dt$.

Solution. Noticing that this function is a composition, we can decompose it as

$$y = \int_2^u \ln(t)\, dt \quad \text{where} \quad u = x^2.$$

Using the Leibniz version of the Chain Rule, we have

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx},$$

where

$$\frac{dy}{du} = \ln(u) \quad \text{and} \quad \frac{du}{dx} = 2x.$$

Therefore,

$$\frac{dy}{dx} = \ln(u) \cdot 2x = \ln\left(x^2\right) \cdot 2x.$$

Example 4

Compute the derivative of $A(x) = \int_{\cos(x)}^5 t^3\, dt$.

Solution. Note that $A(x) = -\int_5^{\cos(x)} t^3\, dt$, which is again a composition. We can decompose it as

$$y = -\int_5^u t^3\, dt \quad \text{where} \quad u = \cos(x).$$

Using the Leibniz version of the Chain Rule, we have

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx},$$

where

$$\frac{dy}{du} = -u^3 \quad \text{and} \quad \frac{du}{dx} = -sin(x).$$

Therefore,

$$\frac{dy}{dx} = -u^3 \cdot -sin(x) = \cos^3(x) \cdot \sin(x).$$

Activity 4.4–3

Compute each of the following derivatives using the Fundamental Theorem of Calculus and the Chain Rule.

(a) $\frac{d}{dx}\left[\int_4^{x^3} \sin(t^2)\, dt\right]$

(b) $\frac{d}{dx}\left[\int_{-3}^{\sqrt{x}} \frac{t}{t+1}\, dt\right]$

(c) $\dfrac{d}{dx}\left[\int_{10}^{\tan(x)} 2te^t\ dt\right]$

Understanding Integral Functions

The FTC provides us with a means to construct an antiderivative of any continuous function. In particular, if we are given a continuous function g and wish to find an antiderivative of G, we can now say that

$$G(x) = \int_c^x g(t)\ dt$$

provides the rule for such an antiderivative, and moreover that $G(c) = 0$. Note especially that we know that $G'(x) = g(x)$. We sometimes want to write this relationship between G and g from a different notational perspective. In particular, observe that

$$\frac{d}{dx}\left[\int_c^x g(t)\, dt\right] = g(x). \tag{4.4}$$

This result can be particularly useful when we're given an integral function such as G and wish to understand properties of its graph by recognizing that $G'(x) = g(x)$, while not necessarily being able to exactly evaluate the definite integral $\int_c^x g(t)\ dt$.

To see how this is the case, let's investigate the behavior of the integral function

$$E(x) = \int_0^x e^{-t^2}\ dt.$$

E is closely related to the well known *error function*[3], a function that is particularly important in probability and statistics. It turns out that the function e^{-t^2} does not have an elementary antiderivative that we can express without integrals. That is, whereas a function such as $f(t) = 4 - 2t$ has elementary antiderivative $F(t) = 4t - t^2$, we are unable to find a simple formula for an antiderivative of e^{-t^2} that does not involve a definite integral. We will learn more about finding (complicated) algebraic formulas for antiderivatives without definite integrals in the chapter on infinite series.

[3] The error function is defined by the rule $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\, dt$ and has the key property that $0 \le \operatorname{erf}(x) < 1$ for all $x \ge 0$ and moreover that $\lim_{x\to\infty} \operatorname{erf}(x) = 1$.

Returning our attention to the function E, while we cannot evaluate E exactly for any value other than $x = 0$, we still can gain a tremendous amount of information about the function E. To begin, applying the rule in Equation (4.4) to E, it follows that

$$E'(x) = \frac{d}{dx}\left[\int_0^x e^{-t^2}\ dt\right] = e^{-x^2},$$

so we know a formula for the derivative of E. Moreover, we know that $E(0) = 0$. This information is precisely the type we were given in problems such as the one in Activity 3.2–1 and others in Section **??**, where we were given information about the derivative of a function, but lacked a formula for the function itself.

Here, using the first and second derivatives of E, along with the fact that $E(0) = 0$, we can determine more information about the behavior of E. First, with $E'(x) = e^{-x^2}$, we note that for all real numbers x, $e^{-x^2} > 0$, and thus $E'(x) > 0$ for all x. Thus E is an always increasing function. Further, we note that as $x \to \infty$, $E'(x) = e^{-x^2} \to 0$, hence the slope of the function E tends to zero as $x \to \infty$ (and similarly as $x \to -\infty$). This tells us that E has horizontal asymptotes as x increases or decreases without bound.

In addition, we can observe that $E''(x) = -2xe^{-x^2}$, and that $E''(0) = 0$, while $E''(x) < 0$ for $x > 0$ and $E''(x) > 0$ for $x < 0$. This information tells us that E is concave up for $x < 0$ and concave down for $x > 0$ with a point of inflection at $x = 0$.

The only thing we lack at this point is a sense of how big E can get as x increases. If we use a midpoint Riemann sum with 10 subintervals to estimate $E(2)$, we see that $E(2) \approx 0.8822$; a similar calculation to estimate $E(3)$ shows little change ($E(3) \approx 0.8862$), so it appears that as x increases without bound, E approaches a value just larger than 0.886. Putting all of this information together (and using the symmetry of $f(t) = e^{-t^2}$), we see the results shown in Figure 4.44.

Again, E is the antiderivative of $f(t) = e^{-t^2}$ that satisfies $E(0) = 0$. Moreover, the values on the graph of $y = E(x)$ represent the net-signed area of the region bounded by $f(t) = e^{-t^2}$ from 0 up to x. We see that the value of E increases rapidly near zero but then levels off x increases since there is less and less additional accumulated area bounded by $f(t) = e^{-t^2}$ as x increases.

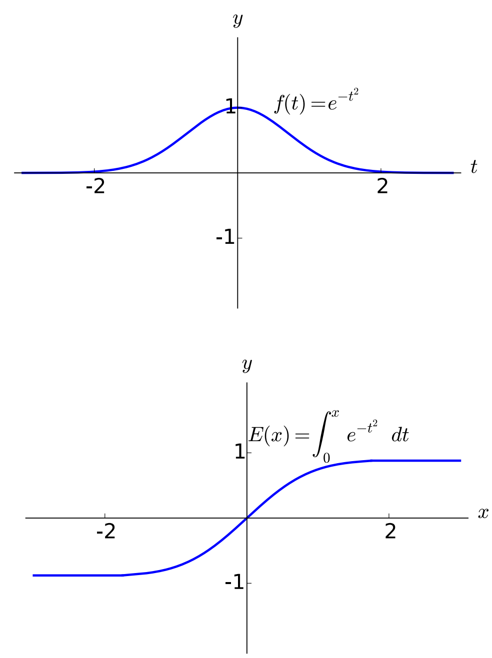


Figure 4.44: At top, the graph of $f(t) = e^{-t^2}$. At bottom, the integral function $E(x) = \int_0^x e^{-t^2}\, dt$, which is the unique antiderivative of f that satisfies $E(0) = 0$.

Activity 4.4–4

Suppose that $f(t) = \dfrac{t}{1+t^2}$ and $F(x) = \int_0^x f(t)\, dt$.

(a) Plot a graph of $f(t) = \dfrac{t}{1+t^2}$ on the interval $-10 \le t \le 10$. Clearly label the vertical axes with appropriate scale.

(b) What is the key relationship between F and f, according to the FTC?

(c) Use the first derivative test to determine the intervals on which F is increasing and decreasing.

(d) Use the second derivative test to determine the intervals on which F

is concave up and concave down. Note that $f'(t)$ can be simplified to be written in the form $f'(t) = \dfrac{1-t^2}{(1+t^2)^2}$.

(e) Using technology appropriately, estimate the values of $F(5)$ and $F(10)$ through appropriate Riemann sums.

(f) Sketch an accurate graph of $y = F(x)$, and clearly label the vertical axes with appropriate scale.

Summary

In this section, we encountered the following important ideas:

- Given a function f, the rule $A(x) = \int_a^x f(t)\,dt$ defines a new function A that measures the net-signed area bounded by f on the interval $[a, x]$. We call the function A the integral function corresponding to f.
- The Fundamental Theorem of Calculus part I is the following: if f is a continuous function and a is any constant, then $A(x) = \int_a^x f(t)\,dt$ is the unique antiderivative of f that satisfies $A(c) = 0$.
- The Chain Rule can be applied to functions of the form: $F\big(g(x)\big) = \int_a^{g(x)} f(t)\,dt$.

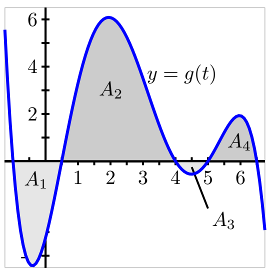
$$\frac{d}{dx}\Big(F\big(g(x)\big)\Big) = F'\big(g(x)\big)g'(x) = f\big(g(x)\big)g'(x).$$

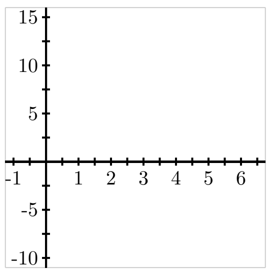
Exercises

Problems

1) Let g be the function pictured below at left, and let F be defined by $F(x) = \int_2^x g(t)\,dt$. Assume that the shaded areas have values $A_1 = 4.3$, $A_2 = 12.7$, $A_3 = 0.4$, and $A_4 = 1.8$. Assume further that the portion of A_2 that lies between $x = 0.5$ and $x = 2$ is 5.4.

   Sketch a carefully labeled graph of F on the axes provided, and include a written analysis of how you know where g is zero, increasing, decreasing, CCU, and CCD.





2) The tide removes sand from the beach at a small ocean park at a rate modeled by the function

$$R(t) = 2 + 5\sin\left(\frac{4\pi t}{25}\right)$$

   A pumping station adds sand to the beach at rate modeled by the function

$$S(t) = \frac{15t}{1+3t}$$

   Both $R(t)$ and $S(t)$ are measured in cubic yards of sand per hour, t is measured in hours, and the valid times are $0 \le t \le 6$. At time $t = 0$, the beach holds 2500 cubic yards of sand.

   (a) What definite integral measures how much sand the tide will remove during the time period $0 \le t \le 6$? Why?

   (b) Write an expression for $Y(x)$, the total number of cubic yards of sand on the beach at time x. Carefully explain your thinking and reasoning.

   (c) At what instantaneous rate is the total number of cubic yards of sand on the beach at time $t = 4$ changing?

   (d) Over the time interval $0 \le t \le 6$, at what time t is the amount of sand on the beach least? What is this minimum value? Explain and justify your answers fully.

3) When an aircraft attempts to climb as rapidly as possible, its climb rate (in feet per minute) decreases as altitude increases, because the air is less dense at higher altitudes. Given below is a table showing performance data for a certain single engine aircraft, giving its climb rate at various altitudes, where $c(h)$ denotes the climb rate of the airplane at an altitude h.

h (feet)	0	1000	2000	3000	4000	5000
c (ft/min)	925	875	830	780	730	685
h (feet)	6000	7000	8000	9000	10,000	
c (ft/min)	635	585	535	490	440	

   Let a new function m, that also depends on h, (say $y = m(h)$) measure the number of minutes required for a plane at altitude h to climb the next foot of altitude.

   a. Determine a similar table of values for $m(h)$ and explain how it is related to the table above. Be sure to discuss the units on m.

   b. Give a careful interpretation of a function whose derivative is $m(h)$. Describe what the input is and what the output is. Also, explain in plain English what the function tells us.

   c. Determine a definite integral whose value tells us exactly the number of minutes required for the airplane to ascend to 10,000 feet of altitude. Clearly explain why the value of this integral has the required meaning.

   d. Determine a formula for a function $M(h)$ whose value tells us the exact number of minutes required for the airplane to ascend to h feet of altitude.

   e. Estimate the values of $M(6000)$ and $M(10000)$ as accurately as you can. Include units on your results.

In exercises 4–7, find $F'(x)$.

4) $F(x) = \displaystyle\int_2^{x^3+x} \frac{1}{t}\,dt$

5) $F(x) = \displaystyle\int_{x^3}^{0} t^3\,dt$

6) $F(x) = \displaystyle\int_{x}^{x^2} (t+2)\,dt$

7) $F(x) = \displaystyle\int_{\ln x}^{e^x} \sin t\,dt$

4.5 *The Fundamental Theorem of Calculus, part II*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- How can we find the exact value of a definite integral without taking the limit of a Riemann sum?
- What is the statement of the Fundamental Theorem of Calculus Part II, and how do antiderivatives of functions play a key role in applying the theorem?
- What is the meaning of the definite integral of a rate of change in contexts other than when the rate of change represents velocity?
- What is an indefinite integral and how is its notation used in discussing antiderivatives?
- How do the First and Second Fundamental Theorems of Calculus enable us to formally see how differentiation and integration are almost inverse processes?

Introduction

Much of our work in Chapter 4 has been motivated by the velocity-distance problem: if we know the instantaneous velocity function, $v(t)$, for a moving object on a given time interval $[a,b]$, can we determine its exact distance traveled on $[a,b]$? In the vast majority of our discussion in Sections 4.1-4.3, we have focused on the fact that this distance traveled is connected to the area bounded by $y = v(t)$ and the t-axis on $[a,b]$. In particular, for any nonnegative velocity function $y = v(t)$ on $[a,b]$, we know that the exact area bounded by the velocity curve and the t-axis on the interval tells us the total distance traveled, which is also the value of the definite integral $\int_a^b v(t)\,dt$. In the situation where velocity is sometimes negative, the total area bounded by the velocity function still tells us distance traveled, while the net signed area that the function bounds tells us the object's change in position. Recall, for instance, the introduction to Section 4.2, where we observed that for the velocity function in Figure 4.45, the total distance D traveled by the moving object on $[a,b]$ is

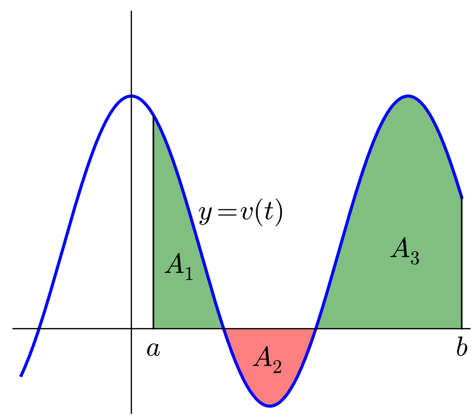


Figure 4.45: A velocity function.

$$D = A_1 + A_2 + A_3,$$

while the total change in the object's position on $[a,b]$ is

$$s(b) - s(a) = A_1 - A_2 + A_3.$$

While the areas A_1, A_2, and A_3, which are each given by definite integrals, may be computed through limits of Riemann sums (and in select special circumstances through familiar geometric formulas), in the present section we turn our attention to an alternate approach, similar to the one we encountered in Activity 4.1–2. To explore these ideas further, we consider the following preview activity.

Preview Activity 4.5

A student with a third floor dormitory window 32 feet off the ground tosses a water balloon straight up in the air with an initial velocity of 16 feet per second. It turns out that the instantaneous velocity of the water balloon is given by the velocity function $v(t) = -32t + 16$, where v is measured in feet per second and t is measured in seconds.

(a) Let $s(t)$ represent the height of the water balloon above the ground at time t, and note that s is an antiderivative of v. That is, v is the derivative of s: $v(t) = s'(t)$. Find a formula for $s(t)$ that satisfies the initial condition that the balloon is tossed from 32 feet above ground. In other words, make your formula for s satisfy $s(0) = 32$.

(b) At what time does the water balloon reach its maximum height? At what time does the water balloon land?

(c) Compute the three differences $s\left(\frac{1}{2}\right) - s(0)$, $s(2) - s\left(\frac{1}{2}\right)$, and $s(2) - s(0)$. What do these differences represent?

(d) What is the total vertical distance traveled by the water balloon from the time it is tossed until the time it lands?

(e) Sketch a graph of the velocity function $y = v(t)$ on the time interval $[0, 2]$. What is the total net signed area bounded by $y = v(t)$ and the t-axis on $[0, 2]$? Answer this question in two ways: first by using your work above, and then by using a familiar geometric formula to compute areas of certain relevant regions.

The Fundamental Theorem of Calculus, Part II

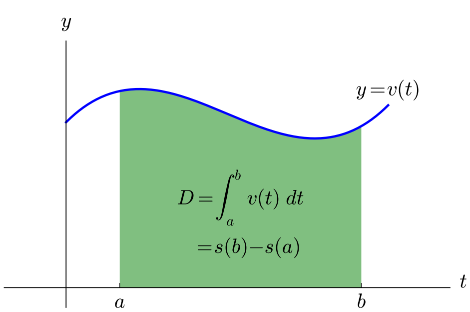


Figure 4.46: Finding the distance traveled when we know an object's velocity function v.

Consider the setting where we know the position function $s(t)$ of an object moving along an axis, as well as its corresponding velocity function $v(t)$, and for the moment let us assume that $v(t)$ is positive on $[a, b]$. Then, as shown in Figure 4.46, we know two different perspectives on the distance, D, the object travels: one is that $D = s(b) - s(a)$, which is the object's change in position. The other is that the distance traveled is the area under the velocity curve, which is given by the definite integral, so $D = \int_a^b v(t)\,dt$.

Of course, since both of these expressions tell us the distance traveled, it follows that they are equal, so

$$s(b) - s(a) = \int_a^b v(t)\,dt. \tag{4.5}$$

Furthermore, we know that Equation (4.5) holds even when velocity is sometimes negative, since $s(b) - s(a)$ is the object's change in position over $[a, b]$, which is simultaneously measured by the total net signed area on $[a, b]$ given by $\int_a^b v(t)\,dt$.

Perhaps the most powerful part of Equation (4.5) lies in the fact that we can compute the integral's value if we can find a formula for s. Remember, s and v are related by the fact that v is the derivative of s, or equivalently that s is an *antiderivative* of v. For example, if we have an object whose velocity is $v(t) = 3t^2 + 40$ feet per second (which is always nonnegative), and wish to know the distance traveled on the interval $[1,5]$, we have that

$$\begin{aligned} D &= \int_1^5 v(t)\,dt \\ &= \int_1^5 (3t^2+40)\,dt \\ &= s(5) - s(1), \end{aligned}$$

where s is an antiderivative of v. We know that the derivative of t^3 is $3t^2$ and that the derivative of $40t$ is 40, so it follows that if $s(t) = t^3 + 40t$, then s is a function whose derivative is $v(t) = s'(t) = 3t^2 + 40$, and thus we have found an antiderivative of v. Therefore,

$$\begin{aligned} D &= \int_1^5 3t^2 + 40\,dt \\ &= s(5) - s(1) \\ &= (5^3 + 40 \cdot 5) - (1^3 + 40 \cdot 1) \\ &= 284 \text{ feet.} \end{aligned}$$

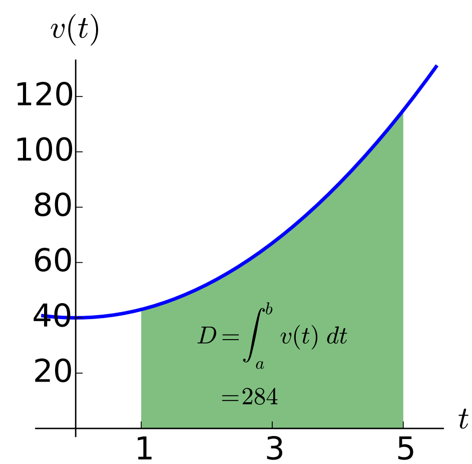


Figure 4.47: The exact area of the region enclosed by $v(t) = 3t^2 + 40$ on $[1,5]$.

Note the key lesson of this example: to find the distance traveled, we needed to compute the area under a curve, which is given by the definite integral. But to evaluate the integral, we found an antiderivative, s, of the velocity function, and then computed the total change in s on the interval. In particular, observe that we have found the exact area of the region shown in Figure 4.47, and done so without a familiar formula (such as those for the area of a triangle or circle) and without directly computing the limit of a Riemann sum. As we proceed to thinking about contexts other than just velocity and position, it turns out to be advantageous to have a shorthand symbol for a function's antiderivative. In the general setting of a continuous function f, we will often denote an antiderivative of f by F, so that the relationship between F and f is that $F'(x) = f(x)$ for all relevant x. Using the notation V in place of s (so that V is an antiderivative of v) in Equation (4.5), we find it is equivalent to write that

$$V(b) - V(a) = \int_a^b v(t)\,dt. \tag{4.6}$$

Now, in the general setting of wanting to evaluate the definite integral $\int_a^b f(x)\,dx$ for an arbitrary continuous function f, we could certainly think of f as representing the velocity of some moving object, and x as the variable that represents time. And again, Equations (4.5) and (4.6) hold for any continuous velocity function, even when v is sometimes negative. This leads us to see that Equation (4.6) tells us something even more important than the change in position of a moving object: it offers a shortcut route to evaluating any definite integral, provided that we can find an antiderivative of the integrand. The second part of the Fundamental Theorem of Calculus (FTCII) summarizes these observations.

The Fundamental Theorem of Calculus, part II

If f is a continuous function on $[a,b]$, and F is any antiderivative of f, then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Proof: We can use the first part of the Fundamental Theorem to prove the second part. In the first part, we showed that the integral, or area, function

$$A(x) = \int_a^x f(t)\,dt \quad \text{for} \quad a \le x \le b$$

is an antiderivative of the function f. We have also shown by the Mean Value Theorem that if two functions have the same derivative, then they must differ by only a constant. In other words, if both A and F are antiderivatives of f, then $F(x) = A(x) + C$, where C is an arbitrary constant.

Recall that $A(a) = 0$, which implies that $F(a) = A(a) + C = C$. Notice also that $F(b) = A(b) + C$, and we have

$$\begin{aligned} F(b) - F(a) &= [A(b) + C] - [A(a) + C] \\ &= A(b) - A(a) \\ &= A(b). \end{aligned}$$

And what is $A(b)$ with respect to the area function? $A(b) = \int_a^b f(t)\,dt$, which is the same as $\int_a^b f(x)\,dx$. So

$$F(b) - F(a) = \int_a^b f(x)\,dx.$$

A common alternate notation for $F(b) - F(a)$ is

$$F(b) - F(a) = F(x)\Big|_a^b$$

where we read the righthand side as "the function F evaluated from a to b." In this notation, the FTCII says that

$$\int_a^b f(x)\,dx = F(x)\Big|_a^b.$$

The FTCII opens the door to evaluating exactly a wide range of integrals. In particular, if we are interested in a definite integral for which we can find an antiderivative F for the integrand f, then we can evaluate the integral exactly. For instance since $\frac{d}{dx}[\frac{1}{3}x^3] = x^2$, the FTCII tells us that

$$\begin{aligned}\int_0^1 x^2\,dx &= \frac{1}{3}\,x^3\Big|_0^1 \\ &= \frac{1}{3}\,(1)^3 - \frac{1}{3}\,(0)^3 \\ &= \frac{1}{3}.\end{aligned}$$

Example 1

We spent a great deal of time in the previous sections studying $\int_0^4 (4x - x^2)\,dx$. Using the second part of the Fundamental Theorem of Calculus (FTCII), let's evaluate this definite integral.

Solution. We need an antiderivative of $f(x) = 4x - x^2$, or a function whose derivative is $4x - x^2$. Such a function, and hence antiderivative of $f(x)$, would be $F(x) = 2x^2 - \frac{1}{3}x^3$.

By FTCII,

$$\int_a^b f(x) = F(b) - F(a);$$

therefore,

$$\begin{aligned}\int_0^4 \left(4x - x^2\right)\,dx &= F(4) - F(0) \\ &= 2x^2 - \frac{1}{3}x^3\Big|_0^4 \\ &= \left(2(4)^2 - \frac{1}{3}4^3\right) - (0-0) \\ &= 32 - \frac{64}{3} = \frac{32}{3}.\end{aligned}$$

Notice that this is the same answer we obtained using limits in the previous sections, just with much less work.

Example 2

Suppose a ball is thrown straight up with velocity given by $v(t) = -32t + 20$ ft/s, where t is measured in seconds. Find, and interpret, $\int_0^1 v(t)\ dt$.

Solution. Using the second part of the Fundamental Theorem of Calculus, we have

$$\begin{aligned}\int_0^1 v(t)\ dt &= \int_0^1 (-32t + 20)\ dt \\ &= -16t^2 + 20t\Big|_0^1 \\ &= 4\end{aligned}$$

Thus if a ball is thrown straight up into the air with velocity $v(t) = -32t + 20$, the height of the ball, 1 second later, will be 4 feet above the initial height. (Note that the ball has *traveled* much farther. It has gone up to its peak and is falling down, but the difference between its height at $t = 0$ and $t = 1$ is 4 feet.)

But finding an antiderivative can be far from simple; in fact, often finding a formula for an antiderivative is very hard or even impossible. While we can differentiate just about any function, even some relatively simple ones don't have an elementary antiderivative. A significant portion of integral calculus (which is the main focus of second semester college calculus) is devoted to understanding the problem of finding antiderivatives.

Basic antiderivatives

The general problem of finding an antiderivative is difficult. In part, this is due to the fact that we are trying to undo the process of differentiating, and the undoing is much more difficult than the doing. For example, while it is evident that an antiderivative of $f(x) = \sin(x)$ is $F(x) = -\cos(x)$ and that an antiderivative of $g(x) = x^2$ is $G(x) = \frac{1}{3}x^3$, combinations of f and g can be far more complicated. Consider such functions as

$$5\sin(x) - 4x^2, \ \ x^2\sin(x), \ \ \frac{\sin(x)}{x^2}, \ \text{ and } \ \sin(x^2).$$

What is involved in trying to find an antiderivative for each? From our experience with derivative rules, we know that while derivatives of sums and constant multiples of basic functions are simple to execute, derivatives involving products, quotients, and composites of familiar functions are much more complicated. Thus, it stands to reason that antidifferentiating products, quotients, and composites of basic functions may be even more challenging. We defer our study of all but the most elementary antiderivatives to later in the text.

We do note that each time we have a function for which we know its derivative, we have a *function-derivative pair*, which also leads us to knowing the antiderivative of a function. For instance, since we know that

$$\frac{d}{dx}[-\cos(x)] = \sin(x),$$

it follows that $F(x) = -\cos(x)$ is an antiderivative of $f(x) = \sin(x)$. It is equivalent to say that $f(x) = \sin(x)$ is the derivative of $F(x) = -\cos(x)$, and thus F and f together form the function-derivative pair. Clearly, every basic derivative rule leads us to such a pair, and thus to a known antiderivative. In Activity 4.7–2, we will construct a list of most of the basic antiderivatives we know at this time. Furthermore, those rules will enable us to antidifferentiate sums and constant multiples of basic functions. For example, if $f(x) = 5\sin(x) - 4x^2$, note that since $-\cos(x)$ is an antiderivative of $\sin(x)$ and $\frac{1}{3}x^3$ is an antiderivative of x^2, it follows that

$$F(x) = -5\cos(x) - 4 \cdot \frac{1}{3}x^3$$

is an antiderivative of f, by the sum and constant multiple rules for differentiation.

Activity 4.5–1

Use your knowledge of derivatives of basic functions to complete the table of antiderivatives. For each entry, your task is to find a function F whose derivative is the given function f. When finished, use the FTCII and the results in the table to evaluate the given definite integrals.

(a) $\int_{-2}^{2} x^3 \, dx$

(b) $\int_{0}^{\pi} \sin x \, dx$

(c) $\int_{0}^{5} e^t \, dt$

(d) $\int_{4}^{9} \sqrt{u} \, du$

(e) $\int_{1}^{5} 2 \, dx$

(f) $\int_{0}^{1} \left(x^3 - x - e^x + 2\right) dx$

(g) $\int_{0}^{\pi/3} (2\sin(t) - 4\cos(t) + \sec^2(t) - \pi) \, dt$

(h) $\int_{0}^{1} (\sqrt{x} - x^2) \, dx$

Function	Antiderivative
k, (k is constant)	
x^n, $n \neq -1$	
$\frac{1}{x}$, $x > 0$	
$\sin(x)$	
$\cos(x)$	
$\sec(x)\tan(x)$	
$\csc(x)\cot(x)$	
$\sec^2(x)$	
$\csc^2(x)$	
e^x	
a^x $(a > 1)$	
$\frac{1}{1+x^2}$	
$\frac{1}{\sqrt{1-x^2}}$	

Table 4.2: Familiar basic functions and their antiderivatives.

We now revisit the fact that each function has more than one antiderivative. Because the derivative of any constant is zero, any time we seek an arbitrary antiderivative, we may add a constant of our choice. For instance, if we want to determine an antiderivative of $g(x) = x^2$, we know that $G(x) = \frac{1}{3}x^3$ is one such function. But we could alternately have chosen $G(x) = \frac{1}{3}x^3 + 7$, since in this case as well, $G'(x) = x^2$. In some contexts later on in calculus, it is important to discuss the most general antiderivative of a function. If $g(x) = x^2$, we say that the *general antiderivative* or *indefinite integral* of g is

$$G(x) = \frac{1}{3}x^3 + C,$$

where C represents an arbitrary real number constant. Regardless of the formula for g, including $+C$ in the formula for its antiderivative G results in the most general possible antiderivative.

General Antiderivatives/Indefinite Integrals

Let a function $f(x)$ be given. A **general antiderivative** or **indefinite integral** of $f(x)$ is the infinite family of functions

$$F(x) + C$$

where C is an arbitrary constant such that

$$F'(x) = f(x).$$

Given a function f and one of its antiderivatives F, we know *all* antiderivatives of f have the form $F(x) + C$ for some constant C. Using the above definition, we can say that

$$\int f(x)\, dx = F(x) + C.$$

Let's analyze this indefinite integral notation.

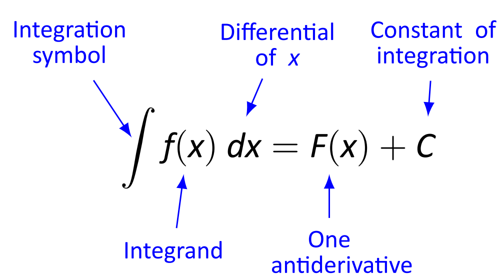


Figure 4.48: Understanding the indefinite integral notation.

Example 3

Evaluate $\int (3x^2 + 4x + 5)\ dx$.

Solution. We seek a function $F(x)$ whose derivative is $3x^2 + 4x + 5$. When taking derivatives, we can consider functions term-by-term, so we can also do that here.

What functions have a derivative of $3x^2$? Some thought will lead us to a cubic, specifically $x^3 + C_1$, where C_1 is a constant.

What functions have a derivative of $4x$? Here the x term is raised to the first power, so we likely seek a quadratic. Some thought should lead us to $2x^2 + C_2$, where C_2 is a constant.

Finally, what functions have a derivative of 5? Functions of the form $5x + C_3$, where C_3 is a constant.

Our answer appears to be

$$\int (3x^2 + 4x + 5)\ dx = x^3 + C_1 + 2x^2 + C_2 + 5x + C_3.$$

We do not need three separate constants of integration; combine them as one constant, giving the final answer of

$$\int (3x^2 + 4x + 5)\ dx = x^3 + 2x^2 + 5x + C.$$

It is easy to verify our answer; take the derivative of $x^3 + 2x^3 + 5x + C$ and see we indeed get $3x^2 + 4x + 5$.

This final step of "verifying our answer" is important both practically and theoretically. In general, taking derivatives is easier than finding antiderivatives so checking our work is easy and vital as we learn. We also see that taking the derivative of our answer returns the function in the integrand. Thus we can say that:

$$\frac{d}{dx}\left(\int f(x)\ dx\right) = f(x).$$

Differentiation "undoes" the work done by antidifferentiation.

Initial Value Problems

We have seen that the derivative of a position function gave a velocity function, and the derivative of a velocity function describes the acceleration. We can now go "the other way:" the antiderivative of an acceleration function gives a velocity function, etc. While there is just one derivative of a given function, there are infinite antiderivatives. Therefore we cannot ask "What is *the* velocity of an object whose acceleration is -32ft/s^2?", since there is more than one answer.

We can find *the* answer if we provide more information with the question, as done in the following example.

Example 4

The acceleration due to gravity of a falling object is -32 ft/s^2. At time $t = 3$, a falling object had a velocity of -10 ft/s. Find the equation of the object's velocity. We want to know a velocity function, $v(t)$. We know two things:

- The acceleration, i.e., $v'(t) = -32$, and
- the velocity at a specific time, i.e., $v(3) = -10$.

Using the first piece of information, we know that $v(t)$ is an antiderivative of $v'(t) = -32$. So we begin by finding the indefinite integral of -32:

$$\int (-32)\ dt = -32t + C = v(t).$$

Now we use the fact that $v(3) = -10$ to find C:

$$\begin{aligned} v(t) &= -32t + C \\ v(3) &= -10 \\ -32(3) + C &= -10 \\ C &= 86 \end{aligned}$$

Thus $v(t) = -32t + 86$. We can use this equation to understand the motion of the object: when $t = 0$, the object had a velocity of $v(0) = 86$ ft/s. Since the velocity is positive, the object was moving upward.

When did the object begin moving down? Immediately after $v(t) = 0$:

$$-32t + 86 = 0 \quad \Rightarrow \quad t = \frac{43}{16} \approx 2.69\text{s}.$$

Recognize that we are able to determine quite a bit about the path of the object knowing just its acceleration and its velocity at a single point in time.

Example 5

Find $f(t)$, given that $f''(t) = \cos t$, $f'(0) = 3$ and $f(0) = 5$. We start by finding $f'(t)$, which is an antiderivative of $f''(t)$:

$$\int f''(t)\ dt = \int \cos t\ dt = \sin t + C = f'(t).$$

So $f'(t) = \sin t + C$ for the correct value of C. We are given that $f'(0) = 3$, so:

$$f'(0) = 3 \quad \Rightarrow \quad \sin 0 + C = 3 \quad \Rightarrow \quad C = 3.$$

Using the initial value, we have found $f'(t) = \sin t + 3$.

We now find $f(t)$ by integrating again.

$$\int f'(t)\ dt = \int (\sin t + 3)\ dt = -\cos t + 3t + C.$$

We are given that $f(0) = 5$, so

$$\begin{aligned} -\cos 0 + 3(0) + C &= 5 \\ -1 + C &= 5 \\ C &= 6 \end{aligned}$$

Thus $f(t) = -\cos t + 3t + 6$.

Activity 4.5–2

Given the following velocity functions and initial positions of some objects, produce the position functions of the objects.

(a) $v(t) = 3t - 5$; $s(1) = 3$

(b) $v(t) = 8\sin(t)$; $s(\pi/6) = 8$

(c) $v(t) = \dfrac{3}{t} + 6$; $s(1) = 8$

Differentiating an Integral Function

We have seen that the Fundamental Theorem of Calculus (FTC) enables us to construct an antiderivative F of any continuous function f by defining A by the corresponding integral function $A(x) = \int_c^x f(t)\,dt$. Said differently, if we have a function of the form $A(x) = \int_c^x f(t)\,dt$, then we know that $A'(x) = \frac{d}{dx}\left[\int_c^x f(t)\,dt\right] = f(x)$. This shows that integral functions, while perhaps having the most complicated formulas of any functions we have encountered, are nonetheless particularly simple to differentiate. For instance, if

$$A(x) = \int_\pi^x \sin(t^2)\,dt,$$

then by the FTC, we know immediately that

$$A'(x) = \sin(x^2).$$

Restating this result more generally for an arbitrary function f, we know by the FTC that

$$\frac{d}{dx}\left[\int_a^x f(t)\,dt\right] = f(x).$$

In words, the last equation essentially says that "the derivative of the integral function whose integrand is f, is f." In this sense, we see that if we first integrate the function f from $t = a$ to $t = x$, and then differentiate with respect to x, these two processes "undo" one another.

Taking a different approach, say we begin with a function $f(t)$ and differentiate with respect to t. What happens if we follow this by integrating the result from $t = a$ to $t = x$? That is, what can we say about the quantity

$$\int_a^x \frac{d}{dt}\left[f(t)\right]\, dt?$$

Here, we use the First FTC and note that $f(t)$ is an antiderivative of $\frac{d}{dt}\left[f(t)\right]$. Applying this result and evaluating the antiderivative function, we see that

$$\begin{aligned} \int_a^x \frac{d}{dt}\left[f(t)\right]\, dt &= f(t)\Big|_a^x \\ &= f(x) - f(a). \end{aligned}$$

Thus, we see that if we apply the processes of first differentiating f and then integrating the result from a to x, we return to the function f, minus the constant value $f(a)$. So in this situation, the two processes almost undo one another, up to the constant $f(a)$.

The observations made in the preceding two paragraphs demonstrate that differentiating and integrating (where we integrate from a constant up to a variable) are almost inverse processes. In one sense, this should not be surprising: integrating involves antidifferentiating, which reverses the process of differentiating. On the other hand, we see that there is some subtlety involved, as integrating the derivative of a function does not quite produce the function itself. This is connected to a key fact we observed, which is that any function has an entire family of antiderivatives, and any two of those antiderivatives differ only by a constant.

Summary

In this section, we encountered the following important ideas:

- We can find the exact value of a definite integral without taking the limit of a Riemann sum or using a familiar area formula by finding the antiderivative of the integrand, and hence applying the Fundamental Theorem of Calculus.
- The Fundamental Theorem of Calculus says that if f is a continuous function on $[a, b]$ and F is an antiderivative of f, then
$$\int_a^b f(x)\, dx = F(b) - F(a).$$
Hence, if we can find an antiderivative for the integrand f, evaluating the definite integral comes from simply computing the change in F on $[a, b]$.
- A slightly different perspective on the FTC allows us to restate it as the Total Change Theorem, which says that
$$\int_a^b f'(x)\, dx = f(b) - f(a),$$
for any continuously differentiable function f. This means that the definite integral of the instantaneous rate of change of a function f on an interval $[a, b]$ is equal to the total change in the function f on $[a, b]$.

Exercises

Terms and Concepts

1) Define the term "antiderivative" in your own words.
2) Is it more accurate to refer to "the" antiderivative of $f(x)$ or "an" antiderivative of $f(x)$?
3) Use your own words to define the indefinite integral of $f(x)$.
4) Fill in the blanks: "Inverse operations do the ________ things in the ________ order."
5) The derivative of a position function is a ________ function.
6) The antiderivative of an acceleration function is a ________ function.
7) How are definite and indefinite integrals related?
8) What constant of integration is most commonly used when evaluating definite integrals?
9) The definite integral can be used to find "the area under a curve." Give two other uses for definite integrals.

Problems

In exercises 4–27, evaluate the indefinite integral.

10) $\int 3x^3\,dx$
11) $\int x^8\,dx$
12) $\int (10x^2-2)\,dx$
13) $\int dt$
14) $\int 1\,ds$
15) $\int \frac{1}{3t^2}\,dt$
16) $\int \frac{3}{t^2}\,dt$
17) $\int \frac{1}{\sqrt{x}}\,dx$
18) $\int \sec^2\theta\,d\theta$
19) $\int \sin\theta\,d\theta$
20) $\int 5e^\theta\,d\theta$
21) $\int 3^t\,dt$
22) $\int \frac{5^t}{2}\,dt$
23) $\int (2t+3)^2\,dt$
24) $\int (t^2+3)(t^3-2t)\,dt$
25) $\int x^2x^3\,dx$
26) $\int e^\pi\,dx$
27) $\int t\,dx$

In exercises 28–38, solve the initial value problem.

28) $f'(x)=\sin x$ and $f(0)=2$
29) $f'(x)=5e^x$ and $f(0)=10$
30) $f'(x)=4x^3-3x^2$ and $f(-1)=9$
31) $f'(x)=\sec^2 x$ and $f(\pi/4)=5$
32) $f'(x)=7^x$ and $f(2)=1$
33) $f''(x)=5$ and $f'(0)=7$, $f(0)=3$
34) $f''(x)=7x$ and $f'(1)=-1$, $f(1)=10$
35) $f''(x)=5e^x$ and $f'(0)=3$, $f(0)=5$
36) $f''(x)=\sin\theta$ and $f'(\pi)=2$, $f(\pi)=4$
37) $f''(x)=24x^2+2^x-\cos x$ and $f'(0)=5$, $f(0)=0$
38) $f''(x)=0$ and $f'(1)=3$, $f(1)=1$

In exercises 39–51, evaluate the definite integral.

39) $\int_1^3 (3x^2-2x+1)\,dx$
40) $\int_0^4 (x-1)^2\,dx$
41) $\int_{-1}^1 (x^3-x^5)\,dx$
42) $\int_{\pi/2}^{\pi} \cos x\,dx$
43) $\int_0^{\pi/4} \sec^2 x\,dx$
44) $\int_1^e \frac{1}{x}\,dx$
45) $\int_{-1}^1 5^x\,dx$
46) $\int_{-2}^{-1} (4-2x^3)\,dx$
47) $\int_1^3 e^x\,dx$
48) $\int_0^4 \sqrt{t}\,dt$
49) $\int_9^{25} \frac{1}{\sqrt{t}}\,dt$
50) $\int_1^8 \sqrt[3]{x}\,dx$
51) $\int_1^2 \frac{1}{x}\,dx$
52) $\int_1^2 \frac{1}{x^2}\,dx$
53) $\int_1^2 \frac{1}{x^3}\,dx$
54) $\int_0^1 x\,dx$
55) $\int_0^1 x^2\,dx$
56) $\int_0^1 x^3\,dx$
57) $\int_0^1 x^{100}\,dx$
58) $\int_{-4}^4 dx$
59) $\int_{-10}^{-5} 3\,dx$
60) $\int_{-2}^2 0\,dx$
61) $\int_{\pi/6}^{\pi/3} \csc x\cot x\,dx$
62) $\int_0^\pi (2\cos x-2\sin x)\,dx$
63) Explain why:
    (a) $\int_{-1}^1 x^n\,dx=0$, when n is a positive, odd integer, and
    (b) $\int_{-1}^1 x^n\,dx=2\int_0^1 x^n\,dx$ when n is a positive, even integer.

In exercises 64–68, a velocity function of an object moving along a straight line is given. Find the displacement of the object over the given time interval.

64) $v(t)=-32t+20$ ft/s on $[0,5]$
65) $v(t)=-32t+200$ ft/s on $[0,10]$
66) $v(t)=2^t$ mph on $[-1,1]$
67) $v(t)=\cos t$ ft/s on $[0,3\pi/2]$
68) $v(t)=\sqrt[4]{t}$ ft/s on $[0,16]$

In Exercises 69–72, an acceleration function of an object moving along a straight line is given. Find the change of the object's velocity over the given time interval.

69) $a(t) = -32$ ft/s^2 on $[0,2]$

70) $a(t) = 10$ ft/s^2 on $[0,5]$

71) $a(t) = t$ ft/s^2 on $[0,2]$

72) $a(t) = \cos t$ ft/s^2 on $[0,\pi]$

73) A function f is given piecewise by the formula

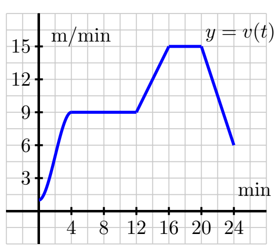
$$f(x) = \begin{cases} -x^2+2x+1, & \text{if } 0 \le x < 2 \\ -x+3, & \text{if } 2 \le x < 3 \\ x^2-8x+15, & \text{if } 3 \le x \le 5 \end{cases}$$

(a) Determine the exact value of the net signed area enclosed by f and the x-axis on the interval $[2,5]$.

(b) Compute the exact average value of f on $[0,5]$.

(c) Find a formula for a function g on $5 \le x \le 7$ so that if we extend the above definition of f so that $f(x) = g(x)$ if $5 \le x \le 7$, it follows that $\int_0^7 f(x)\,dx = 0$.

74) The instantaneous velocity (in meters per minute) of a moving object is given by the function v as pictured below. Assume that on the interval $0 \le t \le 4$, $v(t)$ is given by $v(t) = -\frac{1}{4}t^3 + \frac{3}{2}t^2 + 1$, and that on every other interval v is piecewise linear, as shown.



(a) Determine the exact distance traveled by the object on the time interval $0 \le t \le 4$.

(b) What is the object's average velocity on $[12,24]$?

(c) At what time is the object's acceleration greatest?

(d) Suppose that the velocity of the object is increased by a constant value c for all values of t. What value of c will make the object's total distance traveled on $[12,24]$ be 210 meters?

75) When an aircraft attempts to climb as rapidly as possible, its climb rate (in feet per minute) decreases as altitude increases, because the air is less dense at higher altitudes. Given below is a table showing performance data for a certain single engine aircraft, giving its climb rate at various altitudes, where $c(h)$ denotes the climb rate of the airplane at an altitude h.

h (feet)	0	1000	2000	3000	4000	5000
c (ft/min)	925	875	830	780	730	685
h (feet)	6000	7000	8000	9000	10,000	
c (ft/min)	635	585	535	490	440	

Let a new function called $m(h)$ measure the number of minutes required for a plane at altitude h to climb the next foot of altitude.

(a) Determine a similar table of values for $m(h)$ and explain how it is related to the table above. Be sure to explain the units.

(b) Give a careful interpretation of a function whose derivative is $m(h)$. Describe what the input is and what the output is. Also, explain in plain English what the function tells us.

(c) Determine a definite integral whose value tells us exactly the number of minutes required for the airplane to ascend to 10,000 feet of altitude. Clearly explain why the value of this integral has the required meaning.

(d) Use the Riemann sum M_5 to estimate the value of the integral you found in (c). Include units on your result.

4.6 Integration by Substitution

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- How can we begin to find algebraic formulas for antiderivatives of more complicated algebraic functions?
- How does the technique of u-substitution work to help us evaluate certain indefinite and definite integrals, and how does this process rely on identifying function-derivative pairs?

Introduction

In Section 4.5, we learned the key role that antiderivatives play in the the process of evaluating definite integrals exactly. In particular, we know that if F is any antiderivative of f, then

$$\int_a^b f(x)\, dx = F(b) - F(a).$$

Furthermore, we realized that each elementary derivative rule developed in Chapter 2 leads to a corresponding elementary antiderivative, as summarized in Table 4.2. Thus, if we wish to evaluate an integral such as

$$\int_0^1 \left(x^3 - \sqrt{x} + 5^x\right)\, dx,$$

it is straightforward to do so, since we can easily antidifferentiate $f(x) = x^3 - \sqrt{x} + 5^x$.

Because an algebraic formula for an antiderivative of f enables us to evaluate the definite integral $\int_a^b f(x)\, dx$ exactly, we have a natural interest in being able to find such algebraic antiderivatives.[4]

However, suppose we wish to evaluate the following indefinite integral.

$$\int 2x\sqrt{1+x^2}\, dx$$

Referencing Table 4.2, notice that none of the antiderivatives involve a linear function of x multiplied by a root function of x, which might tell us that none of the derivative rules we learned in Chapter 2 give us a way to antidifferentiate the above integrand. But in fact,

$$\int 2x\sqrt{1+x^2}\, dx = \frac{2}{3}\left(1+x^2\right)^{3/2} + C.$$

[4] Note that we emphasize *algebraic* antiderivatives, as opposed to any antiderivative, since we know by the Fundamental Theorem of Calculus that $G(x) = \int_a^x f(t)\, dt$ is indeed an antiderivative of the given function f, but one that still involves a definite integral.

We know that we can easily check the evaluation of any indefinite integral simply by differentiating. Doing so, we have

$$\begin{aligned}\frac{d}{dx}\left[\frac{2}{3}\left(1+x^2\right)^{3/2}+C\right] &= \frac{2}{3}\cdot\frac{3}{2}\left(1+x^2\right)^{1/2}\cdot(2x)+0\\ &= \left(1+x^2\right)^{1/2}\cdot(2x)\\ &= 2x\sqrt{1+x^2}.\end{aligned}$$

While differentiating, we used a very specific rule from Chapter 2—the Chain Rule! In fact, in every problem we encounter in this section, if we wish to check our integration by differentiating, we must use the Chain Rule. Our goal in this section is to develop a method of integrating such that we "undo" the Chain Rule for differentiation.

Preview Activity 4.6

In Section 2.6, we learned the Chain Rule and how it can be applied to find the derivative of a composite function. In particular, if g is a differentiable function of x, and f is a differentiable function of $g(x)$, then

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

In words, we say that the derivative of a composite function $c(x) = f(g(x))$, where f is considered the "outer" function and u the "inner" function, is "the derivative of the outer function, evaluated at the inner function, times the derivative of the inner function."

For each of the following functions, use the Chain Rule to find the function's derivative.

(a) $f(x) = e^{3x^2}$

(b) $h(x) = \sin(3x+4)$

(c) $p(x) = \arctan\left(\sqrt{x}\right)$

(d) $q(x) = (2-7x)^4$

(e) $r(x) = \ln\left(4-11x^3\right)$

Reversing the Chain Rule: u-substitution

It is important to explicitly remember that differentiation and antidifferentiation are essentially inverse processes; that they are not quite inverse processes is due to the $+C$ that arises when antidifferentiating. This close relationship enables us to take any known derivative rule and translate it to a corresponding rule for an indefinite integral.

Restating the relationship defined by the Chain Rule given in Preview Activity 4.6 in terms of an indefinite integral, we have

$$\int f'(g(x))g'(x)\,dx = f(g(x))+C. \tag{4.7}$$

Hence, Equation (4.7) tells us that if we can take a given function and view its algebraic structure as $f'(g(x))g'(x)$ for some appropriate choices of f and g, then we can antidifferentiate the function by reversing the Chain Rule. It is especially notable that both $g(x)$ and $g'(x)$ appear in the form of $f'(g(x))g'(x)$; we will sometimes say that we seek to *identify a function-derivative pair* when trying to apply the rule in Equation (4.7).

In the situation where we can identify a function-derivative pair, we will introduce a new variable u to represent the function $g(x)$. Observing that with $u = g(x)$, it follows in Leibniz notation that $\frac{du}{dx} = g'(x)$, so that in terms of differentials[5], $du = g'(x)\,dx$. Now converting the indefinite integral of interest to a new one in terms of u, we have

$$\int f'(g(x))g'(x)\ dx = \int f'(u)\ du.$$

To emphasize the importance of this concept, we restate it.

Integration by Substitution

Let f and g be differentiable functions, where the range of g is an interval I and the domain of f is contained in I. Then

$$\int f'(g(x))g'(x)\ dx = f(g(x)) + C.$$

If $u = g(x)$, then $du = g'(x)\ dx$ and

$$\int f'(g(x))g'(x)\ dx = \int f'(u)\ du = f(u) + C = f(g(x)) + C.$$

When f' is an elementary function whose antiderivative is known, we can easily evaluate the indefinite integral in u, and then go on to determine the desired overall antiderivative of $f'(g(x))g'(x)$. We call this process ***u*-substitution**. To see u-substitution at work, we consider the following example.

Example 1

Evaluate the indefinite integral

$$\int 4x^3 \cdot \sin(x^4 + 3)\ dx$$

and check the result by differentiating.

Solution. We can make several key algebraic observations regarding the integrand, $4x^3 \cdot \sin(x^4 + 3)$.

[5] If we recall from the definition of the derivative that $\frac{du}{dx} \approx \frac{\triangle u}{\triangle x}$ and use the fact that $\frac{du}{dx} = g'(x)$, then we see that $g'(x) \approx \frac{\triangle u}{\triangle x}$. Solving for $\triangle u$, $\triangle u \approx g'(x)\triangle x$. It is this last relationship that, when expressed in "differential" notation enables us to write $du = g'(x)\,dx$ in the change of variable formula.

First, $\sin(x^4+3)$ is a composite function; as such, we know we'll need a more sophisticated approach to antidifferentiating.

Second, $4x^3$ is the derivative of (x^4+3). Thus, $4x^3$ and (x^4+3) are a *function-derivative* pair.

Furthermore, we know the antiderivative of $f(u) = \sin(u)$. The combination of these observations suggests that we can evaluate the given indefinite integral by reversing the chain rule through u-substitution.

Let $u = x^4 + 3$, hence $\frac{du}{dx} = 4x^3$. In differential notation, it follows that

$$du = 4x^3\,dx.$$

Observe the original indefinite integral may now be written

$$\int \sin(x^4+3)\cdot 4x^3\,dx,$$

and by substituting the expressions in terms of u for the expressions in terms of x (specifically u for x^4+3 and du for $4x^3\,dx$), it follows that

$$\int \sin(x^4+3)\cdot 4x^3\,dx = \int \sin(u)\cdot\,du.$$

Now we may evaluate the original integral by first evaluating the easier integral in u, followed by replacing u by the expression x^4+3. Doing so, we find

$$\begin{aligned}\int \sin(x^4+3)\cdot 4x^3\,dx &= \int \sin(u)\cdot\,du \\ &= -\cos(u)+C \\ &= -\cos(x^4+3)+C.\end{aligned}$$

To check our work, we observe by the Chain Rule that

$$\begin{aligned}\frac{d}{dx}\left[-\cos(x^4+3)+C\right] &= -(-\sin(x^4+3)\cdot 4x^3) \\ &= \sin(x^4+3)\cdot 4x^3,\end{aligned}$$

which is indeed the original integrand.

An essential observation about our work in Example 1 is that the u-substitution only worked because the function multiplying $\sin(x^4+3)$ was $4x^3$. If instead that function was x^2 or x^4, the substitution process may not (and likely would not) have worked. This is one of the primary challenges of antidifferentiation: slight changes in the integrand make tremendous differences. For instance, we can use u-substitution with $u = x^2$ and $du = 2x\,dx$ to find that

$$\begin{aligned}\int xe^{x^2}\,dx &= \int e^u\cdot\frac{1}{2}\,du \\ &= \frac{1}{2}\int e^u\,du \\ &= \frac{1}{2}e^u + C \\ &= \frac{1}{2}e^{x^2} + C.\end{aligned}$$

If, however, we consider the similar indefinite integral

$$\int e^{x^2}\,dx,$$

the missing x to multiply e^{x^2} makes the u-substitution $u = x^2$ no longer possible. Hence, part of the lesson of u-substitution is just how specialized the process is: it only applies to situations where, up to a missing constant, the integrand that is present is the result of applying the Chain Rule to a different, related function.

Let's look at some more examples of u-substitution.

Example 2

Evaluate $\displaystyle\int \frac{7}{-3x+1}\,dx$.

Solution. First notice this integrand is a composition of

$$f(x) = \frac{7}{x} \quad \text{and} \quad g(x) = -3x+1.$$

Therefore, we begin our substitution by letting $u = -3x+1$, which implies

$$\frac{du}{dx} = -3 \quad \Rightarrow \quad du = -3\,dx.$$

The integrand lacks a -3; hence divide the previous equation by -3 to obtain

$$-\frac{1}{3}du = dx.$$

We can now evaluate the integral through substitution.

$$\begin{aligned}\int \frac{7}{-3x+1}\,dx &= \int \frac{7}{u}\cdot -\frac{1}{3}du \\ &= -\frac{7}{3}\int \frac{1}{u}\,du \\ &= -\frac{7}{3}\ln|u| + C \\ &= -\frac{7}{3}\ln|-3x+1| + C.\end{aligned}$$

Not all integrals that benefit from substitution have a clear "inside" function. Several of the following examples will demonstrate ways in which this occurs.

Example 3

Evaluate $\displaystyle\int \sin(x)\cos(x)\,dx$.

Solution. There is not a composition of functions here to exploit; rather, just a product of functions.

In this example, let $u = \sin(x)$. Then $du = \cos(x)\,dx$, which we have as part of the integrand! The substitution becomes very straightforward:

$$\begin{aligned}\int \sin(x)\cos(x)\ dx &= \int u\ du\\ &= \frac{1}{2}u^2 + C\\ &= \frac{1}{2}\sin^2(x) + C.\end{aligned}$$

One would do well to ask "What would happen if we let $u = \cos(x)$?" The answer: the result is just as easy to find, yet looks very different. The challenge to the reader is to evaluate the integral letting $u = \cos(x)$ and discovering why the answer is the same, yet looks different.

Do not be afraid to experiment; when given an integral to evaluate, it is often beneficial to think "If I let u be *this*, then du must be *that* …" and see if this helps simplify the integral at all.

Example 4

Evaluate $\displaystyle\int \frac{1}{x\ln(x)}\ dx$.

Solution. This is another example where there does not seem to be an obvious composition of functions. Again: choose something for u and consider what this implies du must be. If u can be chosen such that du also appears in the integrand, then we have chosen well.

Suppose

$$u = \frac{1}{x} \quad \text{which implies} \quad du = -\frac{1}{x^2}\ dx;$$

however, that does not seem helpful. Suppose we let

$$u = \ln x \quad \text{which implies} \quad du = \frac{1}{x}\ dx;$$

now that is part of the integrand. Thus:

$$\begin{aligned}\int \frac{1}{x\ln x}\ dx &= \int \underbrace{\frac{1}{\ln x}}_{1/u}\ \underbrace{\frac{1}{x}\ dx}_{du}\\ &= \int \frac{1}{u}\ du\\ &= \ln|u| + C\\ &= \ln|\ln x| + C.\end{aligned}$$

The final answer is interesting; the natural log of the natural log. Take the derivative to confirm this answer is indeed correct.

The next two examples will help fill in some missing pieces of our antiderivative knowledge. We know the antiderivatives of the sine and cosine functions; what about the other standard functions tangent, cotangent, secant and cosecant? We discover these next.

Example 5

Evaluate $\int \tan(x)\,dx$.

Solution. The previous paragraph established that we did not know the antiderivatives of tangent, hence we must assume that we have learned something in this section that can help us evaluate this indefinite integral.

Rewrite $\tan(x)$ as $\frac{\sin(x)}{\cos(x)}$. While the presence of a composition of functions may not be immediately obvious, recognize that $\cos(x)$ is "inside" the $\frac{1}{x}$ function. Therefore, we see if setting $u = \cos(x)$ returns usable results. We have that $du = -\sin(x)\,dx$, hence $-du = \sin(x)\,dx$. We can integrate:

$$\begin{aligned}
\int \tan(x)\,dx &= \int \frac{\sin(x)}{\cos(x)}\,dx \\
&= \int \underbrace{\frac{1}{\cos(x)}}_{u}\underbrace{\sin(x)\,dx}_{-du} \\
&= \int \frac{-1}{u}\,du \\
&= -\ln|u| + C \\
&= -\ln|\cos(x)| + C.
\end{aligned}$$

Some texts prefer to bring the -1 inside the logarithm as a power of $\cos(x)$, as in:

$$\begin{aligned}
-\ln|\cos(x)| + C &= \ln|(\cos(x))^{-1}| + C \\
&= \ln\left|\frac{1}{\cos(x)}\right| + C \\
&= \ln|\sec(x)| + C.
\end{aligned}$$

Thus the result they give is $\int \tan x\,dx = \ln|\sec(x)| + C$. These two answers are equivalent.

Example 6

Evaluate $\int \sec(x)\,dx$.

Solution. This example employs a wonderful trick: multiply the integrand by "1" so that we see how to integrate more clearly. In this case, we write "1" as

$$1 = \frac{\sec x + \tan x}{\sec x + \tan x}.$$

This may seem like it came out of left field, but it works beautifully. Consider:

$$\begin{aligned}
\int \sec x\,dx &= \int \sec x \cdot \frac{\sec x + \tan x}{\sec x + \tan x}\,dx \\
&= \int \frac{\sec^2 x + \sec x \tan x}{\sec x + \tan x}\,dx.
\end{aligned}$$

Now let $u = \sec x + \tan x$; this means $du = (\sec x \tan x + \sec^2 x)\ dx$, which is our numerator. Thus:

$$\begin{aligned} &= \int \frac{du}{u} \\ &= \ln|u| + C \\ &= \ln|\sec x + \tan x| + C. \end{aligned}$$

We can use similar techniques to those used in Examples 5 and 6 to find antiderivatives of $\cot x$ and $\csc x$ (which the reader can explore in the exercises.)

Activity 4.6–1

Evaluate each of the following indefinite integrals by using these steps:

- Find two functions within the integrand that form (up to a possible missing constant) a function-derivative pair;
- Make a substitution and convert the integral to one involving u and du;
- Evaluate the new integral in u;
- Convert the resulting function of u back to a function of x by using your earlier substitution;
- Check your work by differentiating the function of x. You should come up with the integrand originally given.

(a) $\displaystyle\int \frac{x^2}{5x^3+1}\,dx$

(b) $\displaystyle\int e^x \sin(e^x)\,dx$

(c) $\displaystyle\int \frac{\cos(\sqrt{x})}{\sqrt{x}}\,dx$

Evaluating Definite Integrals via u-substitution

This section has focused on u-substitution as a means to evaluate indefinite integrals of functions that can be written, up to a constant multiple, in the form $f(g(x))g'(x)$. However, much of the time integration is used in the context of definite integrals involving such functions. We need to be careful with the corresponding limits of integration. Consider, for instance, the definite integral

$$\int_2^5 xe^{x^2}\ dx.$$

Whenever we write a definite integral, it is implicit that the limits of integration correspond to the variable of integration. To be more explicit, observe that

$$\int_2^5 xe^{x^2}\ dx = \int_{x=2}^{x=5} xe^{x^2}\ dx.$$

When we execute a u-substitution, we change the *variable* of integration; it is essential to note that this also changes the *limits* of integration. For instance, with the substitution $u = x^2$ and $du = 2x\,dx$, it also follows that when $x = 2$, $u = (2)^2 = 4$, and when $x = 5$, $u = (5)^2 = 25$. Thus, under the change of variables of u-substitution, we now have

$$\begin{aligned} \int_{x=2}^{x=5} xe^{x^2}\,dx &= \int_{u=4}^{u=25} e^u \cdot \frac{1}{2}\,du \\ &= \frac{1}{2}e^u \Big|_{u=4}^{u=25} \\ &= \frac{1}{2}e^{25} - \frac{1}{2}e^4. \end{aligned}$$

Alternatively, we could consider the related indefinite integral $\int_2^5 xe^{x^2}\,dx$, find the antiderivative $\frac{1}{2}e^{x^2}$ through u-substitution, and then evaluate the original definite integral. From that perspective, we'd have

$$\begin{aligned} \int_2^5 xe^{x^2}\,dx &= \frac{1}{2}e^{x^2}\Big|_2^5 \\ &= \frac{1}{2}e^{25} - \frac{1}{2}e^4, \end{aligned}$$

which is, of course, the same result.

Substitution with Definite Integrals

Let f and g be differentiable functions, where the range of g is an interval I that contains the domain of f. Then

$$\int_a^b f'\big(g(x)\big)g'(x)\,dx = \int_{g(a)}^{g(b)} f'(u)\,du.$$

Let's look at more examples.

Example 7

Evaluate $\displaystyle\int_0^2 \cos(3x-1)\,dx$.

Solution. In this example, we begin with

$$u = 3x - 1 \quad \text{and} \quad \frac{du}{dx} = 3 \Rightarrow \frac{1}{3}du = dx.$$

Rewriting the bounds of integration with respect to u, we have $u(2) = 3 \cdot 2 - 1 = 5$ and $u(0) = 3 \cdot 0 - 1 = -1$. We now evaluate the definite integral:

$$\begin{aligned}\int_1^2 \cos(3x-1)\;dx &= \int_{-1}^5 \cos(u)\cdot\frac{1}{3}\;du\\ &= \frac{1}{3}\sin(u)\Big|_{-1}^5\\ &= \frac{1}{3}\big(\sin(5)-\sin(-1)\big)\\ &\approx -0.039.\end{aligned}$$

Notice how once we converted the integral to be in terms of u, we never went back to using x.

Example 8

Evaluate $\displaystyle\int_0^{\pi/2} \sin(x)\cos(x)\;dx$.

Solution. We saw the corresponding indefinite integral in Example 3. In that example we set $u = \sin(x)$ but stated that we could have let $u = \cos(x)$. For variety, we do the latter here.

Let $u = g(x) = \cos(x)$, giving $du = -\sin(x)\;dx$. The new upper bound is $g(\pi/2) = 0$; the new lower bound is $g(0) = 1$. Note how the lower bound is actually larger than the upper bound now. We have

$$\begin{aligned}\int_0^{\pi/2} \sin(x)\cos(x)\;dx &= \int_1^0 u\;(-1)du\\ &= \int_1^0 -u\;du \quad \text{(switch bounds \& change sign)}\\ &= \int_0^1 u\;du\\ &= \frac{1}{2}u^2\Big|_0^1\\ &= 1/2.\end{aligned}$$

Activity 4.6–2

Evaluate each of the following definite integrals exactly through an appropriate u-substitution.

(a) $\displaystyle\int_1^2 \frac{x}{1+4x^2}\,dx$

(b) $\displaystyle\int_0^1 e^{-x}(2e^{-x}+3)^9\,dx$

(c) $\displaystyle\int_{2/\pi}^{4/\pi} \frac{\cos\left(\frac{1}{x}\right)}{x^2}\,dx$

Summary

In this section, we encountered the following important ideas:

- To begin to find algebraic formulas for antiderivatives of more complicated algebraic functions, we need to think carefully about how we can reverse known differentiation rules. To that end, it is essential that we understand and recall known derivatives of basic functions, as well as the standard derivative rules.
- The indefinite integral provides notation for antiderivatives. When we write "$\int f(x)\,dx$," we mean "the general antiderivative of f." In particular, if we have functions f and F such that $F' = f$, the following two statements say the exact thing:

  $$\frac{d}{dx}[F(x)] = f(x) \text{ and } \int f(x)\,dx = F(x) + C.$$

  That is, f is the derivative of F, and F is an antiderivative of f.
- The technique of u-substitution helps us evaluate indefinite integrals of the form

  $$\int f(g(x))g'(x)\,dx$$

  through the substitutions $u = g(x)$ and $du = g'(x)\,dx$, so that

  $$\int f(g(x))g'(x)\,dx = \int f(u)\,du.$$

  A key part of choosing the expression in x to let be represented by u is the identification of a function-derivative pair. To do so, we often look for an "inner" function $g(x)$ that is part of a composite function, while investigating whether $g'(x)$ (or a constant multiple of $g'(x)$) is present as a multiplying factor of the integrand.

Exercises

Terms and Concepts

1) Substitution "undoes" what derivative rule?

2) T/F: One can use algebra to rewrite the integrand of an integral to make it easier to evaluate.

Problems

In exercises 3–14, evaluate the indefinite integral to develop an understanding of Substitution.

3) $\int 3x^2 \left(x^3-5\right)^7 dx$

4) $\int (2x-5)\left(x^2-5x+7\right)^3 dx$

5) $\int x\left(x^2+1\right)^8 dx$

6) $\int (12x+14)\left(3x^2+7x-1\right)^5 dx$

7) $\int \frac{1}{2x+7} dx$

8) $\int \frac{1}{\sqrt{2x+3}} dx$

9) $\int \frac{x}{\sqrt{x+3}} dx$

10) $\int \frac{x^3-x}{\sqrt{x}} dx$

11) $\int \frac{e^{\sqrt{x}}}{\sqrt{x}} dx$

12) $\int \frac{x^4}{\sqrt{x^5+1}} dx$

13) $\int \frac{\frac{1}{x}+1}{x^2} dx$

14) $\int \frac{\ln(x)}{x} dx$

In Exercises 15–21, use Substitution to evaluate the indefinite integral involving trigonometric functions.

15) $\int \sin^2(x)\cos(x) dx$

16) $\int \cos(3-6x) dx$

17) $\int \sec^2(4-x) dx$

18) $\int \sec(2x) dx$

19) $\int \tan^2(x)\sec^2(x) dx$

20) $\int x\cos\left(x^2\right) dx$

21) $\int \tan^2(x) dx$

In Exercises 22–28, use Substitution to evaluate the indefinite integral involving exponential functions.

22) $\int e^{3x-1} dx$

23) $\int e^{x^3} x^2 dx$

24) $\int e^{x^2-2x+1}(x-1) dx$

25) $\int \frac{e^x+1}{e^x} dx$

26) $\int \frac{e^x-e^{-x}}{e^{2x}} dx$

27) $\int 3^{3x} dx$

28) $\int 4^{2x} dx$

In Exercises 29–31, use Substitution to evaluate the indefinite integral involving logarithmic functions.

29) $\int \frac{\ln^2(x)}{x} dx$

30) $\int \frac{\ln\left(x^3\right)}{x} dx$

31) $\int \frac{1}{x\ln\left(x^2\right)} dx$

In Exercises 32–37, use Substitution to evaluate the indefinite integral involving rational functions.

32) $\int \frac{x^2+3x+1}{x} dx$

33) $\int \frac{x^3+x^2+x+1}{x} dx$

34) $\int \frac{x^3-1}{x+1} dx$

35) $\int \frac{x^2+2x-5}{x-3} dx$

36) $\int \frac{3x^2-5x+7}{x+1} dx$

37) $\int \frac{x^2+2x+1}{x^3+3x^2+3x} dx$

In Exercises 38–47, use Substitution to evaluate the indefinite integral involving inverse trigonometric functions.

38) $\int \frac{7}{x^2+7} dx$

39) $\int \frac{3}{\sqrt{9-x^2}} dx$

40) $\int \frac{14}{\sqrt{5-x^2}} dx$

41) $\int \frac{2}{x\sqrt{x^2-9}} dx$

42) $\int \frac{5}{\sqrt{x^4-16x^2}}dx$

43) $\int \frac{x}{\sqrt{1-x^4}}dx$

44) $\int \frac{1}{x^2-2x+8}dx$

45) $\int \frac{2}{\sqrt{-x^2+6x+7}}dx$

46) $\int \frac{3}{\sqrt{-x^2+8x+9}}dx$

47) $\int \frac{5}{x^2+6x+34}dx$

In Exercises 48–72, use Substitution to evaluate the indefinite integral involving inverse trigonometric functions.

48) $\int \frac{x^2}{(x^3+3)^2}dx$

49) $\int \left(3x^2+2x\right)\left(5x^3+5x^2+2\right)^8 dx$

50) $\int \frac{x}{\sqrt{1-x^2}}dx$

51) $\int x^2 \csc^2\left(x^3+1\right) dx$

52) $\int \sin(x)\sqrt{\cos(x)}dx$

53) $\int \frac{1}{x-5}dx$

54) $\int \frac{7}{3x+2}dx$

55) $\int \frac{3x^3+4x^2+2x-22}{x^2+3x+5}dx$

56) $\int \frac{2x+7}{x^2+7x+3}dx$

57) $\int \frac{9(2x+3)}{3x^2+9x+7}dx$

58) $\int \frac{-x^3+14x^2-46x-7}{x^2-7x+1}dx$

59) $\int \frac{x}{x^4+81}dx$

60) $\int \frac{2}{4x^2+1}dx$

61) $\int \frac{1}{x\sqrt{4x^2-1}}dx$

62) $\int \frac{1}{\sqrt{16-9x^2}}dx$

63) $\int \frac{3x-2}{x^2-2x+10}dx$

64) $\int \frac{7-2x}{x^2+12x+61}dx$

65) $\int \frac{x^2+5x-2}{x^2-10x+32}dx$

66) $\int \frac{x^3}{x^2+9}dx$

67) $\int \frac{x^3-x}{x^2+4x+9}dx$

68) $\int \frac{\sin(x)}{\cos^2(x)+1}dx$

69) $\int \frac{\cos(x)}{\sin^2(x)+1}dx$

70) $\int \frac{\cos(x)}{1-\sin^2(x)}dx$

71) $\int \frac{3x-3}{\sqrt{x^2-2x-6}}dx$

72) $\int \frac{x-3}{\sqrt{x^2-6x+8}}dx$

In Exercises 73–80, use Substitution to evaluate the indefinite integral involving inverse trigonometric functions.

73) $\int_1^3 \frac{1}{x-5}dx$

74) $\int_2^6 x\sqrt{x-2}dx$

75) $\int_{-\pi/2}^{\pi/2} \sin^2 x \cos x \, dx$

76) $\int_0^1 2x(1-x^2)^4 \, dx$

77) $\int_{-2}^{-1} (x+1)e^{x^2+2x+1} \, dx$

78) $\int_{-1}^{1} \frac{1}{1+x^2} \, dx$

79) $\int_2^4 \frac{1}{x^2-6x+10} \, dx$

80) $\int_1^{\sqrt{3}} \frac{1}{\sqrt{4-x^2}} \, dx$

81) For the town of Mathland, MI, residential power consumption has shown certain trends over recent years. Based on data reflecting average usage, engineers at the power company have modeled the town's rate of energy consumption by the function

$$r(t) = 4 + \sin(0.263t + 4.7) + \cos(0.526t + 9.4).$$

Here, t measures time in hours after midnight on a typical weekday, and r is the rate of consumption in megawatts at time t. Units are critical throughout this problem.

(a) Sketch a carefully labeled graph of $r(t)$ on the interval [0,24] and explain its meaning. Why is this a reasonable model of power consumption?

(b) Without calculating its value, explain the meaning of $\int_0^{24} r(t)\, dt$. Include appropriate units on your answer.

(c) Determine the exact amount of power Mathland consumes in a typical day.

(d) What is Mathland's average rate of energy consumption in a given 24-hour period? What are the units on this quantity?

4.7 *More with Integrals*

Motivating Questions

In this section, we strive to understand the ideas generated by the following important questions:

- How can we use definite integrals to determine the average y-value of a function?
- How can we use definite integrals to measure the area between two curves?
- How do we decide whether to integrate with respect to x or with respect to y when we try to find the area of a region?

Introduction

Early in our work with the definite integral, we learned that if we have a nonnegative velocity function, v, for an object moving along an axis, the area under the velocity function between a and b tells us the distance the object traveled on that time interval. Moreover, based on the definition of the definite integral, that area is given precisely by $\int_a^b v(t)\,dt$. Indeed, for any nonnegative function f on an interval $[a,b]$, we know that $\int_a^b f(x)\,dx$ measures the area bounded by the curve and the x-axis between $x = a$ and $x = b$.

In this section, we will further explore definite integrals including some important properties of integrals. In Preview Activity 4.7, we begin this investigation with a refreshment of certain contexts.

Preview Activity 4.7

Consider the following test scores of a particular student.

$$83, \ 75, \ 92, \ 68$$

(a) Calculate the student's average test score.

(b) Would the method used in (a) to determine the average score change depending on the number of tests?

(c) Is it necessarily the case that the student's average test score must equal exactly one of their actual test scores? Why or why not?

Consider a circle of radius 5 and a square of side length 2.

(a) Compute the areas of the circle and square.

(b) Suppose the square is placed wholly inside the circle. How would you compute the area of the region inside the circle but outside the square? What is the area of that region?

The Average Value of a Function

One of the most valuable applications of the definite integral is that it provides a way to meaningfully discuss the average value of a function, even for a function that takes on infinitely many values. Suppose we wish to take the average of n numbers y_1, y_2, ..., y_n. We do so by computing

$$\text{Avg} = \frac{y_1 + y_2 + \cdots + y_n}{n}.$$

Since integrals arise from Riemann sums in which we add n values of a function, it should not be surprising that evaluating an integral is something like averaging the output values of a function. Consider, for instance, the right Riemann sum R_n of a function f, which is given by

$$\begin{aligned} R_n &= f(x_1)\triangle x + f(x_2)\triangle x + \cdots + f(x_n)\triangle x \\ &= (f(x_1) + f(x_2) + \cdots + f(x_n))\,\triangle x. \end{aligned}$$

Since $\triangle x = \frac{b-a}{n}$, we can thus write

$$\begin{aligned} R_n &= (f(x_1) + f(x_2) + \cdots + f(x_n)) \cdot \frac{b-a}{n} \\ &= (b-a)\frac{f(x_1) + f(x_2) + \cdots + f(x_n)}{n}. \end{aligned} \tag{4.8}$$

Here, we see that the right Riemann sum with n subintervals is the length of the interval $(b-a)$ times the average of the n function values found at the right endpoints. And just as with our efforts to compute area, we see that the larger the value of n we use, the more accurate our average of the values of f will be. Indeed, we will define the average value of f on $[a,b]$ to be

$$f_{\text{AVG}[a,b]} = \lim_{n\to\infty} \frac{f(x_1) + f(x_2) + \cdots + f(x_n)}{n}.$$

But we also know that for any continuous function f on $[a,b]$, taking the limit of a Riemann sum leads precisely to the definite integral. That is, $\lim_{n\to\infty} R_n = \int_a^b f(x)\,dx$, and thus taking the limit as $n \to \infty$ in Equation (4.8), we have that

$$\int_a^b f(x)\,dx = (b-a)\cdot f_{\text{AVG}[a,b]}. \tag{4.9}$$

Solving Equation (4.9) for $f_{\text{AVG}[a,b]}$, we have the following general principle.

Average Value of a Function

If f is a continuous function on $[a, b]$, then its average value on $[a, b]$ is given by the formula

$$f_{\text{AVG}[a,b]} = \frac{1}{b-a} \cdot \int_a^b f(x)\,dx.$$

Observe that Equation (4.9) tells us another way to interpret the definite integral: the definite integral of a function f from a to b is the length of the interval $(b-a)$ times the average value of the function on the interval. In addition, Equation (4.9) has a natural visual interpretation when the function f is nonnegative on $[a, b]$.

Consider Figure 4.49, where we see at the top the shaded region whose area is $\int_a^b f(x)\,dx$, in the middle the shaded rectangle whose dimensions are $(b-a)$ by $f_{\text{AVG}[a,b]}$, and at the bottom these two figures superimposed. Specifically, note that in dark green we show the horizontal line $y = f_{\text{AVG}[a,b]}$. Thus, the area of the green rectangle is given by $(b-a) \cdot f_{\text{AVG}[a,b]}$, which is precisely the value of $\int_a^b f(x)\,dx$. Said differently, the area of the blue region in the top figure is the same as that of the green rectangle in the middle figure; this can also be seen by observing that the areas A_1 and A_2 in the bottom figure appear to be equal. Ultimately, the average value of a function enables us to construct a *single* rectangle whose area is the same as the value of the definite integral of the function on the interval.

The java applet at `http://gvsu.edu/s/az` provides an opportunity to explore how the average value of the function changes as the interval changes, through an image similar to that found in Figure 4.49.

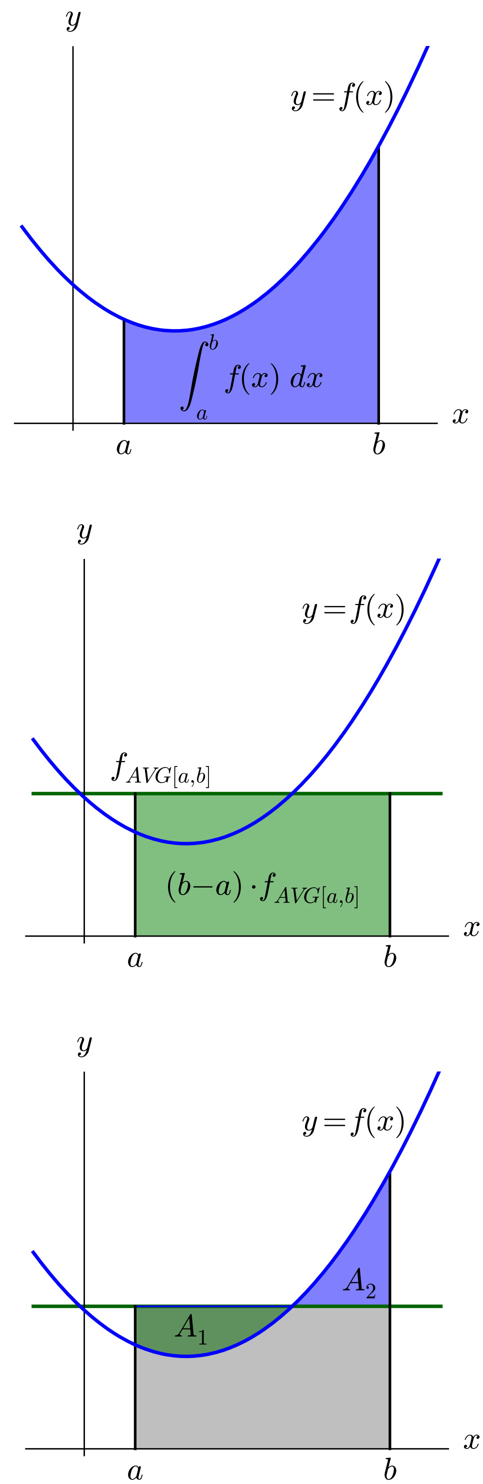


Figure 4.49: A function $y = f(x)$, the area it bounds, and its average value on $[a, b]$.

Activity 4.7–1

Suppose that $v(t) = \sqrt{4-(t-2)^2}$ tells us the instantaneous velocity of a moving object on the interval $0 \le t \le 4$, where t is measured in minutes and v is measured in meters per minute.

(a) Sketch an accurate graph of $y = v(t)$. What kind of curve is $y = \sqrt{4-(t-2)^2}$?

(b) Evaluate $\int_0^4 v(t)\,dt$.

(c) In terms of the physical problem of the moving object with velocity $v(t)$, what is the meaning of $\int_0^4 v(t)\,dt$? Include units on your answer.

(d) Determine the exact average value of $v(t)$ on $[0, 4]$. Include units on your answer.

(e) Sketch a rectangle whose base is the line segment from $t = 0$ to $t = 4$ on the t-axis such that the rectangle's area is equal to the value of $\int_0^4 v(t)\,dt$. What is the rectangle's exact height?

(f) How can you use the average value you found in (d) to compute the total distance traveled by the moving object over $[0, 4]$?

The Mean Value Theorem for Integrals

The average value of a function brings us to an important theoretical result. The Mean Value Theorem for Integrals says that if f is continuous on the interval $[a, b]$, then there exists at least one point c in the interval $[a, b]$ such that $f(c)$ equals the average value of f on $[a, b]$. In other words, the horizontal line $y = f_{\text{AVG}[a,b]}$ intersects the graph of f for some point c in $[a, b]$, or over the interval $[a, b]$, the average value of a function f must equal its actual value at least once.[6]

[6] Compare this statement to the Mean Value Theorem for Derivatives, which says that over the interval $[a, b]$ the instantaneous rate of change of some function must equal its average rate of change at least once.

Mean Value Theorem for Integrals

Let f be continuous on the interval $[a, b]$. There exists a point c in $[a, b]$ such that

$$f(c) = \frac{1}{b-a} \int_a^b f(x)\,dx.$$

Proof: We begin by letting $F(x) = \int_a^x f(t)\,dt$ and noticing that F is continuous on the interval $[a, b]$ and differentiable on the interval (a, b) by the Fundamental Theorem of Calculus. Applying the Mean Value Theorem for Derivatives to F, we can conclude that there exists at least one point c in (a, b) such that

$$F'(c) = f(c) = \frac{F(b) - F(a)}{b-a}.$$

By the FTCII, we know that $F(b) - F(a)$ equals $\int_a^b f(x)\,dx$, so we can write

$$f(c) = \frac{1}{b-a} \cdot F(b) - F(a) = \frac{1}{b-a} \int_a^b f(x)\,dx$$

where c is some point in the interval (a, b). ■

The Area Between Two Curves

In Preview Activity 4.7, we encounter a natural way to think about the area between two curves: the area between the curves is the area beneath the upper curve minus the area below the lower curve.

For the functions $f(x) = (x-1)^2 + 1$ and $g(x) = x + 2$, shown in Figure 4.50, we see that the upper curve is $g(x) = x + 2$, and that the graphs intersect at $x = 0$ and $x = 3$. Note that we can find where these functions intersect by setting the functions equal to each other and solving for x:

$$\begin{aligned} f(x) &= g(x) \\ (x-1)^2 + 1 &= x + 2 \\ x^2 - 2x + 2 &= x + 2 \\ x^2 - 3x &= 0 \\ x(x-3) &= 0 \\ &\Rightarrow x = 0; 3. \end{aligned}$$

On the interval $[0, 3]$, the area beneath g is

$$\int_0^3 (x+2)\,dx = \frac{21}{2},$$

while the area under f on the same interval is

$$\int_0^3 [(x-1)^2 + 1]\,dx = 6.$$

Thus, the area between the curves is

$$A = \int_0^3 (x+2)\,dx - \int_0^3 [(x-1)^2 + 1]\,dx = \frac{21}{2} - 6 = \frac{9}{2}. \quad (4.10)$$

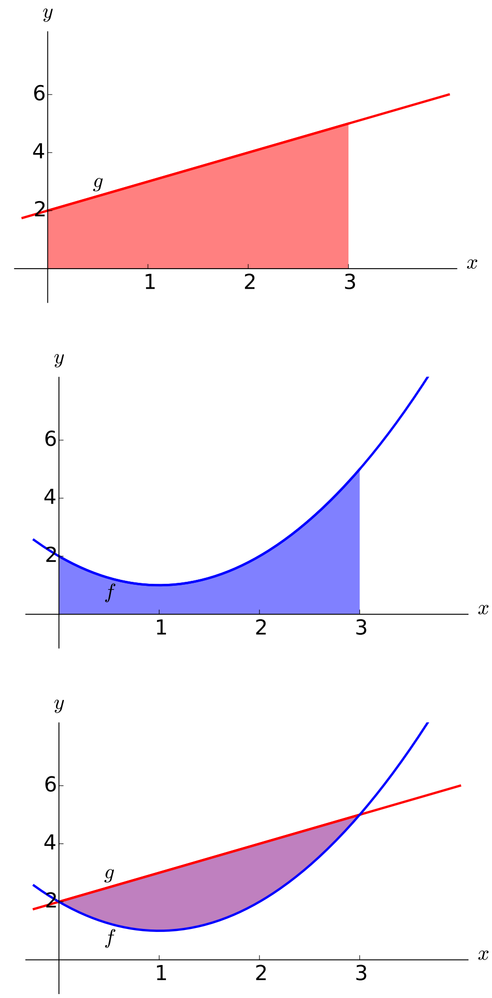


Figure 4.50: The areas bounded by the functions $f(x) = (x-1)^2 + 1$ and $g(x) = x + 2$ on the interval $[0, 3]$.

A slightly different perspective is also helpful here: if we take the region between two curves and slice it up into thin vertical rectangles (in the same spirit as we originally sliced the region between a single curve and the x-axis in Section 4.2), then we see that the height of a typical rectangle is given by the difference between the two functions. For example, for the rectangle shown in Figure 4.51, we see that the rectangle's height is $g(x) - f(x)$, while its width can be viewed as $\triangle x$, and thus the area of the rectangle is

$$A_{\text{rect}} = (g(x) - f(x))\triangle x.$$

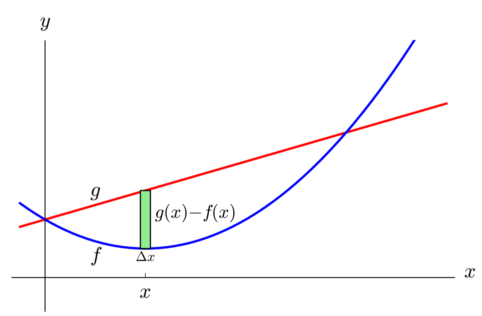


Figure 4.51: The area bounded by the functions $f(x) = (x-1)^2 + 1$ and $g(x) = x + 2$ on the interval $[0, 3]$.

In addition, the area between the two curves on the interval $[0, 3]$ is then approximated by the Riemann sum

$$A \approx \sum_{i=1}^{n} (g(x_i) - f(x_i)) \triangle x,$$

and then as we let $n \to \infty$, it follows that the area is given by the single definite integral

$$A = \int_0^3 (g(x) - f(x))\, dx. \tag{4.11}$$

In our work with applications of the definite integral, we will often find it helpful to think of a "representative slice" and how the definite integral may be used to add these slices to find the exact value of a desired quantity. Here, the integral essentially sums the areas of thin rectangles.

Also, we note that whether we think of the area between two curves as stemming from the difference between the area bounded by the individual curves (as in Equation (4.10)) or as the limit of a Riemann sum that adds the areas of thin rectangles between the curves (as in Equation (4.11)), these two results are the same, since the difference of two integrals is the integral of the difference:

$$\int_0^3 g(x)\, dx - \int_0^3 f(x)\, dx = \int_0^3 (g(x) - f(x))\, dx.$$

Finally, notice that if we had computed

$$\int_0^3 (f(x) - g(x))\, dx \text{ instead of } \int_0^3 (g(x) - f(x))\, dx,$$

then our result would have been $-\dfrac{9}{2}$, which has the same magnitude as the correct result, but the sign is negative. Whereas the net signed area underneath a function can be either positive or negative, the area that is bounded by or between two curves will be positive. Therefore, we can simply take the absolute value of our negative result to get the correct result, leading to the following concept.

Area Between Curves

If two curves $y = f(x)$ and $y = g(x)$ intersect at $x = a$ and $x = b$, then the area between the curves on the interval $[a, b]$

is

$$A = \left| \int_a^b (f(x) - g(x))\, dx \right|.$$

Example 1

Find the area of the region bounded by $y = x^2 + x - 5$ and $y = 3x - 2$.

Solution. We begin by finding the x-values at which the functions intersect.

$$\begin{aligned} x^2 + x - 5 &= 3x - 2 \\ (x^2 + x - 5) - (3x - 2) &= 0 \\ x^2 - 2x - 3 &= 0 \\ (x-3)(x+1) &= 0 \\ x &= -1,\ 3 \end{aligned}$$

Therefore, the area is

$$\begin{aligned} \int_{-1}^{3} \left(3x - 2 - (x^2 + x - 5)\right) dx &= \int_{-1}^{3} (-x^2 + 2x + 3)\, dx \\ &= \left(-\frac{1}{3}x^3 + x^2 + 3x\right)\Bigg|_{-1}^{3} \\ &= -\frac{27}{3} + 9 + 9 - \left(\frac{1}{3} + 1 - 3\right) \\ &= \frac{32}{3} \end{aligned}$$

Activity 4.7–2

In each of the following problems, our goal is to determine the area of the region described. For each region, (i) determine the intersection points of the curves, (ii) sketch the region whose area is being found, (iii) draw and label a representative slice, and (iv) state the area of the representative slice. Then, state a definite integral whose value is the exact area of the region, and evaluate the integral to find the numeric value of the region's area.

(a) The finite region bounded by $y = \sqrt{x}$ and $y = \frac{1}{4}x$.

(b) The finite region bounded by $y = 12 - 2x^2$ and $y = x^2 - 8$.

(c) The area bounded by the y-axis, $f(x) = \cos(x)$, and $g(x) = \sin(x)$, where we consider the region formed by the first positive value of x for which f and g intersect.

(d) The finite regions between the curves $y = x^3 - x$ and $y = x^2$.

Finding Area with Horizontal Slices

At times, the shape of a geometric region may dictate that we need to use horizontal rectangular slices, rather than vertical

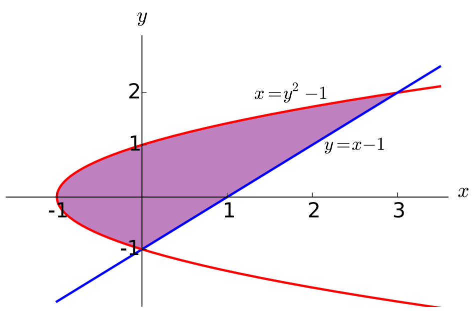


Figure 4.52: The area bounded by the functions $x = y^2 - 1$ and $y = x - 1$.

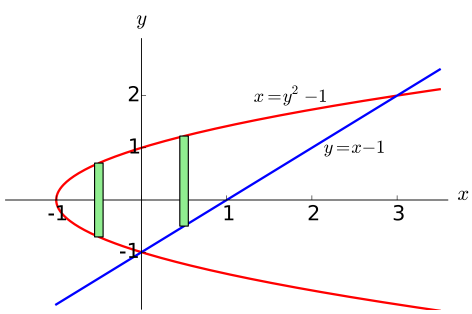


Figure 4.53: The area bounded by the functions $x = y^2 - 1$ and $y = x - 1$ with the region sliced vertically.

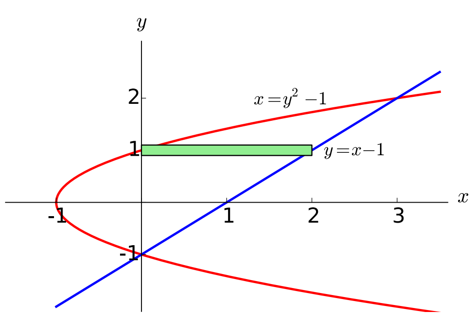


Figure 4.54: The area bounded by the functions $x = y^2 - 1$ and $y = x - 1$ with the region sliced horizontally.

ones. For instance, consider the region bounded by the parabola $x = y^2 - 1$ and the line $y = x - 1$, pictured in Figure 4.52. First, we observe that by solving the second equation for x and writing $x = y + 1$, we can then determine the y-values at which the curves intersect by again setting the curves equal to each other and solving for y:

$$\begin{aligned} y^2 - 1 &= y + 1 \\ y^2 - y - 2 &= 0 \\ (y-2)(y+1) &= 0 \\ &\Rightarrow y = -1; 2. \end{aligned}$$

We can use the y-values to determine the x-values at which the curves intersect, which are $x = 0$ and $x = 3$. We see that if we attempt to use vertical rectangles to slice up the area, at certain values of x (specifically from $x = -1$ to $x = 0$, as seen in Figure 4.53), the curves that govern the top and bottom of the rectangle are one and the same. This suggests, as shown in Figure 4.54, that we try using horizontal rectangles as a way to think about the area of the region.

For such a horizontal rectangle, note that its width depends on y, the height at which the rectangle is constructed. In particular, at a height y between $y = -1$ and $y = 2$, the right end of a representative rectangle is determined by the line, $x = y + 1$, while the left end of the rectangle is determined by the parabola, $x = y^2 - 1$, and the thickness of the rectangle is $\triangle y$.

Therefore, the area of the rectangle is

$$A_{\text{rect}} = [(y+1) - (y^2 - 1)]\triangle y,$$

from which it follows that the area between the two curves on the y-interval $[-1, 2]$ is approximated by the Riemann sum

$$A \approx \sum_{i=1}^{n} [(y_i + 1) - (y_i^2 - 1)]\triangle y.$$

Taking the limit of the Riemann sum, it follows that the area of the region is

$$A = \int_{y=-1}^{y=2} [(y+1) - (y^2 - 1)]\, dy. \tag{4.12}$$

We emphasize that we are integrating with respect to y; this is dictated by the fact that we chose to use horizontal rectangles whose widths depend on y and whose thickness is denoted $\triangle y$. It is a straightforward exercise to evaluate the integral in Equation (4.12) and find that $A = \frac{9}{2}$.

Just as with the use of vertical rectangles of thickness $\triangle x$, we have a general principle for finding the area between two curves, which we state as follows.

Area Between Curves

If two curves $x = f(y)$ and $x = g(y)$ intersect at $y = c$ and $y = d$, then the area between the curves on the interval $[c, d]$ is

$$A = \left| \int_{y=c}^{y=d} (f(y) - g(y))\, dy \right|.$$

Activity 4.7–3

In each of the following problems, our goal is to determine the area of the region described. For each region,

(i) determine the intersection points of the curves,

(ii) sketch the region whose area is being found,

(iii) draw and label a representative slice, and

(iv) state the area of the representative slice.

Then, state a definite integral whose value is the exact area of the region, and evaluate the integral to find the numeric value of the region's area. **Note well:** At the step where you draw a representative slice, you need to make a choice about whether to slice vertically or horizontally.

(a) The finite region bounded by $x = y^2$ and $x = 6 - 2y^2$.

(b) The finite region bounded by $x = 1 - y^2$ and $x = 2 - 2y^2$.

(c) The area bounded by the x-axis, $y = x^2$, and $y = 2 - x$.

(d) The finite regions between the curves $x = y^2 - 2y$ and $y = x$.

Summary

In this section, we encountered the following important ideas:

- The definite integral $\int_a^b f(x)\, dx$ measures the exact net signed area bounded by f and the horizontal axis on $[a, b]$; in addition, the value of the definite integral is related to what we call the average value of the function on $[a, b]$: $f_{\text{AVG}[a,b]} = \frac{1}{b-a} \cdot \int_a^b f(x)\, dx$.
- To find the area between two curves, we think about slicing the region into thin rectangles. If, for instance, the area of a typical rectangle on the interval $x = a$ to $x = b$ is given by $A_{\text{rect}} = (g(x) - f(x))\triangle x$, then the exact area of the region is given by the definite integral

$$A = \int_a^b (g(x) - f(x))\, dx.$$

- The shape of the region usually dictates whether we should use vertical rectangles of thickness $\triangle x$ or horizontal rectangles of thickness $\triangle y$. We desire to have the height of the rectangle governed by the difference between two curves: if those curves are best thought of as functions of y, we use horizontal rectangles, whereas if those curves are best viewed as functions of x, we use vertical rectangles.

Exercises

Problems

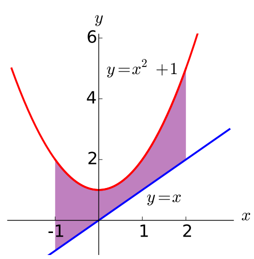
In exercises 1–6, find the average value of the function on the given interval.

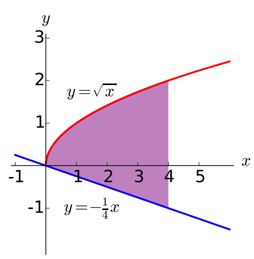
1) $f(x) = \sin(x)$ on $[0, \pi/2]$
2) $y = \sin(x)$ on $[0, \pi]$
3) $y = x$ on $[0, 4]$
4) $f(x) = \dfrac{1}{1+x^2}$ on $[1, \sqrt{3}]$
5) $y = x^3$ on $[0, 4]$
6) $g(t) = 1/t$ on $[1, e]$
7) $f(x) = \sec^2(x)$ on $[-\pi/4, \pi/4]$

In exercises 7–10, find a value c guaranteed by the Mean Value Theorem.

8) $\displaystyle\int_0^2 x^2\,dx$
9) $\displaystyle\int_{-2}^2 x^2\,dx$
10) $\displaystyle\int_0^1 e^x\,dx$
11) $\displaystyle\int_0^{16} \sqrt{x}\,dx$

In exercises 12–13, find the area of the shaded region in each given figure.

12) 

13) 

In exercises 14–17, sketch the given functions and find the area of the enclosed region.

14) $y = 2x$, $y = 5x$, and $x = 3$.
15) $y = -x+1$, $y = 3x+6$, $x = 2$ and $x = -1$.
16) $y = x^2 - 2x + 5$, $y = 5x - 5$.
17) $y = 2x^2 + 2x - 5$, $y = x^2 + 3x + 7$.
18) Suppose that the value of a car in dollars after t years of use is $V(t) = 45,000e^{-0.25t}$. What is the average value of the yacht over its first 8 years of use?
19) Water is run at a constant rate of 3 ft^3/min to fill a cylindrical tank of radius 4 ft and height 6 ft. Assuming that the tank is initially empty, determine the average weight of the water in the tank over the time period required to fill it. Take the weight density of water to be 62.4 lb/ft^3.